SESSION

MOBILITY: MOBILE COMPUTING AND APPLICATIONS + RELATED ALGORITHMS

Chair(s)

TBA

Interference Coordination in Femtocell Networks for QoS Performance Optimization

Jiao Wang[1], Jay Weitzen[1], Volkan Sevindik[1], Oguz Bayat[3], and Mingzhe Li[2]
[1]Department of Electrical and Engineering, University of Massachusetts, Lowell, Massachusetts 01854,USA
[2]Q Factor Communications, 255 Bear Hill Rd. Waltham, MA 02451, USA
[3] Graduate school of Science and Engineering, Istanbul Kemerburgaz University, Istanbul Turkey

Abstract - *Heterogeneous networks with dense small cell deployment cause high inter-cell interference. To mitigate inter-cell interference, many interference coordination (IC) approaches have been proposed. For dense deployment with high interference among cells, traditional forward link IC approaches target on improving edge user throughput for infinite buffer traffic (i.e., FTP download), not necessarily improve QoS performance for delay sensitive (i.e., VOIP) traffic. This research proposes dynamic centralized interference coordination approaches called utility-based SFR (USFR) and utility-based PFR approach (UPFR), to improve forward link performance for both infinite buffer and delay sensitive traffic on dense deployed Enterprise LTE femtocell networks. The cell throughput and user cell edge throughput of infinite buffer traffic, packet loss rate (PLR) of VOIP traffic, have been characterized and compared between proposed approaches and traditional approaches.*

Keywords: femtocell, interference coordination, quality of service (QoS), heterogeneous multi-tier deployment

1 Introduction

Demand for high data rates in wireless mobile networks has triggered the design and development of 4G standards, such as the Third Generation Partnership Project's (3GPP's) Long Term Evolution (LTE), which significantly improves sector capacities and per user data rate, and enables heterogeneous multi-tier deployment. In a typical multi-tier deployment, low power small cells, such as micro-cells, pico-cells and femto-cells, are underlayed within the coverage area of macro-cells to improve coverage and capacity, especially indoors. However, due to the constraint of spectrum resources, operators tend to allocate the same frequency band to neighbor cells or small cells, i.e. with frequency reuse factor of 1, or universal frequency reuse (UFR) [1] to improve the efficiency. This causes higher inter-cell interference (ICI) among UEs who are assigned the same resource blocks (RBs) in neighbor cells. It is especially true for the users located close to the edge of the serving cell and its neighbor cells.

The traditional approach to reduce co-channel ICI is frequency planning among neighbor cells [2]. There are three major static inter-cell interference coordination (IC) techniques: conventional frequency reuse (FFR), Partial frequency reuse (PFR) and Soft frequency reuse (SFR). FFR [3] splits the spectrum among cells, where cells which use the complete set of channels is called a cluster. Cells within the same cluster are assigned different orthogonal frequency band. By dividing the network into clusters, the cells with the same band in different clusters are far away from each other, and this reduces ICI. PFR [4] uses FFR configuration greater than unity for cell edge region and uses FFR configuration of unity for cell center region. The whole bandwidth is divided into N+ 1 segments (N is the number of cells within a cluster). One segment is used by center UEs, other segments are distributed among cells within a cluster for cell edge UE's. SFR [5] splits the frequency band into N segments based on the number of cells within a cluster. For each cell, a dedicated segment (or called prioritized segment) is assigned for cell edge UEs with higher transmit power, other segments (non-prioritized segment) are available for center UEs with lower transmit power. The entire bandwidth is used by all cells, and prioritized segments of different cells within a cluster are orthogonal in frequency.

Given the dynamic nature of network traffic and RF transmission, static IC approaches are sub-optimal solutions and do not provide much overall gain to cell-edge performance without significant penalty to the system performance. Some semi-static and dynamic frequency domain IC approaches are proposed. Adaptive frequency reuse (AFR) [6] is a semi-static SFR technique which adaptively adjusts frequency reuse factor by taking into account traffic loads and data rate requirements of UEs near cell edge. Graphical based interference coordination approaches (GIC) proposed in [7] is dynamic approaches which mitigate ICI by constructing and partitioning interference graph. An interference graph is an undirected graph whose vertices represent UEs, and edges represent interference between UEs, so that UEs connected should avoid the same set of time/frequency resources to reduce ICI. Utility based interference coordination (UIC) proposed in [8] is a technique designed to maximize total network utility that employs a two-level scheme to reduce complexity. At the sector level, the algorithm calculates utilities based on the channel feedback and UEs' demand factor which favor UEs with lower long term average throughput. For coordinated interference control, a central entity processes resource allocation requests from all involved sector elements and resolves conflicting requests based on the utility function values received from sector level algorithm.

With the exception of frequency domain IC (FIC) approaches introduced above, 3GPP-LTE release 10 introduced enhanced IC approaches (eICIC). Almost Blank Sub-frame (ABS) [9] approach is a time domain approach (TIC) allows time sharing of spectrum resources; the basic idea is that the "aggressor" cell (the cell causing severe interference on others) mutes certain sub-frames so that victim cells have a chance to serve their UEs. It is called "almost blank" because the "aggressor" cell still transmit broadcast signals over the ABS. Carrier-aggregation based IC (CBIC) [10] uses multiple component carriers (CCs), inter-cell interference is avoided by transmitting on different CCs among neighboring cells. Every cell always has one component carrier, denoted as the primary component carrier (PCC), which is used for call setup, control channel transmission, and so on. Each cell can dynamically select additional component carriers based on traffic load, referred as secondary component carriers (SCCs). Transmit power can be varied on each CC, and PCC/SCCs are selected in a way that interference among cells are minimized.

The existing IC approaches perform well for sparse deployment of large or medium size cells such as macro cells or micro-cells with a smaller number of interferers and light inter-cell interference. For dense deployment scenario, where each cell receives interference from multiple neighbor cells, the existing dynamic IC approaches, causes large fluctuation in the received power of interference signals. In LTE, eNodeBs transmit data at each transmit time interval (TTI) according to the channel quality indication (CQI) feedback from user equipment (UE). With dynamic IC algorithms and changed RF conditions, the eNodeB has to either re-estimate CQI (base on UE's feedback of signal strength measurements of neighbor cells, which is spectrum consuming) or use the reported CQI. Because the CQI feedback doesn't accurately represent channel condition of current TTI, CQI mismatch [11] happens and the net effect can reduce improvements achieved by interference coordination.

In addition to meeting the high target data rates, heterogeneous LTE networks are also challenged to meet the quality of service (QoS) requirements imposed by an increasing number of real time applications. QoS is defined as the ability of a network to provide a service to an end user at a given service level [12] (priority, latency, bit rate, etc). In LTE networks, each traffic flow is characterized by a set of QoS parameters, such as priority level, maximum acceptable delay, acceptable packet loss rate, etc. Classic interference coordination techniques are focused only on maximizing cell throughput and edge UE throughput for best effort traffic, give little attention to satisfy the QoS requirements of delivered traffic, thus not necessarily improve QoS performance for delay sensitive traffic.

In this research, we propose a novel joint IC approach which combines utility based dynamic IC approach and static IC approaches, with considerations of both infinite buffer traffic (FTP download applications) and delay sensitive traffic (VOIP applications such as VoLTE), in dense deployed LTE femtocell networks. We improve upon the two-level scheme proposed in [8]. The approach introduced in [8] is based on the assumption that the network can be split into pre-defined clusters with predictable inter-cluster dependency, and this 'static" nature of the approach significantly limit the QoS improvement for VOIP traffic, also it is difficult to be implemented in femtocell networks because femtocells are usually deployed at end user locations in an ad-hoc manner. With our approach, the algorithm aims to maximize the network utility by building up graphs to represent interference relationships among femtocells for every resource block (RB). The network is partitioned into clusters dynamically and exhaustively searched for optimized resource allocation solutions among femtocells within each cluster; at femtocell level, the algorithm calculates utilities and schedules traffic flows according to the resource allocated by the central unit, and employs existing static approaches, such as SFR or PFR to reduce RF fluctuation and further improve UE's QoS performance. We demonstrated better QoS performance of our proposed approaches than compared static and dynamic IC approaches.

The remainder of the paper is organized as follows. In Section II, we describe the problem of improving QoS for VOIP traffic. In Section III, femtocell network architecture is explained, and the main components of proposed USFR and UPFR algorithms are presented. In Section IV, simulation methodology and simulation assumptions are described. Simulation results are shown in Section V, complexity analyses are shown in Session VI, and Section VII concludes our paper.

2 Interference coordination for QoS

Voice quality is measured by mean option score (MOS) and according to [17], MOS depends on network delay and PLR. With network delay of 0 to 400 ms, at least 2% of PLR is required to achieve a fair (MOS = 3) quality of voice.

In this session we start with the existing IC approaches, investigate their improvement on PLR for VOIP traffic. Three IC approaches (TIC, CBIC and UIC) introduced in previous session have been simulated and compared with ReUse1. The simulation setup detail is described in session 4 and table 2. For fair comparison, the same bandwidth was used for all IC approaches.

The dynamic TIC method is implemented base on the method introduced in [11]. For simplicity, all femto cells within the network form one cooperative set and jointly decide which cell should enter mute state according to system utility calculation. Each femto cell pre-schedule UEs to its resource blocks (RBs) based on CQI feedbacks and utility is calculated by summing over all RBs. Utility per RB is multiplied by the weight of the UE pre-scheduled on the RB, and the weight is calculated based on spectrum efficiency, average data rate, head of line delay for delay sensitive traffic, and a demand factor that favors UEs with lower long term average throughput.

The dynamic CBIC method is implemented base on the method introduced in [13]. Each femto cell has three CCs and one CC is assigned as PCC. The PCC is assigned in a way to keep interference among neighbors as low as possible. The rest two CCs are used as SCC and will be allocated only when the allocation increases total network utility. Utility is calculated the same way as TIC introduced above.

The dynamic UIC method is implemented base on the method introduced in [8]. Two set of pre-defined clusters are used to cover the femto network. The first set has the max cluster size of 2 (UIC approach 1) and the 2nd set has max cluster size of 4 (UIC approach 2). The same utility calculation as TIC/CBIC is used.

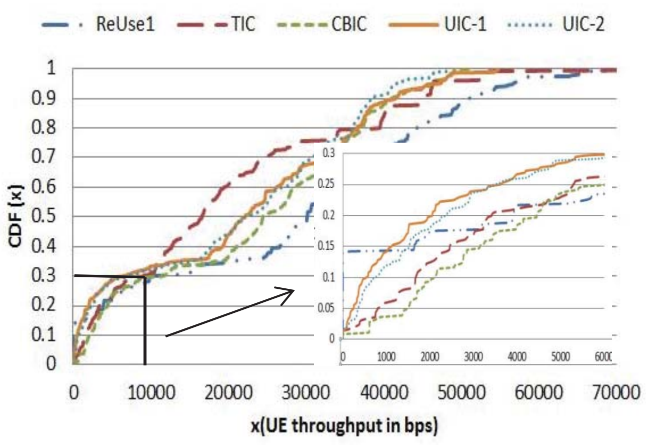


Figure 1: CDF of UE throughput for IC methods

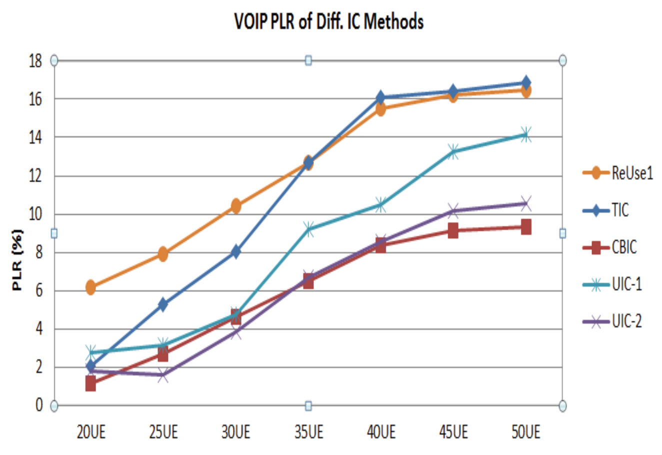


Figure 2: VOIP PLR for IC Methods

Figure 1 shows the cumulative distribution function (CDF) of UE throughput (figure inset shows 30% of lowest throughput samples) for a scenario with 40 UEs per cell, with penetration loss of internal walls set at 7dB. The IC algorithms runs in period of 5 ms.

Figure 2 shows PLR for VOIP traffic for different IC methods, with penetration loss set to 9dB. We can see that even with small number of UEs, most of the IC approaches can't achieve the 2% target.

The observation shows that traditional IC approaches perform well with best effort traffic, but improve little on QoS for VOIP traffic.

The less PLR improvement of TIC/CBIC approaches due to their coarse grained nature, i.e., the IC is performed at per cell or per carrier granularity. For VOIP traffic, each packet has a delay budget, if the packet can't be sent within a time window (in our case, 100 ms), it will be dropped. So muting a cell or carrier can significantly increase PLR, especially for deployment with large number of UEs per cell.

The less PLR improvement of UIC approach dues to its "static" nature of pre-designed cluster set. The RBs of each cell are allocated to UEs dynamically, and UEs, even within the same cell, have different neighbor interferers. The static partitioning of the network can't optimally mitigate interference which varies among RBs.

The above simulation results assume perfect CQI re-estimation, i.e., each UE feedbacks all neighbors' signal strength measurements during IC execution period, and at per RB level. For real implementation, dues to spectrum limitation, perfect CQI re-estimation is hard (if not impossible) to achieve.

Based on above observation and analysis, in order to improve QoS for VOIP traffic, we should develop centralized IC approach to fulfill the following criteria:

1. IC should to be performed for each RB. Complexity problem need to be addressed.

2. For each IC execution, network should be partitioned into clusters dynamically.

3. Dues to dynamic nature of the algorithm, new methodology should be developed to overcome RF fluctuation and CQI mismatch.

3 Description of method

In this session, we present our proposed two-level interference coordination approaches. The two-level scheme refers to the individual femtocell level and central level (runs on a central coordinating entity). Figure 3 illustrates the network topology in which femtocells connect to a central entity directly in a dedicated switching network that is used to coordinate interference among femtocells, for dense deployment, i.e., femtocells within an enterprise. The diagram also shows that a UE camped to femtocell1 and detects femtocell2 as the most dominant interferer on assigned resource blocks.

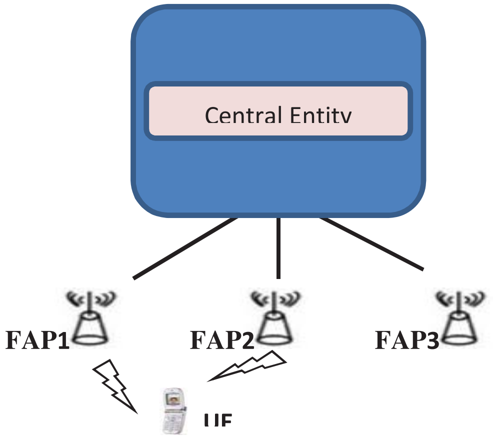


Figure 3: network topology

3.1 UE Feedback and utility calculation

The algorithm starts on each femtocell with a preparatory phase. Consider the scenario in which a UE is camped to femtocell1 and receives interference from its most significant interferer femtocell2 and other neighbors.

For each resource block (RB), the UE feeds back to femtocell two measured SINR values: SINRLow (if femtocell2 does not turn off RB s) and SINRHigh (if femtocell2 turns off RB s). The SINRs are calculated as:

$$SINRLow_s^{u,1}(p_s) = \frac{g_s^{u,1} * p_s^1}{g_s^{u,2} * p_s^2 + \sigma_s^u + \sum_{m\neq 1,2} g_s^{u,m} * p_s^m}$$

$$SINRHigh_s^{u,1}(p_s) = \frac{g_s^{u,1} * p_s^1}{\sigma_s^u + \sum_{m\neq 1,2} g_s^{u,m} * p_s^m} \tag{1}$$

where $g_s^{u,n}$ is channel gain between cell n and user u on RB s; σ_s^u is the thermal noise, and p_s^n as power transmitted by cell n on RB s.

Based on the SINR values, femtocell will calculate utilities using the following equations:

$$UtilityLow_s^{u,1}(p_s) = w_u * \log_2[1 + SINRLow_s^{u,1}(p_s)]$$

$$UtilityHigh_s^{u,1}(p_s) = w_u * \log_2[1 + SINRHigh_s^{u,1}(p_s)] \tag{2}$$

where $w_u$is the weight of UE u, calculated based on spectrum efficiency, average data rate, head of line delay for delay sensitive traffic, and a demand factor [8] that favors UEs with lower long term average throughput. After this step, two matrices, utilityLow and utilityHigh will be available for each femtocell.

To reduce air-link communication complexity, UE feedbacks CQI values instead of SINR values, utility function can be calculated based on CQI values instead.

3.2 Pre-Scheduling

In this step, a temporary user resource allocation indicator matrix is determined, by applying the Hungarian algorithm to the utilityLow matrix to temporarily assign resource to UEs. Hungarian algorithm [8] is a scheduling technique optimal for one to one resource allocation. After this step, each RB of every femtocell will be attached with a temporary scheduled UE, a pair of utility values (utilityLow, utilityHigh) and the most significant interfered femtocell, refers as conflict femtocell. The following table 1 shows an example of the output of step B for femtocell n: the RB1 of femtocell n is assigned to user 1, and it can achieve utility of $UH_1^{1,n}$ if its conflict femtocell (femtocell2) turns off RB1; otherwise, the utility it can achieve $UL_1^{1,n}$.

Table 1: Pre-scheduling results

Resource Block	User	UtilityLow	UtilityHigh	Conflict femtocell
RB_1	U_1	$UL_1^{1,n}$	$UH_1^{1,n}$	Femtocell 2
RB_2	U_3	$UL_2^{3,n}$	$UH_2^{3,n}$	Femtocell 4
⋮	⋮	⋮	⋮	⋮
RB_S	U_x	$UL_S^{x,n}$	$UH_S^{x,n}$	Femtocell y

3.3 Central processing

The results of step B are forwarded to the central entity to find resource allocated among femtocells for each RB. The central level algorithm includes the following three steps:

Step1- Graph construction: build up graph for every resource block: graph G = (N, V), where N nodes represent N femtocells and V edges represent most significant interference relationship between femtocells. The graph is directed; if femtocell A has most significant interferer B, the direction is indicated by drawing an arrow from A (source node) to B (destination node). Figure 4 shows an example graph.

Step2- Graph Partitioning: Breadth First Search [14] is used to partition the graph into clusters, where cluster is defined as a set of nodes within graph that interfere significantly to each other. Femtocells that cause the most interference are treated as distinguished node or head of cluster. The search starts from adding the distinguished nodes into a cluster, and keep adding source nodes of any femtocell already in the cluster, till the cluster reaches a maximum size. Maximum size of cluster is limited by a small value to avoid high computation complexity.

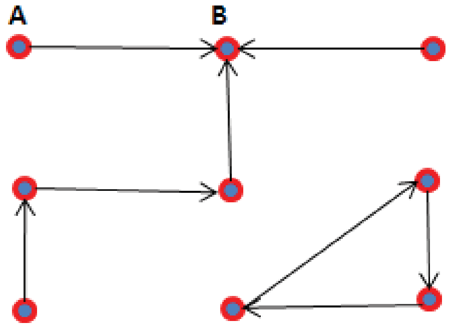


Figure 4: an example of constructed graph

Figure 5 shows the partitioning of the example graph, with maximum cluster size set to 5 (letters within parentheses in cluster C2 are used to denote femtocell. C2 also shows the scenario that if femtocells are equal, any of them can be used as cluster head. Our algorithm chooses the first one (femtocell a) as head).

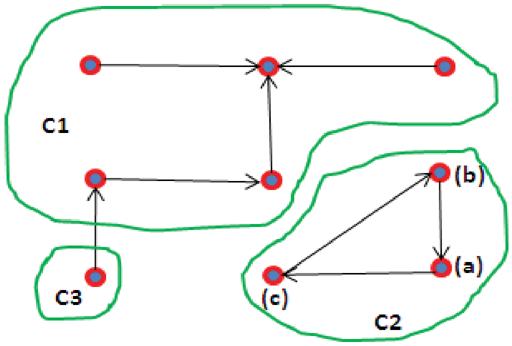


Figure 5: graph partitioning

Step 3- Search within each cluster for an optimal resource allocation solution which achieves the highest utility. For example, the cluster C2 in figure 5 includes 3 femtocells (a, b and c).Thus, 8 permutations exist, and each corresponds to a utility. If (a) is not allocated, both (b) and (c) are allocated, utility corresponding to this permutation will be:

$$Utility_s = \ UtilityHigh_s^{u_x,b} + UtilityLow_s^{u_y,c} \qquad (3)$$

where u_x and u_y are RBs assigned to UEs in step B.

3.4 Data transmission

To reduce complexity, the Central algorithm should run at a period of multiple TTIs. After the central unit determines the resources allocated for each femtocell, it informs femtocells a list of RBs used for next scheduling period. In order to reduce RF fluctuation, existing static IC approaches, such as SFR, PFR have been implemented at femtocell level to further improve edge UE's performance, thus named utility-based SFR (USFR) and utility-based PFR approach (UPFR).

4 Simulation methodology

To evaluate our proposed interference coordination scheme, the open source framework, LTE-Sim [15] was used. The simulation configuration is summarized in Table 2: 9 femto-cells have been mapped into a 3x3 grid layout. Users are randomly generated within each cell, 10 realizations have been used to randomize simulations. The penetration loss of internal walls between femtocells was varied, between 7 dB and 10 dB to vary inter-cell interference levels. For scheduling strategies, proportional fair (PF) scheduling was used for best effort traffic and the Log rule (LOG) was used for VOIP traffic. For VOIP traffic, network delay budget is set to 100 ms (packet will be dropped if it is not received within 100 ms) and target PLR is set to 1%. Each packet transmitted through air-link has probability to drop according to a packet error model and re-transmission will be scheduled with higher priority than normal transmission.

Table 2: Simulation setup

Cell Parameters	
Number of Cells	9, 3x3 grid
Cell Size	square cell with size of 30m x 30m
Cell-center Radius	14 meter
OFDMA Parameters	
Carrier Frequency	2 GHz
Bandwidth	5 MHz
Number of RBs	25
Sub-carrier per RB	12
Sub-carrier Bandwidth	180 KHz
Channel Model	
Path Loss (dB)	46.4+20log(d)+20log(2/5), d in meter
MultipathChannelModel	Jakes model
Shadow Fading Std	8 dB
Penetration Loss-ext. wall	20 dB
Penetration Loss-int. wall	Varies
Power Control Parameters	
Transmit power per cell	20 mW
Edge/Center Power Ratio	2:1
Other Parameters	
Best Effort Traffic	Infinite buffer
VOIP Traffic	on/off Markov Proc., 3s Avg. "on" time
Utility Function	log(.)

5 Simulation results

Figures 6 to figure 9 summarize the performance of the proposed approaches, in comparison with static approaches, including UFR (ReUse1), FFR with frequency reuse factor of one third (ReUse3), PFR and SFR. The same per cell transmit power is used among all methods for fair comparison. Cells are group into clusters, where cluster is a set of cells which use the complete set of channels. For the PFR approach, the frequency band is split into 2 parts: 1st part is used for "center" UEs (10 resource blocks) and 2nd part is used for "edge" UEs, which is further split into 3 segments (5 resource blocks each). Cells are grouped into clusters of size 3, each cell in a cluster uses one segment of the 2nd part for "edge" UEs. The ratio of power per resource block in edge band and in center band is set to 2:1. For the SFR approach, the frequency band is split into 3 segments; cells are grouped into clusters of size 3; cells within a cluster use different segments of the band for "edge" UEs and the rest segments for "center" UEs. The same edge/center power ratio as PFR is used. And within "edge" or "center" band, power is evenly redistributed among RBs used for next scheduling period.

For ReUse3, PFR and SFR, the cluster is arranged in a way that cells use the same "edge" band in different clusters are separated as far as possible to reduce inter-cluster interference. The same setup of SFR and PFR was employed in our proposed USFR and UPFR approaches.

The UIC approach compared with USFR/UPFR is different from [8] in that instead of using "static" cluster set, it dynamically partitions the network into clusters for each IC execution.

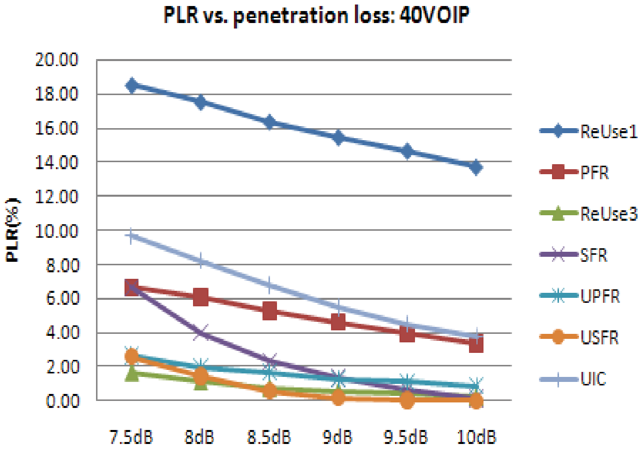


Figure 6: VOIP PLR vs. penetration loss

Figures 6 to figure 8 show simulation results for ideal conditions that femtocell re-estimate CQIs (assuming UE feedback signal strength measurements of all neighbor cells) and schedule data traffic according to re-estimated CQIs. Figure 6 shows PLR for VOIP traffic for different IC methods, with 40 UEs per cell, and varied penetration loss. As penetration loss increases, USFR performs close or even better than ReUse3, which improves PLR performance by sacrificing cell throughput performance.

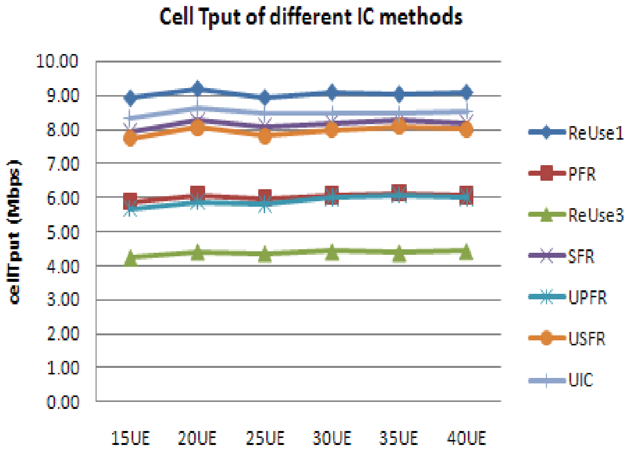


Figure 7: Cell Throughput for IC Methods

Figure 7 compares the average cell throughput for BE traffic for different IC methods with penetration loss set to 9dB. We see that ReUse3, as expected, delivers the worst cell throughput performance. Our proposed approaches, USFR and UPFR, show slight performance drop compare with SFR and PFR.

Our proposed methods show better cell edge throughput than static methods. Figure 8 shows the cumulative distribution function (CDF) of UE throughput (figure inset shows 25% of lowest throughput samples) for a scenario with 40 UEs per cell, with penetration loss of internal walls reduced to 7dB. We can see that cell edge performance are improved by using SFR and UIC, and can be further improved by USFR.

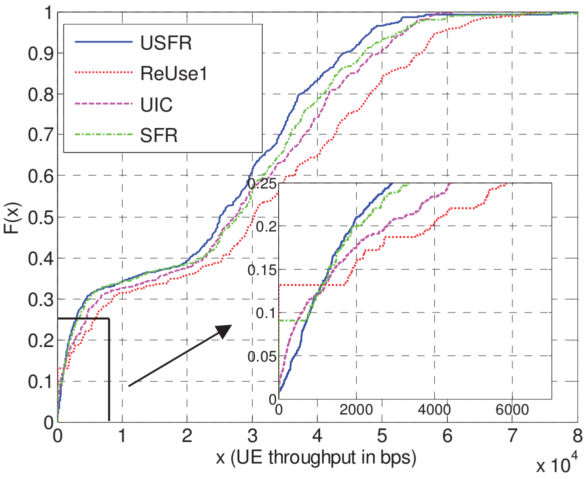


Figure 8: CDF of UE throughput for IC methods

Figures 9 show simulation results with realistic consideration that muting RBs transmit reduced interference caused by reference signal and femtocells scheduling data traffic according to UE's feedback CQI without CQI re-estimation. Penetration loss was set to 9dB and the central level algorithm run at a period of 10 TTI. We see that with high RF fluctuation, UIC's PLR performance drops significantly, but both UPFR and USFR still show significant PLR performance. USFR performs better than all other static and dynamic approaches.

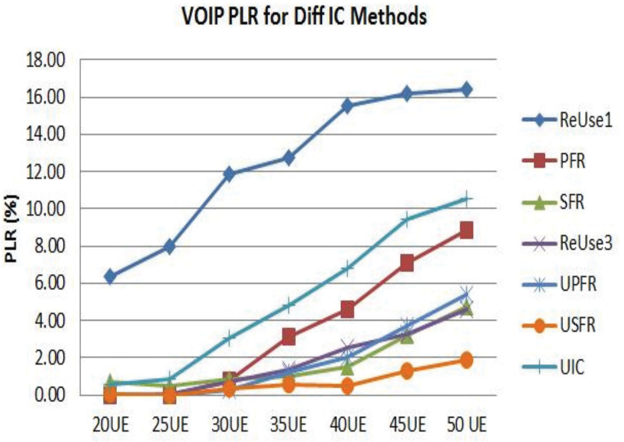


Figure 9: PLR for IC Methods with UE CQI feedback

6 Complexity analysis

In this section, the communication complexity and computational complexity are analyzed. For communication complexity, one extra CQI feedback message is needed per processing period since only most significant interferer's interference are considered, the air-link overhead is small. Pre-scheduling results and central processing results are exchanged between BTS and central unit per processing period, consider BTS and central unit are connected in dedicated switching network such as fiber links or switch, both delay [16] and data volume are less of an issue.

For computational complexity, the sector-level algorithm is dominated by the complexity of the Hungarian algorithm, which is upper bounded by $O(\min(S,U)^2 * \max(S,U))$ [8], where S is the number of RBs and U is number of UEs per cell. At central level, the algorithm runs for each RB independently, thus parallel processing can be used to reduce delays between BTS and central unit, assuming high performance central device with parallel computing capacity. The complexity of the algorithm per RB is dominated by the graph building with complexity of $O(N)$, breadth first cluster partitioning with complexity of $O(N^2)$, where N represents number of cells and exhausted search within each cluster with complexity of $O(2^l)$, where l represents the max cluster size, the max cluster size can be set to a small number to reduce the computational complexity.

7 Conclusion

This paper described novel ICI reduction algorithms: UPFR and USFR. The proposed joint approaches show improved cell edge throughput for best effort traffic and improved packet loss rate for delay sensitive traffic, than corresponding UIC, SFR and PFR approaches. Especially under realistic consideration without CQI re-estimation, our proposed USFR approach outperforms all other approaches in PLR performance, which will have superior effect on delivering high quality VoIP traffic to UEs.

8 References

[1] Ki Tae Kim, Seong Keun Oh. "A Universal Frequency Reuse System in a Mobile Cellular Environment", In Proc. of IEEE VTC 2007-Spring, pp.2855-2859.

[2] A. Eisenbltter, M. Grtschel, and A. M. Koster, "Frequency planning and ramifications of coloring," ZIB-Report 00–47, December 2000.

[3] R. Kwan and C. Leung, "A Survey of Scheduling and Interference Mitigation in LTE," Journal of Electrical and Computer Engineering Article ID 273486, doi:10.1155/2010/273486, vol. 2010, 2010.

[4] Nokia "OFDMA Downlink Inter-Cell Interference Mitigation", 3GPP Project Document R1-060291, Feb. 2006.

[5] Huawei "Soft Frequency Reuse Scheme for UTRAN LTE", 3GPP Project Document R1-050507, May 2005.

[6] Texas Instruments, "Performance of Inter-Cell Interference Mitigation with Semi-Static Frequency Planning for EUTRA Downlink" Tech. Rep. R1-060368, 3rd Generation Partnership Project (3GPP), 2006.

[7] M. C. Necker, "Coordinated fractional frequency reuse," in Proceedings of the 10th ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, October 2007.

[8] M. Rahman and H. Yanikomeroglu, "Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination," IEEE Trans. Wireless Communication., pp. 1414–1425, April 2010.

[9] S. Deb, P. Monogioudis, J. Miernik, and J. P. Seymour, "Algorithms for Enhanced Inter-Cell Interference Coordination (eICIC) in LTE HetNets," IEEE/ACM Transactions on Networking,, 2013.

[10] L. Lindbom, R. Love, S. Krishnamurthy, C. Yao, N. Miki, and V. Chandrasekhar. Enhanced inter-cell interference coordination for heterogeneous networks in LTE-Advanced: A survey (http://arxiv.org/abs/1112.1344). CoRR, abs/1112.1344, 2011.

[11] Wang J, She X, Chen L. Enhanced dynamic inter-cell interference coordination schemes for LTE-advanced. Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th, 2012; 1–6.

[12] D. Soldani, M. Li, and R. Cuny, Eds., "QoS and QoE Management in UMTS Cellular Systems". John Wiley & Sons Ltd, 2006.

[13] B. Wang, et al. "A Cooperative Downlink Power Setting Scheme for CA-Based Femtocells," IEEE Vehicular Technology Conference (VTC Spring), 2012.

[14] E.S. Schaeffer, "Survey: Graph Clustering", Comput. Sci. Rev. 1 (2007)

[15] Giuseppe Piro, Luigi Alfredo Grieco, Gennaro Boggia, Francesco Capozzi, and Pietro Camarda", Simulating LTE Cellular Systems: an Open Source Framework", IEEE Trans. Veh. Technol., vol. 60, no. 2, Feb, 2011, doi: 10.1109/TVT.2010.2091660.

[16] Cisco "Understanding Switch Latency", white paper, June 2012.

[17] L. Sun and E. Ifeachor, "New Methods for Voice Quality Evaluation for IP Networks," in Proceedings of 18th International Teletraffic Congress (ITC-18), (Berlin, Germany), pp. 1201-1210, Sep 2003.

Clusterization-Based Resource Leverage in Hybrid Access Femtocell Networks

Dlamini Thembelihle, Kai-Ten Feng, Jui-Hung Chu, and Pei-Rong Li
Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan
lihles.eed00g@nctu.edu.tw, ktfeng@mail.nctu.edu.tw, jhchu.cm98g@g2.nctu.edu.tw,
and shockbowwow.cm01g@nctu.edu.tw

Abstract— *In this paper, we propose a resource allocation scheme using geometric programming and a novel clustering scheme for femtocells using hybrid access mode in order to maximize the uplink capacity for non-closed subscriber group (Non-CSG) users. We manage the given resource block by sharing the resources between Non-CSG and CSG users. The proposed dynamic resource percentage threshold (D-RPT) scheme reserves resources for Non-CSG users based on the number of CSG users currently being served per femtocell. We also proposed a hybrid femtocell clustering (H-FC) scheme for cluster formation which selects cluster head based on timestamp information. Numerical results show that the proposed algorithms not only improves the uplink capacity but enables more Non-CSG users to be served while still guaranteeing the data rate for CSG users.*

1. Introduction

Femto access points (FAPs) are wireless access points, which provide cost effective means of providing multi-connectivity in the next generation networks [1]. They are low powered, low cost, plug and play devices that are normally installed by the end user and they are connected to the network via a backhaul cable. They are administered by operators and make use of the licensed spectrum technology. The main goal behind this is to improve indoor coverage in current cellular systems due to the fact that consumers are demanding more data [2].

With the explosive growth of mobile data traffic, the femtocell technology is one of the proper solutions to enhance mobile service quality and system capacity for cellular networks. However, the appeal for femtocells gives rise to unsolved problems such as interference, coordination and resource allocation. In dense environment, the interference becomes severe since they are deployed in a small area in large quantities. Therefore, interference minimization remains a major challenge in femtocells operation [3], [4]. Obtaining the optimal resource allocation in dense environment is a non-linear non-convex NP-hard optimization problem [5], [6], which leads to existing work focusing on centralized resource allocation heuristics algorithms.

Furthermore, in [7] [8], different clustering suboptimal heuristic algorithms have been proposed to mitigate inter-femtocell interference. However, they do not mention how the femto head (FH) is elected except for [9]. In our work, we provide a new method for electing the FH using timestamp which avoids the frequent change of leadership expected in [9]. Research studies in [3], [4] do not show how they can overcome the challenges of how to design an effective hybrid access scheme to equilibrate the quality of services of different users since they do not consider different types of users, namely CSG and Non-CSG users, which will be addressed in this work.

Therefore, we propose a resource allocation scheme that reserves resources for Non-CSG users taking into account the total CSG users being served and a suboptimal clustering scheme to mitigate inter-femtocell interference. Our motivation stems from the fact that few studies focus on distributed resource allocation in femtocells that use hybrid access mode. The hybrid access mode allows femtocells to provide preferential access to femtocell owners and subscribers while other public users can access femtocells with certain restrictions [1], [2]. In order to manage the allocated resource block (RB), each FAP will define its own resource percentage threshold (RPT), i.e., the percentage threshold for resources to be reserved for Non-CSG users. Numerical results illustrate that our proposed algorithm can enhance system capacity for hybrid femtocell networks.

2. System Model and Problem Formulation

We consider the uplink (UL) of a dual-tier system, where a dense femtocell network is overlaid on top of the macrocell, and the femtocell network employs frequency division duplexing (FDD). Femtocell scenario takes the form of an enterprise deployment area where there is high density of femtocells, as shown in Fig. 1. Here we assume only a single floor space and femtocells use hybrid access mode. We assume a split spectrum between femtocells and macrocell thus eliminating interference between femto and macro users. Each FAP is responsible for allocation of subcarriers to its femtocell user equipment (FUE). Table I summarizes the notations used in this paper.

Our chief aim is to maximize the UL capacity for Non-CSG users by reserving resources for Non-CSG users based on

[1] This work was in part funded by the Aiming for the Top University and Elite Research Center Development Plan, NSC 102-2221-E-009-018-MY3, the MediaTek research center at National Chiao Tung University, the Industrial Technology Research Institute (ITRI), and the Telecommunication Laboratories at Chunghwa Telecom Co. Ltd, Taiwan.

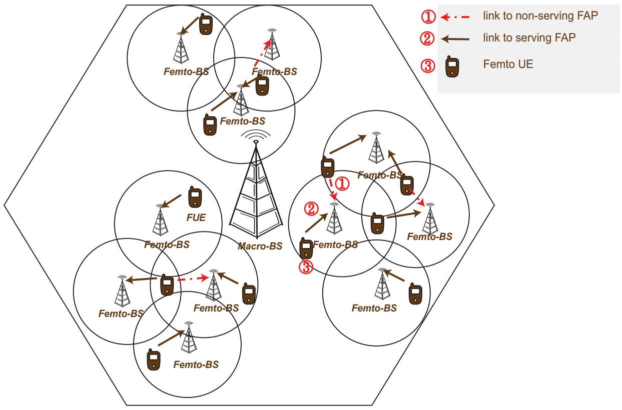


Fig. 1

Clustered femtocells embedded in a macrocell.

the total number of CSG users currently being served; while still guaranteeing the required CSG user's data rate. Also, we will mitigate femto-to-femto interference by clustering hybrid femtocells. We maximize the UL capacity under the constraint of maximum transmission power of FUEs and guarantee data rate for CSG users per femtocell. Our optimization problem can be formulated as

$$\mathbf{P}^* = \arg\max_{\mathbf{P}} \sum_{n=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} C_{n,j,k}^{ncsg} \tag{1}$$

subject to:

$$C1 : \sum_{j=1}^{J} P_{j,k} \leq P_{max}, \qquad \forall k \in \mathbf{K}, \tag{2}$$

$$C2 : \sum_{j=1}^{J} C_{n,j,k}^{csg} \geq C_{req}^{csg}, \qquad \forall k \in \mathbf{K}, \tag{3}$$

where $P_{j,k}$ represent the UL transmission power for user k in subcarrier j and P_{max} is the maximum allowed UL power for each FUE in (2). $\mathbf{P}$ is the set of $P_{j,k}$. C_{req}^{csg} is the minimum data rate threshold to guarantee the data rate for CSG users and $C_{n,j,k}^{csg}$ is the data rate for CSG user k being served by FAP n in subcarrier j. (2) imposes a per FUE constraint on the maximum power, that is, UL transmission power must be lower than maximum power. Constraint (3) denotes that the minimum required data rate for CSG users must be satisfied. The objective function in (1) can be expressed as

$$C_{n,j,k}^{ncsg} = \frac{\beta \gamma_n^{th}}{\lambda_n} \log_2 [1 + \Gamma_{n,j,k}]. \tag{4}$$

The signal to interference plus noise ratio (SINR) $\Gamma_{n,j,k}$ is

$$\Gamma_{n,j,k} = \frac{P_{j,k} L_{n,j,k}}{N_0 + \sum_{m \neq k, m \in \mathbf{K}} P_{j,m} L_{n,j,m}}, \tag{5}$$

where N_0 is the Gaussian noise power and the other term represent the total interference due to other femto users. In the numerator, $P_{j,k}$ is the UL power for user k in subcarrier j and $L_{n,j,k}$ is the path loss between user k and FAP n in subcarrier j.

Since (1) is non-convex due to interference, we use the geometric programming (GP) approach to transform it into a concave function. On the other hand, the optimal clustering problem has been proved to be a NP-hard in [10]. Therefore, we propose a suboptimal clustering scheme for hybrid cells using timestamp and distance. A detailed description of our proposed schemes will be given in the next section.

Table 1

Notation

Symbols	Definition
γ_n^{th}	RPT for Non-CSG users, $\gamma_n^{th} = 1 - \rho_n$
ρ_n	RPT for CSG users, $\rho_n = \frac{C_{req}^{csg}}{\sum_{u=1}^{N_{csg}} \frac{\beta}{N_{csg}} log_2(1+\Gamma_{n,j,u})}$
$\mathbf{N}$	Set of femtocells, $\{f_1, ..., f_N\}$
$\mathbf{J}$	Set of sub-carriers
$\mathbf{K}$	Set of Non-CSG FUE's, $\{U_1, ..., U_K\}$
β	System bandwidth (MHz).
λ_n	Total of Non-CSG users accessing FAP_n
$\mathbf{M}_n^{neighbor}$	Set of FUEs served by neighboring FAPs of FAP_n
$\mathbf{M}_n$	Set of FUEs served by FAP_n
N_{csg}	Total of CSG users accessing FAP_n
$C_{n,j,k}^{ncsg}$	Uplink capacity for Non-CSG user k in subcarrier *j*
$\Gamma_{n,j,k}$	SINR experienced by user k in sub-carrier j in FAP_n
d_{max}	Max distance between Femto Head (FH) and interfering FAP_{new}
d^{th}	Interference distance between FH and 1-hop FAP_n
$\mathcal{M}_T$	Threshold for members per cluster
RPT	Threshold for resources reserved for Non-CSG users
M_{cnt}	Total number of current members in a cluster

3. Proposed Resource Allocation and Clustering Schemes

In this section, we divide our research work into resource allocation and cluster formation. In resource allocation, we perform resource block sharing in a dense environment considering interference in both non-clustered and clustered femtocells. In cluster formation, we propose a clustering method by electing FH based on timestamp.

When a Non-CSG FUE request UL connection, the FAP has to check if the admission control condition is still satisfied by accepting the new Non-CSG. This can be done by calculating the new admission condition, $\beta_{new} = \frac{\beta\gamma_n^{th}}{\lambda_n+1}$. The purpose of admission control is to prevent femtocell overloading resulting to low data rate. Then, we compare the new admission condition with an admission bandwidth threshold, β_n^{ncsg}, which is the minimum equi-spaced channel per FUE of bandwidth 180 KHz similar to 3GPP LTE definition [11]. We impose the following admission constraint to protect CSG users:

$$\beta_{new} \geq \beta_n^{ncsg}, \quad \forall n \in \mathbf{N}. \tag{6}$$

Non-CSG users can either be admitted or rejected. To balance the load over a cluster of femtocells, the system can employ an immediate retry procedure by which the rejected user attempt's to gain service from nearby FAP that still has available resources. This can be further illustrated using Proposition 1 as follows:

Proposition 1 (Admission Control Probability)
When a Non-CSG FUE request uplink connection, the FAP has to check if by accepting the new Non-CSG user it will still meet its admission control condition. Within the cluster, the admission probability, ψ_n, is given by

$$\psi_n = \begin{cases} 1, & \beta_{new} \geq \beta_n^{ncsg}, \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $\beta_{new} = \frac{\beta\gamma_n^{th}}{\lambda_n+1}, \quad \forall n \in \mathbf{N}$.
Proof: The total number of Non-CSG FUEs attempting to connect to FAP_n is given by

$$\lambda_n = \frac{\beta\gamma_n^{th}}{\beta_{new}} - 1 = \beta\gamma_n^{th}(\beta_n^{ncsg})^{-1}. \tag{8}$$

If $0 \leq \lambda_n = \frac{\beta\gamma_n^{th}}{\beta_{new}} - 1 \leq \beta\gamma_n^{th}(\beta_n^{ncsg})^{-1}$, the FAP_n is underloaded thus the admission probability is equal to 1. If $\frac{\beta\gamma_n^{th}}{\beta_{new}} - 1 > \beta\gamma_n^{th}(\beta_n^{ncsg})^{-1}$, the FAP_n is overloaded and the FUE that request uplink connection is blocked/rejected. Thus, the admission probability will be equal to $\frac{\beta\gamma_n^{th}}{\lambda_n+1}$. ■

3.1 Dynamic- Resource Percentage Threshold (D-RPT) Scheme

We propose a resource allocation scheme that allocates resources based on the number of CSG users that are currently being served. The FAP dynamically adjust the resources allocated for Non-CSG users. This scheme guarantees the data rate for CSG users first before allocating the remaining resources to Non-CSG users. Furthermore, Non-CSG users will be admitted only if they meet the admission control condition set in (6). The FAP has to compute the RPT currently dedicated for CSG users, ρ_n, based on the number of CSG users being served and the required data rate for CSG users, C_{req}^{csg}. Then, the FAP has to compute the RPT for Non-CSG users, γ_n^{th} (formula for ρ_n and γ_n^{th} can be obtained in Table I). Once resources have been reserved, we then compute the optimized UL capacity for Non-CSG user using GP after lower bound substitution and variable transformation as the same manner in [12]. This can be further illustrated in Algorithm 1.

Algorithm 1: D-RPT Scheme

```
Input: $\beta, P_{max}, N_0, \beta_n^{ncsg}$
Output: $C_{n,j,k}^{ncsg}$
Each time a CSG FUE is admitted,
compute new $\rho_n$ under the current SINRs
Recompute $\gamma_n^{th}$ under the current
SINR's of its associated FUEs
If Non-CSG FUE request uplink then
    Compute new admission condition, $\beta_{new}$
    If ($\beta_{new} \geq \beta_n^{ncsg}$) then
        Accept Non-CSG FUE
        Compute path loss, $L_{n,j,k}$
        Compute $C_{n,j,k}^{ncsg}$ after lower bound and
        variable
        transformation. Then, maximize $C_{n,j,k}^{ncsg}$.
    else
        Block Non-CSG FUE
    end if
end if
```

3.2 Hybrid-Femtocell Clustering (H-FC) Scheme

Before conducting with the H-FC scheme, the distance between each femtocell pair is necessary to be defined. A femtocell is in general located indoor, and interference occurs when adjacent femtocell use the same subcarrier. The interference level between non-adjacent femtocells is negligible because any signal coming from one FUE travels through at least two walls to reach the FAP of a non-adjacent femtocell. Therefore, we propose a suboptimal heuristic algorithm for cluster formation. We use the Euclidian distance measure as

$$d(f_a, f_b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}, \tag{9}$$

where $d(f_a, f_b)$ is the distance between femtocell f_a and femtocell f_b which are located at (x_a, y_a) and (x_b, y_b), respectively. (x_a, y_a) represents the femto head (FH) and (x_b, y_b) is the neighbor FAP within 1-hop distance.

In this subsection, we propose a suboptimal heuristic algorithm for cluster formation in a dense environment to mitigate femto-to-femto interference based on distance, 1-hop interference, and timestamp, τ_D. Each cluster should have a FH that is elected based on timestamp which avoids the frequent changes of leadership expected in [9]. We assume over-the-air (OTA) coordination to save energy for the close proximity of cluster members. The proposed H-FC scheme overcomes the limitations of other methods by considering how the femto head can be more effectively elected, by having the FH setting d_{max} and by setting the cardinality of cluster members. This scheme is suitable for clustering femtocells in a dense environment where femtocell uses hybrid access mode.

If FH becomes inactive, another FAP is elected as a new FH based on the same criteria. The femto-gateway (F-GW) keeps records of newly deployed FAP and this includes deployment time and date. First, we assume an initially deployed FAP n_0 with no members. FAP n_0 sets d^{th} as the interference distance between FH and 1-hop FAP_n, and the threshold for members per cluster $\mathcal{M}_T$. If a *new* FAP is deployed and interferes with FH, the FH measures the maximum distance, d_{max}. If $d_{max} \leq d^{th}$ and $M_{cnt} \leq \mathcal{M}_T$, then the new FAP joins the cluster, $FAP_{new} \in \phi$, else it may join another cluster or become a new FH. The duty of FH is to form and maintain the cluster, that is, the FH keeps track of active and non-active members. The cluster formation can be described using pseudo-code as in Algorithm 2.

Algorithm 2: H-FC Scheme

```
Assume the presence of $n_0$ as the FH
(label(n)=$\mathcal{H}$) with no members
$n_0$ sets the $d^{th}$ and $\mathcal{M}_T$ [measurements obtained
from “HeNB Sniffer”]
If a new FAP joins a network, i.e., $FAP_{new}$
is switched on within the area, and interfere with
FAP $n_0$ then
    Find the max distance, $d_{max}$
    If ($d_{max} \leq d^{th}$) then
        If ($M_{cnt} \leq M_T$) then
            $FAP_{new}$ becomes the member of the
            cluster, $FAP_{new} \in \phi$
            label(n)=$M$: the status update that node
            new is a member
            Increase membership count, $M_{cnt}$ +1
        else
            $FAP_{new}$ becomes a new FH,
            $FAP_{new} \rightarrow \mathcal{H}$
        end if
    else
        $FAP_{new}$ joins another FH ($FAP_{new} \rightarrow \mathcal{H}'$,
        another cluster) or $FAP_{new} \rightarrow \mathcal{H}$
    end if
FAP $n_0$ updates and shares membership list
Wait for all members to respond
FH periodically checks its active members and
if a member is not determined, status update
becomes X (label(n) =$X$) and $M_{cnt}$ is updated
```

3.3 Dynamic-Resource Threshold (D-RPT) and Hybrid-Femtocell Clustering (H-FC) Scheme

By combining H-FC and D-RPT schemes for hybrid femtocells, the UL capacity of Non-CSG users can be enhanced. We consider inter-femtocell interference caused by femto-users in neighboring femtocells of the serving femtocell. The UL co-tier interference caused by neighboring femto-users, $m \in \mathbf{M}_n^{neighbor}$, at the receiver can be expressed as

$$I_{n,j}^{inter} = \sum_{m \in \mathbf{M}_n^{neighbor}} P_{j,m} L_{n,j,m}, \tag{10}$$

where $P_{j,m}$ is the UL transmission power of user *m* in subcarrier *j* and $L_{n,j,m}$ is the link gain from user *m* to FAP_n in subcarrier *j* . We consider *k* users within the cluster members that are 1-hop from the serving FAP. Therefore, the overall SINR at the FAP_n is

$$\Gamma_{n,j,k} = \frac{P_{j,k} L_{n,j,k}}{N_0 + \sum_{m \neq k, m \in \mathbf{M}_n} P_{j,m} L_{n,j,m} + I_{n,j}^{inter}}. \tag{11}$$

The proposed scheme is named D-RPT + H-FC, and its a combination of Algorithms 1 and 2. Our optimization problem is non-convex due to the presence of inter-cell interference. To transform (1) into a concave formulation, we make use of the GP concept where we introduce alternative variables and approximations similar to [12]. In our problem, we employ the following lower bound as

$$\alpha \log \Gamma_0 + \chi \leq \log(1 + \Gamma_0), \tag{12}$$

which is tight at Γ_0 when the approximation parameters are chosen as

$$\alpha = \frac{\Gamma_0}{1 + \Gamma_0}, \tag{13}$$

$$\chi = \log(1 + \Gamma_0) - \frac{\Gamma_0}{1 + \Gamma_0} \log \Gamma_0, \tag{14}$$

where α and χ are fixed parameters. Therefore, (4) can be reformulated as

$$\hat{C}_{n,j,k}^{ncsg} = \frac{\beta \gamma_n^{th}}{\lambda_n} \left[\alpha \log_2(\Gamma_{n,j,k}) + \chi \right], \tag{15}$$

where $\hat{C}_{n,j,k}^{ncsg}$ can be viewed as the lower bound of $C_{n,j,k}^{ncsg}$. Therefore, the original optimization problem can be transformed to maximize the UL capacity under the constraint of maximum power transmission of FUEs and guarantee data rate for CSG users per femtocell. Nevertheless, (15) is still non-convex which still requires further transformation into a concave function. The lower bound can be transformed into concave by letting $P_{j,k}$ in (15) to be equal to $e^{(\hat{P}_{j,k})}$ and $\hat{P}_{j,k} = \ln(P_{j,k})$. Then we have transformed (15) into

$$\tilde{C}_{n,j,k}^{ncsg} = \frac{\beta \gamma_n^{th}}{\lambda_n} \left[\frac{\alpha}{\ln(2)} \left(\ln(L_{n,j,k}) + \hat{P}_{j,k} - \varphi_{n,j,k} \right) + \chi \right], \tag{16}$$

where $\varphi_{n,j,k} = \ln(\sum_{m \neq k} e^{(\hat{P}_{j,m})} L_{n,j,m} + N_0 + \eta_{n,j,k})$ and $\eta_{n,j,k} = \sum_{m \in \mathbf{M}_n^{neighbor}} e^{(\hat{P}_{j,m})} L_{n,j,m}$. Observing (16), we find a *log-sum-exp* function which has been proven to be convex in [5]. After lower bound variable transformation and approximation, our initial optimization problem in (1) can be reformulated as

$$\mathbf{P}^* = \arg\max_{\mathbf{P}} \sum_{n=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} \tilde{C}_{n,j,k}^{ncsg}(e^{\hat{P}_{j,k}}; \alpha, \chi) \tag{17}$$

Table 2

SIMULATIONS PARAMETERS

System Parameters	Value
Femtocell radius, R_f	10 m
Max number of CSG FUEs per FAP	13
Shadowing effect, ω	4 dB
Wall loss, η	20 dB
Rayleigh fading, ξ	8 dB
FAP transmit power	20 dBm
FUE min. transmit power, P_{min}	0 dBm
FUE max. transmit power, P_{max}	18 dBm
Channel bandwidth, β_f^{ncsg}	180 KHz
Max FUE-FAP distance, D	5 m
Noise, N_0	-174 dBm

subject to:

$$C1 : \sum_{j=1}^{J} P_{j,k} \leq P_{max}, \qquad \forall k \in \mathbf{K}, \qquad (18)$$

$$C2 : \sum_{j=1}^{J} \tilde{C}_{n,j,k}^{csg} \geq C_{req}^{csg}, \qquad \forall k \in \mathbf{K}. \qquad (19)$$

Since the original non-convex problem is transformed into a concave one, $\mathbf{P}^*$ can be readily obtained by the conventional concave optimization methods.

4. Perfomance Evaluation

In this section, we present the results of our proposed schemes considering femtocells using hybrid access mode with a system bandwidth of 10 MHz. We apply the FDD system level simulation assumptions and parameters given in 3GPP specification [11] as summarized in Table II. We consider an single floor enterprise building with room size 10 m x 10 m, where one FAP deployed in each room that uses hybrid access mode.

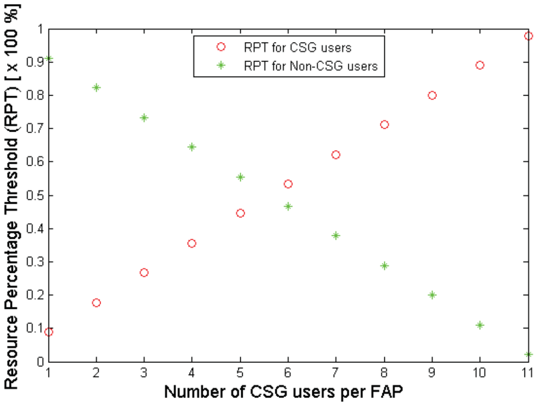


Fig. 2

COMPARISON OF RPT FOR NON-CSG AND CSG USERS.

Fig. 2 shows the comparison of RPT values for Non-CSG and CSG users, which illustrates the variation between the values of γ_n^{th} and ρ_n as the number of admitted CSG users increases. What can be observed is that as FAP_n keeps on admitting CSG users, the value of γ_n^{th} decreases to $\gamma_n^{th} \leq 0$ when user number is greater than 11. Nevertheless, the possibility of having an overloaded FAP with significant amount of CSG users might not be common in a dense environment when the FAPs use hybrid access mode.

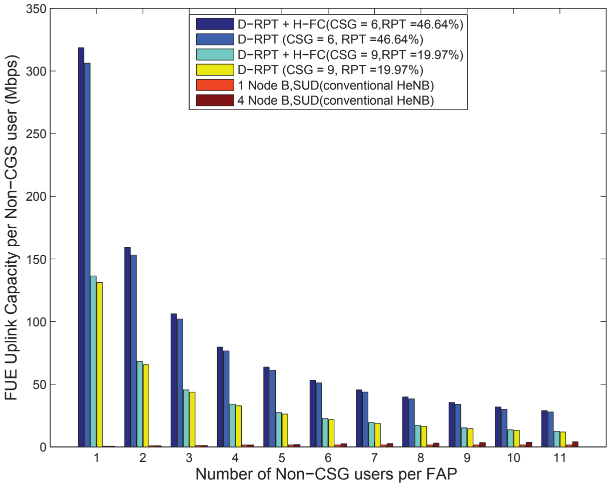


Fig. 3

PERFORMANCE COMPARISON OF FUE UL CAPACITY PER NON-CSG USER.

Fig. 3 illustrates the performance of our proposed D-RPT and D-RPT + H-FC schemes compared with the algorithm used in [13] based on UL capacity per Non-CSG user. In [13], the UL capacity was analyzed by using conventional HeNB and single user detector (SUD) to determine cochannel interference as well as received SINR and a closed loop power control. It is observed that the achievable UL capacity per Non-CSG users decreases with increased number of Non-CSG users being served considering that FUEs performance is limited by interference from other FUEs. However, by combining clustering with our proposed resource allocation scheme, D-RPT, the UL capacity can be greatly improved as clustering reduces the interference impact amongst femtocells. The D-RPT + H-FC outperforms the other schemes and enables the ability to serve a large number of Non-CSG users while still serving CSG users. For example, when the FAP is serving 9 CSG users, the resources reserved for Non-CSG users is 19.97 % and this results to about 11 Non-CSG users being served concurrently with an UL capacity of more than 10 Mbps. The poor performance for conventional HeNBs results from noise saturation at the receiver due to increased number of accepted FUEs.

5. Conclusion

We have proposed a novel clustering scheme for hybrid femtocells where the cluster heads are elected based on timestamp information. The resource allocation scheme is also proposed to dynamically adjust the number of admitted Non-CSG users in femtocells. Simulation results show that the proposed schemes can maximize uplink capacity for Non-CSG while also increasing the number of Non-CSG users being served. The proposed schemes can overcome the challenges of how to design an effective hybrid access scheme to equilibrate the quality of service for Non CSG and CSG users thus optimizing Non-CSG user's uplink capacity.

References

[1] A. Golaup, M. Mustapha, and L. B. Patanapongpibul, "Femtocell Access Control Strategy in UMTS and LTE," *IEEE Communications Magazine*, vol. 47, no. 9, pp. 117–123, Sep. 2009.

[2] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell Networks: a Survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, Sep. 2008.

[3] K. Sundaresan and S. Rangarajan, "Efficient Resource Management in OFDMA Femto Cells," in *Proc. ACM Interational Symposium on Mobile Ad Hoc Networking and Computing*, 2009, pp. 33–42.

[4] V. Chandrasekhar and J. G. Andrews, "Spectrum Allocation in Tiered Cellular Networks," *IEEE Communications Magazine*, vol. 57, no. 10, pp. 3059–3068, Oct. 2009.

[5] M. Chiang, "Geometric Programming for Communications Systems," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 2, pp. 1–156, Aug. 2005.

[6] T. Zhang, "Multi-stage Convex Relaxation for Non-convex Optimization," *Technical report, Rutgers Tech Report*, 2009.

[7] H. Li, X. Xu, D. Hu, X. Qu, X. Tao, and P. Zhang, "Graph Method Based Clustering Strategy for Femtocell Interference Management and Spectrum Efficiency Improvement ," in *Proc. Wireless Communications Networking and Mobile Computing*, Sep. 2010.

[8] F. Tariq, L. S. Dooley, and A. S. Poulton, "Virtual Clustering for Resource Management in Cognitive Femtocell Networks," in *Proc. IEEE International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*, Oct. 2011.

[9] A. Hatoum, N. Aitsaadi, R. Langar, and R. Boutaba, "Femtocell Cluster-based Resource Allocation Scheme for OFDMA Networks," in *Proc. IEEE International Conference on Communications*, Jun. 2011.

[10] S. Sahni and T. Gonzalez, "P-complete Approximation Problems," *Journal of the Association for Computing Machinery*, vol. 23, no. 3, pp. 555–565, Jul. 1976.

[11] *Evolved Universal Terrestrial Radio Access (E-UTRA); FDD Home eNode B (HeNB) Radio Frequency (RF) Requirements Analysis*, 3GPP TR 36.921 V9.0.0, Mar. 2010.

[12] W. C. Ho, L. P. Tung, T. S. Chang, and K. T. Feng, "Enhanced Component Carrier Selection and Power Allocation in LTE-Advanced Downlink System," in *Proc. IEEE Wireless Communications and Networking Conference*, Apr. 2013.

[13] Z. Shi, M. Zhao, M. C. Reed, and H. Wang, "On the Uplink Coverage and Capacity of UMTS Femtocells in Enterprise Environment," in *Proc. 2nd Internation Conference on Femtocells*, 2010.

Measurements of Inter-Femtocell Reverse Link Interference

Jay. Weitzen[1,3], **Theodore Grosch**[2]
[1]Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, Ma, USA
[2] Electrical Engineering, Kennesaw State University, Alpharetta, GA, USA
Airvana, Chelmsford Ma, USA

Abstract - *This paper presents results from a series of experiments using commercial CDMA femtocells to characterize and better understand reverse link inter-femtocell interference. Under conditions of significant RF dragging which can occur with closely spaced femtocells, reverse link interference and potential instability were observed when doing high speed data uploads. The experiments reproduced some of the effects described theoretically in the literature, but only under a very limited set of scenarios..*

Keywords: Femtocells, Interference, Wireless Communication, Cellular Communication

1 Introduction

Residential small cell technology (also known as femtocell) has rapidly evolved from research, development and operator trials [2-11] to deployments exceeding one million units [11]. Small cells now represent an important tool for mobile operators seeking to provide high quality indoor coverage required for data enabled smart phones in suburban and rural residences that are not close to a macrocell. Smallcells use the licensed spectrum of mobile operators, but unlike macro cells that are deployed in a strictly planned way residential femtocells are deployed totally ad-hoc. Early researchers modeled the deployment of femtocells as uniform within the macrocell coverage area, but research described in [11], illustrates that femtocell placement tends to be concentrated in the annular region at the fringe of macro cell coverage. The models used in the literature also asserted that the placement of one femtocell was totally independent of the next femtocell. This also has been shown to not be correct because there likely will be multiple customers of a mobile operator that have a femtocell to fill the same coverage holes. When a coverage hole intersects a large apartment block or other dense residential area, it is likely that there will be multiple customers in the same building or complex having femtocells. Data presented in [11] shows statistics of the distance between nearest neighbor femtocells taken from one of the major operators. It shows that there are now a relatively large number of femtocells within the mutual coverage and interference zone of each other (50 meters or less).

CDMA based third generation (3G) wireless technologies such as CDMA-2000 and UMTS/WCDMA operate using a frequency reuse of 1 and depend for reverse link stability on fast, efficient, soft-handoff power-control to minimize the amount of reverse link transmitter power and therefore the amount of reverse link interference in a CDMA network. In 3G systems, an Access Terminal (AT or UE) in soft handoff can be simultaneously power controlled by up to 8 different sectors and the algorithms select the lowest reverse link power that satisfies the target bit-error rate. Current generation residential femtocells generally do not support soft-handoff either between femtocells or between femtocells and the macro network (most support voice hard handoff to the macro network but not to other femtocells, and generally no data handoff) and there is potential for reverse link interference between femtocells as the density of deployment grows.

"RF dragging" occurs when a handset device at the cell edge actively connected to one cell wants to handoff to another cell because it senses a stronger signal from the new cell, but because handoff is not supported, the handset remains attached to the weaker cell and therefore power controlled by it until the connection drops. The greatest potential for inter-femtocell reverse link interference exists under conditions of RF dragging between closely spaced femtocells when the mobiles are operating at high reverse data rates using either 1xEV-DO or HSUPA in which case the mobile transmitter power is significantly higher than that used for voice calls.

As the density of femtocells increases, the requirement to better understand interference between densely packed femtocells has grown from an academic problem to one of concern to operators of large femtocell networks. This paper presents results from a series of experiments using commercial femtocells and access terminal devices to characterize and better understand the problem of reverse link interference between femtocell mobiles.

A detailed discussion of interference scenarios encountered in femtocell deployments is presented [1]. What is missing from the body of literature described in theoretical studies is a set of measurements using commercial femtocells, and commercial handsets to understand the conditions and severity of these effects. This paper describes experiments using commercial 1xEV-DO Revision A (the high speed data for CDMA-2000) to measure and better understand the effects of femtocell to femtocell reverse link interference due to RF dragging. The experiments focus on uplink data services as opposed to voice because high speed data requires significantly higher uplink transmitter power than low rate voice services.

2 Experiment Design

Figure 1 illustrates how to create the conditions for significant reverse link interference between femtocells. A semi-permanent calibrated test bed that can be used for experimenting when needed was created in the Ball Engineering Building at University of Massachusetts Lowell. Two commercial femtocells, we call F1 and F2 are located in adjacent classrooms separated by concrete brick walls. This use case represents the situation of co-channel femtocells located in adjacent apartments. Each of the femtocells are instrumented with purpose built software tools to periodically query and record the received signal level (RSSI) and the reverse link (uplink) data rate at the femtocell. From these measurements we determine the effective reverse link noise rise-over-thermal and the associated reverse link throughput. A series of calibrated waypoints were created in the halls and classrooms of the building. These are labeled “d” through “i” in figure 1. Generally a connection initiated on femtocell F1 will drop in the region between waypoints denoted “f1” and “f2” because the signal strength from femtocell F2 is much greater than that from femtocell F1 (handoff between femtocells is not supported). At the point denoted "f", the pathloss between femtocells F1 and F2 are both approximately 72 to 77 dB.

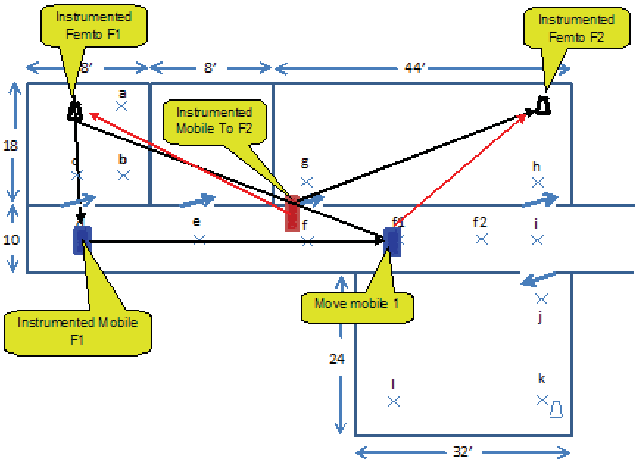


Figure 1. Basic Setup of Inter-femtocell Reverse link Interference Test.

Laboratory calibrated handset devices instrumented with purpose built logging software established connections to their respective femtocells F1 and F2. There is no active connection handoff, either hard or soft, between femtocells other than the call dropping and the handset reselecting the new femtocell.

To serve as reference measurement, a handset device is connected to femtocell F2 (call it AT_{F2}) and is positioned in a static location in the vicinity of the handoff zone between F1 and F2 as shown in Figure 1. Because AT_{F2} is relatively far from F2, it must transmit more power to maintain a constant upload rate. Repeated ftp uploads of large files serve to create a near constant offered load. The AT_{F2} mobile is power-controlled only by femtocell F2, even though it may be able to receive both femtocells F1 and F2 on the forward link. The change in the received signal level at F2 relative to times when there is no activity on femtocells F1 and F2 are recorded and analyzed. The handset device is instrumented to log the forward link requested data rate (DRC), the forward link Received Signal Level, the reverse link transmitter power (both pilot and total power), and all signaling information.

Next, a second handset data device, call it AT_{F1} establishes a connection to femtocell F1 and starts an ftp upload. Device AT_{F1} is moved away from femtocell F1 starting at point “d” towards point “f2” to a point beyond “f1”, just before the call drops, creating a condition known as “RF dragging” where device AT_{F1} is closer in terms of pathloss to femtocell F2 than it is to femtocell F1. Since there is no soft or hard handoff between F1 and F2, AT_{F1} is power controlled only by F1 and a situation of “RF dragging” is created in which F2 should be power-controlling AT_{F1} but cannot. Again the reverse link transmitter power, forward link received signal power and link quality are measured at the mobile terminals and at the femtocells.

The third step is to turn on reverse link interference mitigation and repeat the second step. The data is analyzed by timestamp synchronizing the measurements from the 4 devices (2 femtocells F1 and F2, and two mobile devices AT_{F1} and AT_{F2}).

3 Measurement Results

The first set of experiments was conducted using two laboratory calibrated CDMA-2000 femtocells operating with 1xEV-DO Revision. A. A second identical set of measurements were conducted for CDMA-2000 1xRTT voice, but no significant interference was observed due to the relatively low reverse link transmitter power.

The first step in the experiment was to create the reference or baseline as shown in Figure 2 in which a simple upload from AT_{F2} with AT_{F1} idle is performed. Assuming that the femtocell receiver noise figure is on the order of 10 dB, the base noise level at the femtocells should be on the order of approximately -104 dBm plus or minus one or two dB (the blue trace in figure 2). Looking at the femtocell F2 RSSI we see that the level varies from about -104 dBm when all transmitters are turned off to a maximum high of about -92 dBm (12 dB above the baseline) when there is a single mobile device uploading at about 900 kbps and power controlled by the femtocell. We observe that the AT_{F2} transmitter power (the yellow trace in figure 2) stays in the -20 to -30 dBm range for 900 kbps uplink throughput (the purple trace in figure 2).

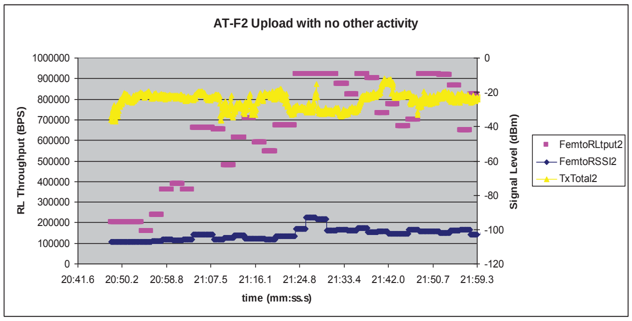


Figure 2. Baseline ftp upload of Femtocell F2 with no other activity.

Figure 3 (see figure caption for explanation of the colors that represent different simultaneously logged parameters) shows the case of two mobile devices simultaneously uploading to their respective femtocells. In the figure AT_{F1} is moving from waypoint “d” towards waypoint “f1” as shown Figure 1 and then stays for a few moments at point “f1” then moves back towards point “d”. As the mobile moves away from its home femtocell it powers up to maintain the required E_b/N_o set point. AT_{F1} causes interference to femtocell F2 which instructs AT_{F2} to power up to overcome the interference from the other mobile, which in turn causes interference to femtocell F1, etc. until both mobiles have reached their maximum power of 23 dBm.

We also notice in Figure 3 that both handset 1 and handset 2 transmitter powers tend to synchronously oscillate. At the time we observed that the femtocell received power level has in some instances increased by close to 40-50 dB over the background thermal noise level of the receiver. This can be seen from a basic link budget calculation in which a mobile with73 dB of pathloss to the closer femtocell and transmitting at maximum 23 dBm produces a received signal level of approximately -50 dBm. Clearly, in this corner case, the reverse link of both femtocells has become unstable and the effective reverse data rate drops to close to zero. This situation represents the limiting case in which handsets connected to closely spaced femtocells and in conditions of RF dragging and high power simultaneous data uploads causes significant mutual reverse link interference to both femtocells.

Correlation analysis between the total uplink transmitted power of ATF2 and AT_{F1} showed that under the conditions of RF dragging the transmitter powers of the two handsets became on the average 92% correlated. This explains the mutual oscillatory behavior as the power control of the two femtocells tries to compensate for the interference rise.

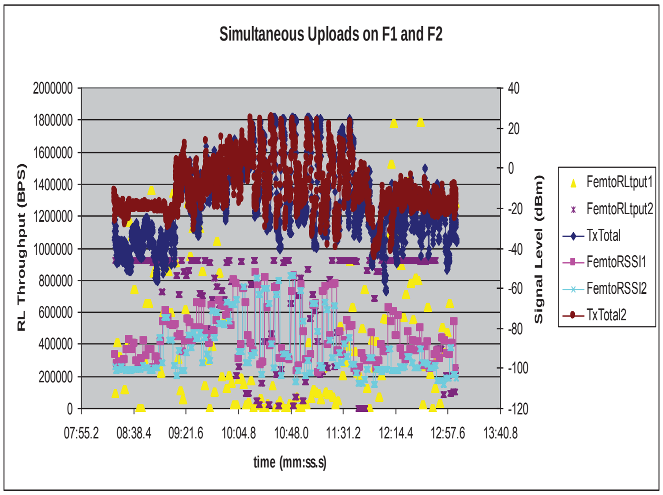


Figure 3. Simultaneous uploads on AT_{F1} and $AT_{F2.}$ Mobile moves from point d to point f1 in Figure 1. Brown trace: Total uplink Tx Power for UE2. Dark Blue: Total Tx Power for UE1. Yellow: Femtocell F1 Reverse link throughput, Magenta: Femtocell 2 Reverse Link Throughput, Light Blue: Femtocell 2 Total Reverse Link RSSI, Light purple Femtocell 1 Reverse link RSSI.

3.1 3.1 Reverse link interference mitigation

The ideal technique to control this type of reverse link instability would be to form ad-hoc clusters of femtocells that can support either hard or soft handoff within the cluster. In the absence of inter-femtocell handoff, other forms of reverse link interference mitigation must be considered. The first technique can be implemented in the Automatic Network Planning Function [5] when femtocells are initially provisioned so as to pull in the coverage of closely spaced units so there is minimal overlap of the forward link coverage and calls drop much earlier before creating significant interference.

A second technique for controlling this type of mutual interference is to significantly reduce the reverse link transmitter power (and therefore the data rate) of handsets when then the conditions of mutual RF dragging are sensed. There are two elements to successful reverse link interference mitigation algorithms using this technique: detection and reaction. In most cases detection attempts to sense high levels of interference on the forward link at the handset combined with high reverse link transmitter power that imply that there is mutual interference on the reverse link and the algorithm reacts to reduce the data rate of the mobile terminal (even reducing to 0) to reduce the total transmitter power and therefore the interference.

The forward link measurements can be done either based on primary measures of interference such as Ec/Io reported by the handset or from secondary measures of signal quality such as DRC (1xEVDO data rate control) or CQI (HSDPA channel quality index). When a femtocell detects that the handset is seeing significant forward link interference, it reduces the uplink data rate and therefore the transmitter power from the handset, potentially reducing throughput to 0 depending on the level of interference.

In the limiting case the femtocell terminates the data call in which the AT either attempts to reselect and then register on the other femtocell (open access) or go to the macro network, but in either case the RF dragging induced interference will be stopped.

4 Conclusions

Measurements described in this paper were designed to create and better understand the scenarios described in [1] in real deployments. The use case of greatest concern in real world deployments occur when "RF dragging" between handsets connected to closely spaced femtocells can lead to significant reverse link interference when users are doing simultaneous high speed uploads.

For voice calls where the total reverse link transmitter power is low and given the path losses involved between handsets and femtocells, we were unable to reproduce the conditions of reverse link instability. In addition almost all femtocells support voice handoff which moves handsets off the femtocell near its cell edge and eliminates the problem of RF dragging.

For the case of high speed reverse link data upload with two handsets in condition of RF dragging, reverse link data rates and transmitter power were observed to oscillate under conditions of sustained upload. In other words, the reverse link became unstable. Constraining the handset transmitter power by reducing uplink data rates, including the pilot, was able to mitigate some, but not all, reverse link interference. In the limit the best approach other than forming ad-hoc clusters that support handoff is to require the algorithms that manage forward link coverage (the centralized network planning function) pull in the coverage of both of the femtocells so that RF dragging is minimized. Sensing and then disconnecting (dropping) calls that are in conditions of RF dragging is a final approach to interference management.

The good news from the study is that the conditions leading to inter-femtocell reverse link interference scenarios described in [1] are relatively rare and can be mitigated through a combination of RF planning (having the two femtocells pull in their coverage to minimize the RF dragging potential) and reverse link interference mitigation. As the femtocell density increases, the importance of algorithms in the femtocell to manage reverse link femtocell to femtocell reverse link interference increases.

5 References

[1] Femto Forum Online White Paper: "Interference Management in UMTS femtocell systems", December 2008, www.femtoforum.com

[2] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell Networks: A survey," *IEEE Comm. Magazine*, vol. 46, pp. 59–67, Sep. 2009

[3] J. Weitzen and T. Grosch, "Comparing Coverage Quality for Femtocell and Macrocell Broadband Data Services, *IEEE Communication Magazine*, Jan 2010, pp 40-44

[4] L. Ho and H. Claussen, "Effects of User-Deployed, Co-Channel Femtocells on the Call Drop Probability in a Residential Scenario," *The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Communications*, pp.1 – 5, Sep 2007.

[5] J. Weitzen, B. Raghothaman, A. Srinivas "Managing Coverage and Interference in UMTS Femtocell Deployments", Book Chapter in *"Evolved Cellular Network Planning and Optimization for UMTS and LTE*" published March 2011, CRC Press, L. Song and J. Shen Ed.

[6] G. de la Roche, A. Valcarce, D. López-Pérez, and J, Zhang, Access Control Methods for Femtocells, *IEEE Communication Magazine*, pp 33-39, January 2010

[7] D. Calin, H. Claussen, and H, Uzunalioglu, On Femto Deployment Architectures and Macrocell Offloading Benefits in Joint Macro-Femto Deployments, *IEEE Communication Magazine*, January 2010, pp 26-32

[8] D. López-Pérez, A. Valcarce, G. de la Roche, and Jie Zhang**,** OFDMA Femtocells: A Roadmap on Interference Avoidance, *IEEE Communications Magazine*, September 2009, pp 41-48

[9] P. Humblet, B. Raghothaman, and A. Srinivas, S, Balasubramanian, C. Patel, and M. Yavuz, System Design of CDMA2000 Femtocells, *IEEE Communications Magazine*, Sept 2009, pp 92-100

[10] M. Yavuz, F. Meshkati, and S. Nanda, A. Pokhariyal and . Johnson, B. Raghothaman and A. Richardson**,** Interference Management and Performance Analysis of UMTS/HSPA+ Femtocells, *IEEE Communications Magazine*, Sept 2009, pp. 102-109

[11] Weitzen, J.A, Li M., Anderlend, E and Eyuboglu, V., Managing and Measuring the performance of large femtocell networks, *IEEE Proceedings, Nov 2013*

Case Study for a HighLy Portable Mesh nEtwork (H.L.P.-M.E.)

L. P. O. Sousa[1], S. J. Bachega.[3], J. Martins Jr.[4], A. C. Oliveira Jr.[2],
M. A. Batista[2], T. A. Santos Filho[2], S. F. da Silva[2] and D. M. Tavares[2]
[1]Industrial Mathematics Department, Federal University of Goiás (UFG), Catalão, Goiás, Brazil
[2]Computer Science Department, Federal University of Goiás (UFG), Catalão, Goiás, Brazil
[3]Production Engineering Department, Federal University of Goiás (UFG), Catalão, Goiás, Brazil
[4]Computer Engineering Department, College of Campinas (FACAMP), Campinas, São Paulo, Brazil

Abstract— *In this paper, we explore a prelude implementation for a portable wireless mesh network, intended to enable multimedia communication with no onsite infrastructure. This is intended as a perimeter network for the fast and secure communication of devices (e.g. robots, IP cameras, notebooks, wifi sensors, etc.) in an environment with no network coverage (e.g. due to a natural disaster, as communication support during a sting operation etc). This kind of environment must be simple to configure, and it must support some kind of mesh network implementation for easy deployment. We estimate that by owning such communication infrastructure, for instance, law enforcement agencies would be able to perform a diverse scope of operations in an easy and efficient manner, preferably in the context of a MAN, which must be independent of landlines, and would allow for the transmission of multimedia data seamlessly (e.g. audio, video, GPS coordinates etc).*

Keywords: portable wireless mesh network, B.A.T.M.A.N.

1. Introduction

Wireless Mesh Networks or WMNs are computer networks that interconnect a set of nodes, where each node is capable of forwarding packets, until they reach a given destination. Therefore, each node can act as a router or client allowing for more mobility and flexibility regarding the infrastructure organization [1].

Mesh routers are capable of communicating heterogeneous networks, like sensor networks (assuming one of the mesh nodes acts as a sink) and usual wifi devices. Besides, one of the nodes can share Internet access to a whole section of the mesh (depending on the size of the mesh network). Mesh nodes can also automatically establish a backbone network and keep the connectivity among mesh clients [2]. In comparison to a conventional router, a mesh router achieves the same range at a lower transmission power, thanks to multi-hop communication. Mesh clients usually have only one network interface and act as both end users (i.e. with Internet access) and routers [3]. Mesh nodes traditionally use the IEEE 802.11 standard [4] in order to communicate.

After the rise of Wi-Fi, lots of applications that partly used landlines were developed. Using as motivation the need to improve the services offered by wireless networks and also to reduce the dependency of landlines, the mesh concept emerged [5]. This technology is already in widespread use, for example, in community or food squares, airports, shoppings, hotels, isolated places (e.g. mountainous regions), universities etc. There are scenarios where this technology is used in a more broad fashion, as in the Dharamsala community in India, where a mesh network was deployed. According to [6], even with a mountainous terrain and with more than two thousand computers interconnected, the performance was satisfactory in the devised tests. Microsoft's Self Organizing Wireless Mesh Networks project uses the user's computer with a Windows driver, which creates a virtual layer between the network and data link layers. This project also has a framework to manage mesh network failures. The analysis is done by event simulations allowing the diagnostics of problems and traffic conditions [7]. RoofNet is another project that deploys a mesh network in a densely populated 4 square kilometer area at Cambridge, Massachusetts, using volunteer users and 37 mesh node kits, in order to share a fraction of their DSL lines [8]. According to [9], systems based on mesh architecture are a viable solution when compared to a hypothetical single-hop network. In this sense they increase the connectivity and the data transfer rate.

The rise of wireless mesh networks is due to its advantages when compared to the traditional wireless network model. The main advantage is the easiness of expansion thanks to the possibility of a mesh client acting also as a router. This turns this network model easy to deploy and low cost allowing access to places where cabled networking would be impracticable [6].

The effectiveness of any network architecture, including mesh networks, depends on the routing protocol used. The routing protocol is the responsible for transmitting information from a source to a destination hopping through intermediate nodes [10]. The challenge is to find the most effective route. In this paper we present the behaviour and the features of the Better Approach To Mobile Ad-hoc Networking (B.A.T.M.A.N.) protocol and define as hour research hypothesis the possibility to implement a highly portable mesh network using off-the-shelf cost effective equipment with minimum downtime for configuration.

In the next sections we will present the theoretical background behind mesh networks, a brief explanation of the B.A.T.M.A.N. protocol, given it is used in our experimental testbed, the application context for our proposal and some final thoughts regarding our proposal.

2. Theoretical Background

A traditional computer network contains a centralised controller for each node. In a mesh network there is no need for a controller, taking into account that the users themselves can expand the coverage area [12]. Therefore, mesh networks present a dynamic feature, in which by adding or removing nodes in the network does not compromise the network connectivity. That happens because the nodes can be connected to more than one node, and that way, the network cost decreases considering there is no need for a more “formal” maintenance policy [5].

The topology of a traditional network obeys a hierarchy where the devices can only be accessed inside of their coverage area. In a mesh network, the network topology is defined in such a way that all the devices in the network can be a part of the transmission path [15], resulting in a more effective transmission. Besides, mesh networks are also fault tolerant [1], due to the mesh nodes’ capabilities to act as clients or routers, allowing a variety of paths among nodes during packet transmission. Other mesh network feature is the support to ad-hoc networking, which is an operational mode that provides the ability to self-generation, self-maintenance and self-organization [16]. Note that the main characteristics of mesh networks, like flexibility and lack of a predefined infrastructure, are appropriate to the proposal of this article. In the US, this technology is already being used in military applications, seeking a communication infrastructure that is independent of the traditional landlines and also fault tolerant [18].

A lot of research fields in mesh networks involve the the study of routing protocols. Although there are several protocols, there is no universal choice [10]. The routing protocols operate generally in the network layer, where their main function is to issue packets from a source node to a destination node. The protocol also specifies the way the routers communicates among themselves, giving access between any two nodes in the network [19]. The problem of classic routing protocols is that they were not created considering the features of ad hoc wireless networks. This genre of network changes its topology according to the inclusion/exclusion of nodes, fact that was not envisioned in traditional routing protocols. For instance, Optimized Link State Routing Protocol (OLSR) had to go through some changes in its original specification, due to the specifics of a link state algorithm which has to recalculate all the topology for each node [20].

In this context, our research group studied applications and the principles involved in creating routing protocols applied to mesh networks [13], [14]. Each routing protocol is devised using different principles and features. To help comprehend these differences, the protocols are classified in proactive, reactive and hybrid. The proactive protocols are based on predefined tables that keep track of the routes for any possible destination and are updated at each topology change. Protocols like Wireless Routing Protocol (WRP) and OLSR are examples of proactive protocols. Reactive protocols stipulate that each node only keeps track of its neighbours when there is the need for it to communicate, a bigger delay is only generated if a new path is necessary. Dynamic Source Routing (DSR) is an example of such type of protocol. Hybrid protocols use conveniently the features of both proactive and reactive protocols, in such a way that in a set of nodes, only some of them do a periodic update of the possible destinations. An example of such protocols is Zone Routing Protocol (ZRP) [5], [14]

B.A.T.M.A.N. is a proactive protocol that identifies only the best next hop instead of discovering the complete route [21], [22], [23], [24]. Therefore, there is no need for the global knowledge of all the changes in the network topology. Besides, the overall number of messages that floods the mesh topology is limited, avoiding control traffic overload [20]. Considering the intended scenario (most likely some kind of sting operation performed by the authorities), B.A.T.M.A.N. seems as one of the possibilities for a routing protocol with its performance improved given the use of a limited quantity of mesh nodes for temporary coverage of an area for a short period of time. Internet connection is not an issue for the sake of the depicted scenario. Besides, according to [28], a high node density limits the network ability to cope with a large amount of hops in the transmission path. Therefore, the relatively short number of hops for this kind of deployment favours the use of B.A.T.M.A.N.

2.1 B.A.T.M.A.N. Protocol

B.A.T.M.A.N. routing protocol was devised to operate in non-reliable media with high levels of instability and packet loss, instead of the stable and reliable media used by traditional cabled networks. The protocol’s algorithm proposes the decentralization of the knowledge about routes among B.A.T.M.A.N. nodes. These nodes have no information whatsoever regarding the overall network routing, allowing low battery and CPU consumption for each node. Instead of discovering the complete route to the destination node, a router only identifies the best next hop to achieve a given node. A node detects the presence of B.A.T.M.A.N.-Originators, regardless of the number of hops (single-hop or multi-hop) to/from an B.A.T.M.A.N.-Originator. It also keeps track of new B.A.T.M.A.N.-Originators and informs its neighbours about their existence [25].

Originator Messages (OGMs) inform neighbouring nodes about their existence. The messages must be transmitted in a given time interval (ORIGINATOR_INTERVAL). An OGM packet has a field for: its version, a field to inform if the node is a direct neighbour or not, an unidirectional flag, a desired value for the Time-To-Live (TTL), a gateway flag (to inform if it is a node with Internet access), a sequence number used for the packet identification and an originator address (IPv4 address of the B.A.T.M.A.N. interface on which the OGM has been generated). When a node receives an OGM it must check: if the OGM contains the same version, if the OGM address is not the broadcast address of a B.A.T.M.A.N. interface and if the OGM is defined as a bidirectional link (capable of full-duplex communication) [25].

If the previous conditions are met, OGM information must be updated. If the sequence number of the received OGM packet is more recent than the one seen before, the new sequence number must be defined to the sequence number of the received OGM packet, and the last TTL of this neighbour must be updated. The window of all known links of the OGM packet must be updated to reflect the new boundaries of the classification range, and the sequence number of the received OGM must be added to the window that represents the link that was held. If the link window whose OGM was received contains the sequence numbers bigger than in its range table, this link is said to be the new best binding to the OGM originator; otherwise, there are no changes. When an OGM is retransmitted, its TTL must be reduced (in case it becomes zero, the packet must be discarded) [25].

Each node that receives an OGM must retransmit the message, therefore flooding the network. The network is flooded until each node has received an OGM at least once, or until happens packet loss (that can happen due to interference, collision or traffic congestion), or until its TTL value expires. Using the data obtained from each OGM, it is possible to distinguish new messages from duplicates, assuring that all OGMs are counted only once. The amount of OGMs received is used to estimate the quality of a route (single-hop or multi-hop). That way, B.A.T.M.A.N. protocol allows each node to keep a table with the best neighbouring nodes in the network [25].

3. Application Context

This paper is inserted in the context of a major project called “Mobile mEsh Network to Aid in CountEring drug TRAffiCKing (M.E.N.A.C.E-TRACK)”, which is intended to suggest improvements to the communication model used by the Brazilian authorities in order to improve reaction to security threats [12], [13]. The system currently in use by the authorities (based on radio transmitters), although reliable, is too limited considering complex operations, as for example, when tracking tactical teams (personnel and vehicles) in real time, with no possibility to access video feeds and GPS coordinates. The primary intention of M.E.N.A.C.E-TRACK is the creation of a dynamic mesh network, intended to interconnect field personnel to a base of operations whenever possible. This type of network accepts the dynamic disconnection and reconnection of nodes. Therefore, it is paramount to research technologies intended to improve the availability of information resources to the authorities (e.g. audio, video, GPS coordinates etc) similarly to [11].

This paper has a different objective considering the original M.E.N.A.C.E-TRACK concept: we propose the creation of a HighLy Portable Mesh nEtwork (or H.L.P-M.E. for short) using off-the-shelf cost effective equipment with minimum downtime for configuration. The idea behind this proposal is to have a number of pre-configured mesh nodes, which can be deployed in the field, in order to provide an *in promptum* mesh network to be used anywhere, anytime. With this infrastructure it would be possible to share multimedia data (e.g. video feeds, GPS coordinates, audio communication etc) in the field without any dependency on landlines or any preexistent infrastructure.

Considering the intended user is not necessarily a computer network specialist, and public safety has a decreasing budget in Brazil [17], [26], [27], the main prerequisites for the intended system are: it must be cost effective and it must be easy to use and deploy. To achieve the proposed objective, we created an experimental environment using off-the-shelf equipment from Open-Mesh, which provided a standard networked environment (i.e. not tampered with in any way) with native support to mesh networks. The steps intended to achieve the proposed objective are: 1) study the Open-Mesh infrastructure, which use B.A.T.M.A.N. routing protocol and 2) explore several mesh network configurations in order to test the flexibility of the devices in establishing meshes. Section 4 discusses the proposed testbed in detail.

4. Experimental Environment

At this time, we chose to use a manufactured B.A.T.M.A.N. access point (AP) instead of using an open source environment (i.e. proprietary hardware + open source firmware), so we can compare this setting to a previous experimental OpenWrt testbed we used with OLSR [12]. Our main objective does not concern the routing protocol used with OpenWrt *per se*, but the difficulties faced when using a completely configurable open source environment. Using OpenWrt we have complete control over the development/production environment, but it is also true that the configuration downtime and the possibilities for unforeseen situations are more prone to happen. Therefore, we chose a manufactured (proprietary hardware + proprietary software + open source firmware) AP which natively supports the B.A.T.M.A.N. protocol: the Open-Mesh OM2P access point (Fig. 1) [29].

Fig. 1: Open-Mesh AP OM2P.

4.1 AP OM2P

Each AP OM2P is enabled to form a mesh infrastructure. That way, it is possible to install units with traditional access (i.e. as Internet gateways) and add other units that can extend the network coverage. This AP has an external 2.4 GHz antenna with 23 dBm (200 mW) with a RP-SMA standard connector. Other aspect is that it can be managed using a cloud service called CloudTrax, which is provided free of charge by Open-Mesh [30]. The AP also has the ability to use passive power over Ethernet (incompatible with 802.3af). The specifications for the device are in Tab. 1 [31].

Table 1: Features of Open-Mesh AP OM2P.

Speed (max.)	150 Mbps
Radio	802.11b/g/n 2.4 GHz
Range (approx.)	75-150' indoor or 600' outdoor
Processor	400 MHz Atheros AR9331 MIPS 24k
Plug and play	yes
Memory	64 MB DRAM
Ethernet (WAN e LAN)	2 x 100 Mbps

4.1.1 CloudTrax Environment

The CloudTrax environment is a free cloud network controller that helps building, managing and monitoring wireless networks from any place in the world. This controller can manage an unlimited number of APs and networks, simply by registering the devices. Even if the devices lose connectivity with the cloud controller, the registered networks aren't affected. This happens because no network traffic passes through the cloud controller [31]. Another advantage is it provides access to network usage statistics graphics (containing number of users, amount of upload and download traffic, the relationship between each of the nodes and details of each node) [32].

To configure a network it is necessary to create a master login at CloudTrax homepage, which allows the administrator to access the configuration of several networks at one place, to create a network, to add any amount of nodes, to install them physically as gateways (connected via Ethernet) or as repeaters. Among the many configurations we can set, we can manually adjust the transmitting power of the antenna (allowing the AP to work indoors), configure cryptography via WPA/WPA2 or use vouchers to regulate user access and protect network traffic, define download and upload limits, restrict access using MAC filtering and, for a more general configuration, to determine if the network will be public or private [32].

4.2 Experimental Data

The acquired APs OM2P were configured initially in very simple scenarios. These APs use B.A.T.M.A.N. advanced (often referenced as batman-adv), which implements the B.A.T.M.A.N. routing protocol in the form of a linux kernel module operating on layer 2. Batman-adv operates entirely on ISO/OSI layer 2, meaning not only the routing information is transported using raw Ethernet frames but also the data traffic is handled by batman-adv. It encapsulates and forwards all traffic until it reaches the destination, hence emulating a virtual network switch of all participating nodes. Therefore all nodes appear to be linked locally and are unaware of the network's topology as well as unaffected by any network changes [33].

Regarding our first experiment, we configured an AP separately as a gateway and in the second one, we configured one AP as gateway and one as a repeater. As expected, there were no difficulties in this first set of experiments. Notebooks were connected to the SSID of AP N01 and the Internet was accessible. Our objective in this first set of experiments was to try the basic functions of this devices and assess the difficulties in using the CloudTrax environment. The environment is practically self explanatory simplifying the described tasks.

After this first stage, we created scenarios that emphasized the mesh topology. For the second stage, we used three nodes (N01, N02 and N03), each of which presenting specific configurations, depending on the created scenario. The first scenario consisted in the configuration of a mesh with one gateway and two repeaters (Fig. 2).

In Fig. 2, we can verify that N01 is configured as a gateway (N01(g)) and the other nodes are configured as repeaters. This configuration demonstrates a first example of increased network coverage. The CloudTrax controller offers meaningful visual data as shown in Fig. 5. We highlight the hop count each repeater AP performs to the gateway (last column). We only presented here the APs tab of the generated graphics, given the data provided by the other tabs are not useful for the mesh evaluation (except for the network diagram tab – as shown in Fig. 2). The "network map" shows the AP and its current configurations in a Google Map like environment, "all networks map" offers a Google Map like environment with all the CloudTrax managed networks, "clients" show client statistics and "site survey" shows information on neighbouring network APs (e.g. signal strength, channel, SSID, current mode – b/g/n etc).

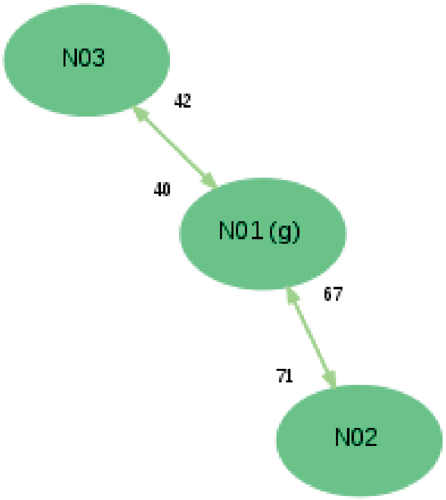


Fig. 2: Network diagram generated in the CloudTrax environment.

Given the natural mesh auto-configuration feature, it is possible to obtain different paths with the same infrastructure. Fig. 3 demonstrates a new organization of the same three nodes. Comparing Fig. 2 and Fig. 3, we can see that in this new organization, the devices connected to the node N02 can now communicate with devices connected to N03 without passing through N01, only because we added a new path between N02 and N03.

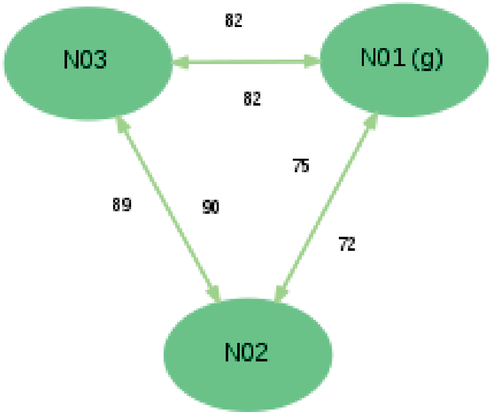


Fig. 3: Network diagram after adding a new path between N02 and N03.

In the next experiment we tinkered again with the paths of the mesh testbed and configured one gateway and two repeaters, but now, connecting the gateway to one repeater and this repeater, to another AP also configured as a repeater (Fig. 4).

The network diagram presented in Fig. 4 demonstrates that the APs have the ability to communicate through multiple hops. Observing the network data presented in Fig. 6, it is clear that node N02 is two hops away from the N01 gateway.

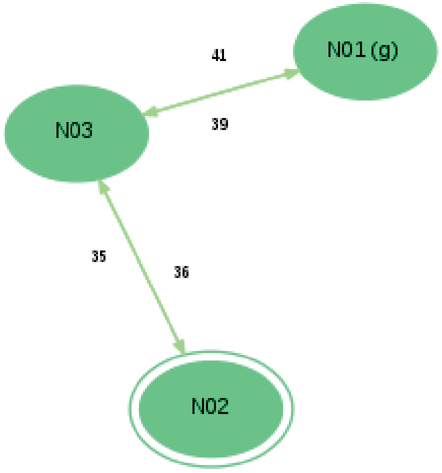
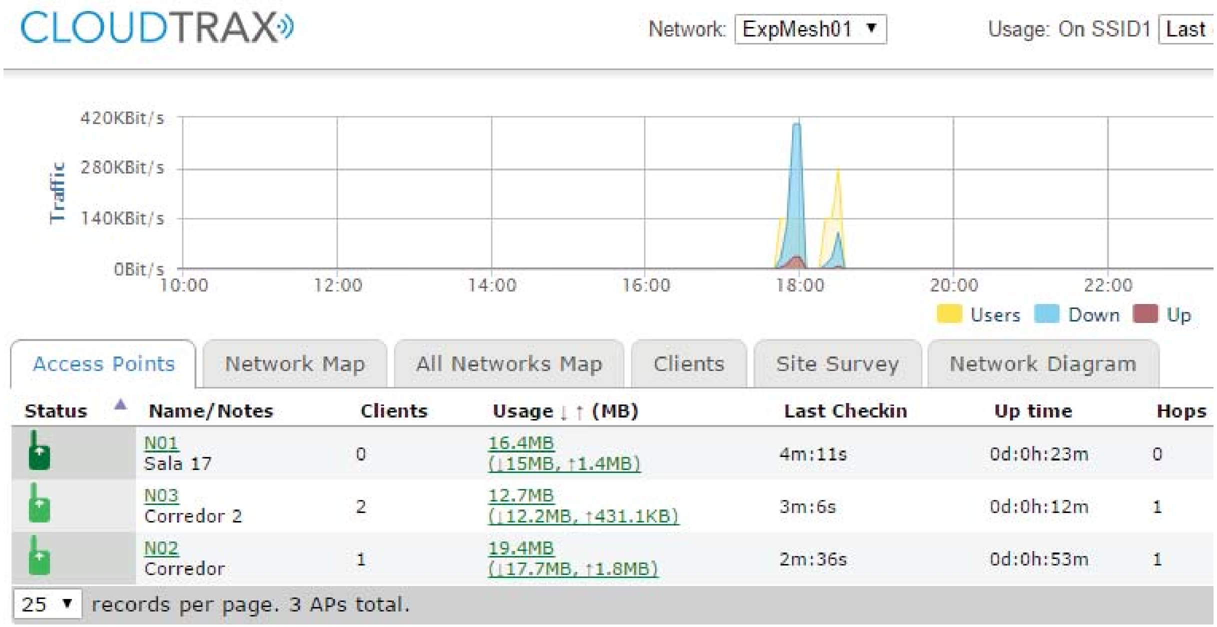


Fig. 4: Network diagram for the new topology of N01, N02 and N03.

One last thought regarding the presented topologies is that all the links established are bidirectional (i.e. full duplex). All the experimental setups presented were tested connecting devices to each SSID and using the ping tool to verify their connectivity (simultaneously) and verifying mainly if the 1 hop and 2 hop distance did not interfere in the reachability of each device. Besides, we also made another simple test: we disconnected the gateway (i.e. N01) from the Ethernet network, therefore rendering it unreachable from/to the Internet (and therefore, unavailable to CloudTrax). Given we disabled the feature "access point isolation" (which prevents wireless users from accessing each other's computers) in the advanced tab, as the infrastructure was already configured in CloudTrax, it keeps its configured characteristics. Therefore, we still can access the SSID of the mesh network and we can still reach every single device that is using the network locally. Considering this APs are extremely portable, by adding a battery module (like a portable powerbank) in each node, we have an almost zero configuration mesh network environment that is ready for use in any environment (indoor or outdoor), as we intended for this paper.

5. Conclusion

The main objective of this paper was to present the basis for the creation of a HighLy Portable Mesh nEtwork (or H.L.P-M.E.) using an off-the-shelf device, which implements B.A.T.M.A.N. layer 2. Given our experience with OpenWrt, we know it is possible to achieve a similar environment using only open source software (i.e. hardware + open source firmware) but when comparing to the functionalities available in the Open-Mesh OM2P and in the CloudTrax network management tool, we raise questions

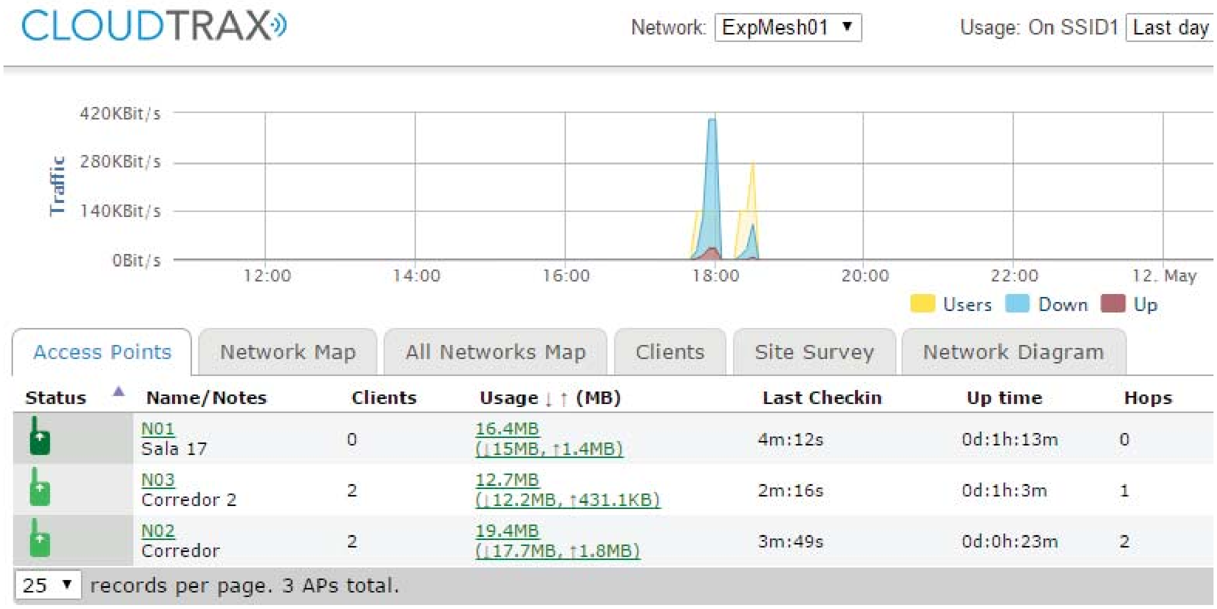


Network: ExpMesh01 ▼ Usage: On SSID1 Last

Status	Name/Notes	Clients	Usage ↓ ↑ (MB)	Last Checkin	Up time	Hops
	N01 Sala 17	0	16.4MB (↓15MB, ↑1.4MB)	4m:11s	0d:0h:23m	0
	N03 Corredor 2	2	12.7MB (↓12.2MB, ↑431.1KB)	3m:6s	0d:0h:12m	1
	N02 Corredor	1	19.4MB (↓17.7MB, ↑1.8MB)	2m:36s	0d:0h:53m	1

25 ▼ records per page. 3 APs total.

Fig. 5: Network test data CloudTrax graphic.



Network: ExpMesh01 ▼ Usage: On SSID1 Last day

Status	Name/Notes	Clients	Usage ↓ ↑ (MB)	Last Checkin	Up time	Hops
	N01 Sala 17	0	16.4MB (↓15MB, ↑1.4MB)	4m:12s	0d:1h:13m	0
	N03 Corredor 2	2	12.7MB (↓12.2MB, ↑431.1KB)	2m:16s	0d:1h:3m	1
	N02 Corredor	2	19.4MB (↓17.7MB, ↑1.8MB)	3m:49s	0d:0h:23m	2

25 ▼ records per page. 3 APs total.

Fig. 6: Network data for the new topology of N01, N02 and N03.

regarding the development time and the amount of training we would need to put the intended audience through (i.e. law enforcement agents) to use effectively the system. Using OM2P + CloudTrax, the creation of the mesh topologies is almost effortless and we see almost now downtime considering the learning curve to use this infrastructure. Using minor adaptations (i.e. adding a portable battery module) the configured mesh topology is available on the go to enable a perimeter network anytime/anywhere as we wanted to demonstrate. Our next experiments will involve field testing with the battery modules and outdoor testing regarding the transmission of multimedia data in real life situations (e.g. as in the fast deployment of the infrastructure in a sting operation).

Acknowledgment

The authors would like to thank DECIT/SCTIE/MS/CNPq/FAPEG for the full sponsorship of this research (edict numbers 006/2012 and 12/2013).

References

[1] R. T. do Valle e D. C. Muchaluat-Saade, "MeshAdmin: An integrated platform for wireless mesh network management", in *Proc. Network Operations and Management Symposium (NOMS)*, 2012, pp. 293–301.

[2] Mi. Kim, I. Ra, J. Yoo, D. Kim, and H. Kim, "QoS Mesh Routing Protocol for IEEE 802.16 based Wireless Mesh Networks", in *Proc. 10th International Conference on Advanced Communication Technology ICACT*, 2008, pp. 812–817.

[3] G. A. Cabral, and G. R. Mateus, "Simulation-Based Optimization for Wireless Mesh Network Planning", in *Proc. 2010 Third International Conference on Advances in Mesh Networks (MESH)*, 2010, pp. 28–34.

[4] *IEEE Standard for Information technology – Telecommunications and information exchange between systems Local and metropolitan area networks – Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Std. 802.11, 2012.

[5] M. M. Farias. "Routing protocol for wireless mesh networks [Protocolo de roteamento para redes wireless mesh]," M. Comp. Science thesis, Informatics College PUCRS [Faculdade de Informática PUCRS], Porto Alegre, Brazil, 2008.

[6] T. M. Cardoso, and P. C. F. Marques, "Mesh network: topology and application [Rede Mesh: topologia e aplicação]", *iTEC Magazine [Revista iTEC]*, vol. IV, n. 4, pp. 16–25, Jul. 2012.

[7] (2015) Microsoft Self Organizing Wireless Mesh Networks, website. [Online]. Available: http://research.microsoft.com/en-us/projects/mesh/

[8] J. Bicket, D. Aguayo, S. Biswas and R. Morris, "Architecture and Evaluation of an Unplanned 802.11b Mesh Network," in *Proc. 11th annual international conference on Mobile computing and networking (MobiCom'05)*, 2005, pp. 31–42.

[9] S. A. Mahmud, Shahbaz Khan, Shoaib Khan and H. Al-Raweshidy, "A comparison of MANETs and WMNs: commercial feasiblity of community wireless networks and MANETs", in *Proc. 1st international conference on Access networks (AcessNets'06)*, 2006, paper 18.

[10] S. Barakovi? and J. Barakovi?, "Comparative performance evaluation of Mobile Ad Hoc routing protocols," in *Proc. 33rd International Convention, MIPRO*, 2010, pp. 518–523.

[11] D. PADI, "Vehicular Information & Communications Technology (VICT) System," in *Proc. 2nd International Conference on Adaptive Science & Technology*, 2009, pp. 390–394.

[12] D. M. Tavares, M. J. Lima, R. V. Aroca, G. A. P. Caurin, A. C. De Oliveira Jr, T. A. Santos Filho, S. J. Bachega, M. A. Batista and S. F. Da Silva, "Access Point Reconfiguration Using OpenWrt," in *Proc. The 2014 International Conference on Wireless Networks (ICWN'14)*, 2014, pp. 254–260.

[13] D. M. Tavares, A. P. Da Silva, S. J. Bachega, R. V. Aroca, J. Ueyama, G. A. P. Caurin and A. C. De Oliveira Jr., "A Practical Evaluation of Smartphone Application on Mesh Networks," in *Proc. The 2014 International Conference on Wireless Networks (ICWN'14)*, 2014, pp. 247–253.

[14] S. J. Bachega and D. M. Tavares, "Simulation of Reactive Routing Protocols in Wireless Mesh Networks: a Systematic Literature Review", in *Proc. The 2014 International Conference on Wireless Networks (ICWN'14)*, 2014, pp. 235–239.

[15] C. L. Chan, S. C. Lee, K. C. Yeong and V. Jeewa, "Innovations to improve wireless mesh network performance: A survey," in *Proc. IEEE Symposium on Wireless Technology and Applications (ISWTA)*, 2013, pp. 80–84.

[16] I. F. Akyildiz and X. Wang, "Innovations to improve wireless mesh network performance: A survey," *IEEE Communications Magazine*, pp. S23–S30, 2005.

[17] (2015) Journal of Brazil [Jornal do Brasil], RJ: cuts of R$ 2,6 billion in the annual budget concerns public safety [RJ: corte de R$ 2,6 bilhões no orçamento anual preocupa segurança pública]. [Online]. Available: http://www.jb.com.br/rio/noticias/2015/01/27/rj-corte-de-r-26-bilhoes-no-orcamento-anual-preocupa-seguranca-publica/

[18] (2015) Mesh Dynamics, Mobile mesh networks for military, defense and public safety. [Online]. Available: http://www.meshdynamics.com/military-mesh-networks.html

[19] P. Garnepudi, T. Damarla, J. Gaddipati and D. Veeraiah, "Proactive, reactive and hybrid multicast routing protocols for Wireless Mesh Networks," *Proc. IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2013, pp. 1–7.

[20] (2015) Open-mesh, B.A.T.M.A.N. Protocol concept. [Online]. Available: http://www.open-mesh.org/projects/open-mesh/wiki/BATMANConcept

[21] F. Zeiger, N. Kraemer and K. Schilling, "Commanding mobile robots via wireless ad-hoc networks – A comparison of four ad-hoc routing protocol implementations," *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 2008, pp. 590–595.

[22] D. Johnson, N. Ntlatlapa and C. Aichele, "A Simple pragmatic approach to mesh routing using BATMAN," in *Proc. 2nd IFIP International Symposium on Wireless Communications and Information Technology in Developing Countries*, 2008.

[23] D. Murray, M. Dixon and T. Koziniec, "An experimental comparison of routing protocols in multi hop ad hoc networks," *Proc. Telecommunication Networks and Applications Conference (ATNAC)*, 2010, pp. 159–164.

[24] R. Sanchez-Iborra and Maria-Dolores Cano, "Qoe-based performance evaluation of video transmission using the BATMAN routing protocol," *Proc. 10th ACM symposium on QoS and security for wireless and mobile networks (Q2SWinet'14)*, 2014, pp. 9–16.

[25] A. Neumann, C. Aichele and S. Wunderlich. (2015) Better Approach To Mobile Ad-hoc Networking (B.A.T.M.A.N.) draft-wunderlich-openmesh-routing-00. [Online]. Available: http://tools.ietf.org/html/draft-wunderlich-openmesh-manet-routing-00

[26] (2015) Sul 21, Civil Police on Strike Against Budget Cuts in RS Public Safety [Polícia Civil paralisa atividades contra cortes na segurança pública do RS]. [Online]. Available: http://www.sul21.com.br/jornal/policia-civil-paralisa-atividades-contra-cortes-na-seguranca-publica-do-rs/

[27] (2015) GoiásReal, Budget cuts in public safety instills violence high tide [Com cortes na Segurança Pública, violência segue em alta]. [Online]. Available: http://www.goiasreal.com.br/noticia/15/com-cortes-na-seguranca-publica-violencia-segue-em-alta

[28] J. Xu, L. Wang, Y. Li, Z. Qin and M. Zhu, "An Experimental Study of BATMAN Performance in a Campus Deployment of Wireless Mesh Networks," *Proc. Seventh International Conference on Mobile Ad-hoc Sensor Networks (MSN)*, 2011, pp. 341–342.

[29] (2015) Open-Mesh, OM2P 150 Mbps Access Point with External Antenna. [Online]. Available: http://www.open-mesh.com/products/access-points/om2p.html

[30] (2015) CloudTrax, Part 1: CloudTrax Guide Overview. [Online]. Available: https://help.cloudtrax.com/hc/en-us/articles/202465650-Part-1-CloudTrax-Guide-Overview

[31] (2015) Open Mesh, OM2P Access Point with External Antenna. [Online]. Available: https://www.open-mesh.com/skin/frontend/default/open-mesh/images/OM-2015-04.pdf

[32] (2015) CloudTrax, Creating your first CloudTrax network. [Online]. Available: http://cloudtrax-static.s3.amazonaws.com/docs/quick_start_guide.pdf

[33] (2015) batman-adv, Doc-overview B.A.T.M.A.N. advanced. [Online]. Available: http://www.open-mesh.org/projects/batman-adv/wiki/Wiki

IP Network Mobility using OSPF Area

Sunghyun Yoon, Ho-Yong Ryu

Smart Network Research Department, Electronics and Telecommunications Research Institute, Daejeon, Korea
shy72@etri.re.kr, hyryu@etri.re.kr

Abstract - *Network mobility is a capability that all devices within the mobile network can communicate with external peers without any restrictions according to moving network. The NEMO is a current leading technology for IP network mobility. Since, however, the NEMO have been defined and implemented by extending the mobile IP (MIP), the mobile router must implement the MIP protocol in addition to existing routing protocol. As a result, of course, the IP core network also needs to deploy MIP elements including home agent and foreign agent. In this paper, we propose an IP network mobility scheme using mobile router equipped with extended OSPF routing protocol. Since our scheme can provide network mobility only by extending the OSPF protocol, mobile router and the network do not need to implement the MIP protocol and network element system for network mobility. Furthermore, the OSPF routing protocol is commonly used in many routers. Therefore, it is possible to implement cost-effective network mobility through minimizing changes in the network.*

Keywords: network mobility, mobile router, OSPF Area

1. Introduction

In general, network includes several kinds of network element systems for data transmission such as router, switch, repeater, etc. Especially, the router transmits packets through the Internet Protocol (IP) layer and routing protocol is used for its operations.

The routing protocol is to exchange routing information with each other. There are many kinds of routing protocol according to the purpose such as coverage (interior gateway protocol, exterior gateway protocol), behaviour (dynamic, static), implementation of the algorithm (distance vector, link state). Therefore, many routing protocols have been developed and used depending on the different types of routing (e.g. Routing Information Protocol (RIP), Internet Gateway Routing Protocol (IGRP), Open Shortest Path First (OSPF), Enhanced Internet Gateway Routing Protocol (EIGRP), Intermediate System to Intermediate System routing protocol (ISIS), and Border Gateway Protocol (BGP), etc.). From among these, the OSPF is the current most commonly used routing protocol for IP networks [1-2]. It is an interior gateway protocol that routes IP packets solely within a single routing domain (autonomous system). It gathers link state information from available routers and constructs a topology map of the network.

On the other hand, the growing use of IP devices in portable applications has created the demand for mobility support for entire networks of IP devices. The network mobility (NEMO) solves this problem by extending Mobile IP. The IP network generally uses mobile router for supporting network mobility. A representative technique that supports network mobility using the mobile router is the Network Mobility (NEMO) proposed by Internet Engineering Task Force (IETF) [3-5].

The NEMO requires legacy routing protocol as well as mobile IP. The mobile IP provides mobility to mobile router by expanding legacy IP. As the mobile IP is used, however, the NEMO additionally needs to include additional network element systems for mobile IP such as home agent (HA) and foreign agent (FA). As such, if the mobile IP used in the mobile router, the mobile network systems are additionally needed (e.g. HA and FA).

In this paper, we propose a network mobility using OSPF area concept, which is adaptable conventional network.

2. Related Works

2.1 OSPF routing protocol

The OSPF is a link-state routing protocol that employs a version of Dijkstra's shortest path first protocol, and is an open standard. It allows collections of contiguous networks and hosts to be grouped together and labels such a group, together with the routers having interfaces to any one of the included networks, an area. OSPF areas are interconnected via a backbone area, which is called area zero. OSPF requires that IP datagram exchanges between areas must traverse the OSPF backbone area.

Routing bandwidth utilization should be minimized. However, the ability to detect, advertise, and route around network outages should be timely. In addition, the grouping of networks within OSPFs hierarchical architecture should be carefully designed to ensure network reconfigurations can be performed with ease [6].

OSPF supports a number of messages that allow routing information to be exchanged within and between areas. OSPFs area architecture limits the required number and size of routing messages through IP route aggregation at area boundaries. The OSPF IP routing database is maintained through the exchange of Link State Update (LSU) messages. LSU messages encapsulate a variety of OSPF messages called Link State Advertisements (LSAs).

The OSPF uses multicast addressing for route flooding on a broadcast domain. For non-broadcast networks, special provisions for configuration facilitate neighbour discovery. The OSPF multicast IP packets never traverse IP routers (never traverse Broadcast Domains), they never travel more than one hop. The OSPF is therefore a Link Layer protocol in the Internet Protocol Suite.

2.2 Network Mobility

NEMO is a novel thought for handling a bunch of nodes within a moving vehicular area. Namely, this protocol upholds continuous internet connectivity to nodes by establishing a bi-directional tunnel between Mobile Router (MR) and HA, when the MR of a mobile network changes its point of attachment. The bi-directional tunnel is set up as soon as the mobile router sends a successful Binding Update (BU) to its HA in order to inform the home agent about its current point of attachment. All traffic flow between the nodes in the mobile network and correspondent node must pass through the HA. This leads to sub-optimal routing that can surely disrupt and deteriorate all communications to and from the Mobile Network Nodes (MNN). Even the overheads can be further amplified if mobile networks are nested which is unacceptable for real-time applications that require certain Quality of Service (QoS) restrictions [7].

Routing is one of the key challenges that arises in compound internetworks: indeed, while specific routing protocols are typically used for wired networks on one hand, and for wireless mesh networks on the other hand, it has been observed that operating a single routing protocol to manage a compound internetwork as a whole brings several advantages. In this realm, the Internet Engineering Task Force (IETF) has thus standardized protocol extensions to OSPF, enabling OSPF to operate simultaneously on wired networks, and on wireless mesh or moderately mobile ad hoc networks (MANETs) [8].

Congestion Aware Selection of Path with Efficient Routing (CASPER) [9] aims at providing a solution to the problem of network congestion that arises when huge amount of data such as multimedia data is transferred in mobile ad hoc networks. This issue has been addressed by designing a protocol that performs routing intelligently and minimizes the delay in data transmission. The objective of CASPER is to move the traffic away from the shortest path that is obtained by a suitable shortest path calculation algorithm to a less congested path so as to minimize the number of packet drops during data transmission and to avoid unnecessary delay. Here, a router runs the shortest path algorithm after pruning those links that violate a given set of constraints. The proposed protocol has been compared with two link state routing protocols namely, OSPF and Optimized Link State Routing Protocol (OLSR) [10].

3. Network Mobility Support using OSPF

3.1 Mobile Router

Fig. 1 shows an example of network mobility service with MR. As shown in Fig.1, an architecture for network mobility service with mobile routers includes AAA (Authentication, Authorization, and Accounting) server, a number of access routers (AR), some MRs, and terminals. The ARs are connected to an IP network, and the MR is conventionally connected to the IP network through ARs.

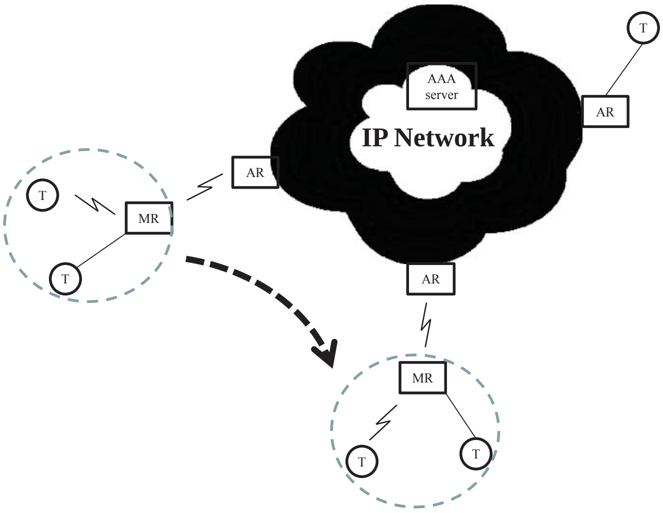


Figure 1. Network mobility service with mobile router

The MR constitutes at least one network and also a mobile network because its network itself is mobile. Additionally, the terminals are connected to the MR through various forms such as wire and wireless and the MR may include various interfaces to support communication with the terminals connected through various forms.

The MR performs an additional authentication procedure in order to connect with an IP network system. That is, the IP network system recognizes the MR as one terminal until the MR is recognized as a router. Thus, the IP network system performs an authentication procedure for the first network connection.

The MR performs an authentication procedure through the AAA server and the authenticating of the MR is regarded as a network mobility service such that the lower level terminals can receive service through the MR.

Additionally, the terminals to be connected to the IP network system accomplish the authentication procedure through the

AAA server. The authentication procedure includes terminal authentication and network authentication. During the authentication procedure, the AAA server allocates an IP address for IP communication to a terminal that requires connection. Also, the AAA server allocates authentication information in preparation for a case that re-authentication occurs later on and may recognize types of a terminal according to a level of an authentication procedure.

The MR that performs the authentication procedure is connected to the IP network through the first AR. According to the current location of the MR, the service connection point may shifts from the previous AR to the new AR. Then the terminals in the MR recognize the MR as a default router, i.e., a gateway, and maintain communication.

The terminals move together as the MR moves, but each of the terminals maintains communication without recognition about mobility. Moreover, among routing protocols used in the IP network, the OSPF routing protocol is a general-purpose routing protocol that does not belong to a specific vendor. The OSPF routing protocol is accepted as being more efficient compared to a conventional RIP (Routing Information Protocol). In this paper, mobility is provided to the MR by expanding the OSPF routing protocol.

3.2 OSPF area

Fig. 2 shows an example of routing information transfer for network mobility support using OSPF. As shown in Fig. 2, the network system to which the OSPF routing protocol is applied includes several core routers (CR), ARs and MRs.

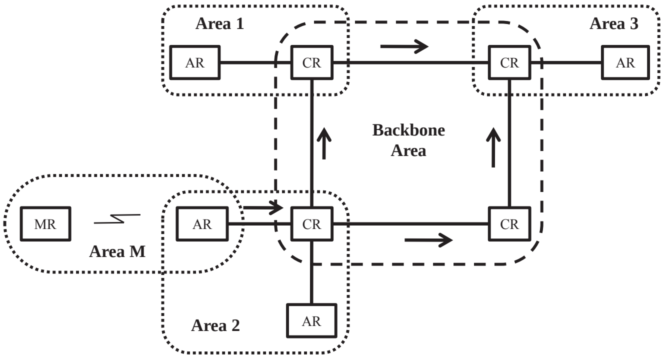


Figure 2. Routing information transfer

The MR may be directly connected to CRs but for convenience of description, it is assumed that the MR is connected to the network system through one of the ARs. In fact, this is a very dangerous case in real network world and rarely occurs due to extreme situation.

The OSPF routing protocol is an IGP (Interior Gateway Protocol) that exchanges routing information between routers in an AS (Autonomous System) classified as one routing management area. The OSPF routing protocol divides the AS into some areas (i.e., OSPF areas) and performs layer 2 routing that connect the areas through a backbone network, i.e., a backbone area.

Each area has an original area identifier (ID) which is transmitted through a header of an OSPF packet. Especially, since the backbone area has a special area ID, it is distinguished from other areas.

The CRs are responsible for the centre of the network system, and an area where the CRs are included is called as a backbone area. The area 1, 2, and 3 include the ARs together with the CRs. The area M is an area including the AR and the MR based on the connection of the MR.

The MR generates an area ID about the area M where the MR and the AR are included once MR is connected to the AR. The MR generates an area ID by using its own ID. The MR may generate an area ID through various methods.

For example, it is assumed that an area ID has the length of 32 bits. If the MR has a 32 bit ID, it may use its own ID as an area ID. If the MR has an ID of more than 32 bits, it selects only the 32 bits from the entire ID (based on the MSB (Most Significant Bit) or the LSB (Least Significant Bit)) and may use it as an area ID. If the MR has an ID of less than 32 bits, the MR adds predetermined bits to have the length of 32 bits and uses it as an area ID.

The MR uses a new area ID to support the OSPF routing protocol and is classified as one area in the OSPF routing protocol. The MR is allocated with an area ID from the AAA server when an authentication procedure is performed using the AAA server.

The MR uses an area ID to identify the area M of the MR in the network system. The routing information including the area ID is transmitted to the CR of the backbone area through the AR.

However, since the area M is not directly connected to the backbone area like the area 1, 2, and 3, routing information is transmitted to the backbone area through the virtual link. Likewise, since the routing information is delivered through a virtual link, the routing table of each router may be updated rapidly. The routing information is transmitted through a virtual link passing through the MR, the new AR, and CR sequentially. This virtual link is created from the MR to the CR through the AR.

The AR transmits the routing information including an area ID of the MR and the CR transmits the routing information to other CRs in the backbone area. Also, the CR receives the routing information including an area ID of the MR from other CRs. That is, the CR propagates the routing information including an area ID of the MR to other CRs in the backbone area.

The OSPF delivers routing information according to a change of a link state. As the MR moves, the AR to which the MR is connected is changed and routing information is transmitted into a corresponding AR.

If ARs and CRs are disconnected from the MR when the MR moves, the routing information of the corresponding MR is removed.

The MR adds information about effective time to the routing information including an area ID for the AR and CR receiving corresponding routing information to delete the routing information if the effective time of corresponding routing information is expired.

3.3 Example of network mobility

Fig. 3 shows an example for flow of a network mobility support. As shown in Fig. 3, an MR begins an authentication procedure in order to connect with the network system. The MR requests authentication to a network node, i.e., a first AR, according to an initial connection in operation.

The first AR requests authentication of the MR to AAA server. The AAA server confirms whether authentication about the MR is appropriate or not, and if not, the AAA server approves the authentication and allocates authentication information necessary for instant re-authentication later on to transmit it to the first AR. The authentication information may be issued in a form of an authentication key and may be issued to the MR and ARs (e.g., ARs). Therefore, if re-authentication is required later on during service, the authentication is possible without intervention of the AAA server.

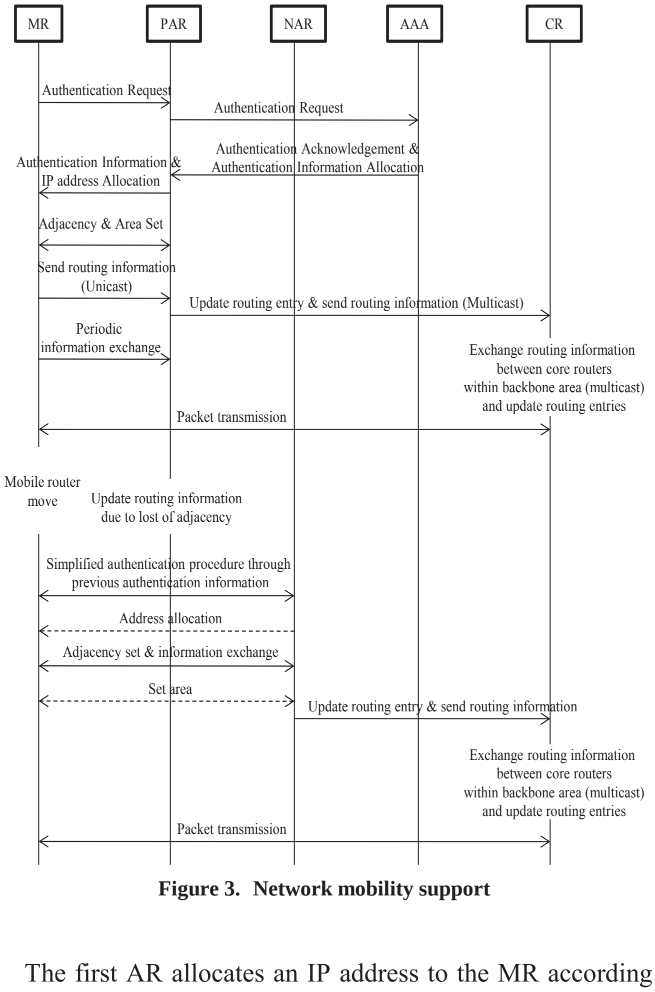

MR
PAR
NAR
AAA
CR
Authentication Request
Authentication Request
Authentication Acknowledgement &
Authentication Information Allocation
Authentication Information &
IP address Allocation
Adjacency & Area Set
Send routing information
(Unicast)
Update routing entry & send routing information (Multicast)
Periodic
information exchange
Exchange routing information
between core routers
within backbone area (multicast)
and update routing entries
Packet transmission
Mobile router
move
Update routing information
due to lost of adjacency
Simplified authentication procedure through
previous authentication information
Address allocation
Adjacency set & information exchange
Set area
Update routing entry & send routing information
Exchange routing information
between core routers
within backbone area (multicast)
and update routing entries
Packet transmission

Figure 3. Network mobility support

The first AR allocates an IP address to the MR according to a network connection authentication of the MR.

When the MR receives the IP address and authentication information, it creates an adjacency with the first AR and sets an area in operation. The MR creates an adjacency with the first AR by transmission a "Hello" message and sets up an area according to an adjacency formation. At this point, the MR generates an area ID using its own information, for example, an original identifier. Also, the MR may be allocated with an area ID through the above authentication procedure.

The MR transmits the routing information including an area ID to the first AR. The area that the MR and the first AR create is not directly connected to a backbone area and thus is connected through a virtual link. Accordingly, the routing information is transmitted to the first AR through a unicast form. The MR transmits the routing information to the first AR through LSD (a Link State Update) message.

Once the first AR receives the routing information, it modifies its routing table and transmits the routing information to the connected CR simultaneously. At this point, the routing information is transmitted through a multicast form.

The CR simultaneously modifies its routing table and transmits the routing information received from other CRs in the backbone area in order to propagate a routing table to other CRs in the backbone area.

The MR transmits information, for example, a "Hello" message, periodically in order to maintain a connection state in operation.

Terminals connected to the MR communicate with other terminals through the MR and the MR transmits a packet between the communicating terminals in operation.

Next, once the MR moves in operation, it is disconnected from the first AR that is currently connected thereto. The first AR detects that the MR moves or communication errors occur if the "Hello" message is not periodically received from the MR. Then, the first AR removes the corresponding routing information and the adjacency.

The MR can perform a simplified authentication procedure using the previously received authentication information in operation. The MR can perform a prompt re-authentication with the new AR.

The MR acquires a new IP address through the new AR in operation. At this point, if the new AR uses the same network prefix as the first AR, operation can be omitted.

Once the MR sets up the adjacency with the new AR, it generates an area ID to set up an area in operation. If the new AR has the same area ID as the previous AR, operation can be omitted.

When the new AR receives routing information, it transmits the routing information to the CR in operation. The MR transmits the "Hello" message periodically to the new AR in operation.

Terminals connected to the MR communicate with other terminals through the MR and the MR transmits a packet between communicating terminals in operation.

The ARs or CRs, which receive the routing information including an area ID of the MR, update their routing tables using the routing information such that it is possible to identify the MR.

4. Conclusions

Generally, an IP-based network mobility is most appropriate technology for providing IP mobility in the vehicles such as cars, trains, buses, planes, ferries and etc. However, the mobile IP technology should be implemented to provide network mobility in conventional technology using mobile router, which is most common technology is IETF NEMO. This means that NEMO has to overcome the limitations of the mobile IP in order to provide network mobility.

The proposed network mobility support regards an MR as an AR (to which the MR is connected) and OSPF area. For this, the MR uses an area ID that identifies an area where the MR itself is included. Therefore, this paper uses an OSPF area concept in a routing domain where OSPF routing protocol is possible, thereby proving mobility to a network system.

Furthermore, the proposed network mobility in this paper extends the OSPF routing protocol in MR to support mobility. As a result, a network change can be minimized. This means the network mobility can be provided with low cost. And, the conventional network element system does not require an additional mobile IP routing protocol and additional network components in behalf of MR. Also the proposed network mobility can provide easy virtual private network configuration due to simple architecture of the OSPF.

Acknowledgment

This research was supported by the IT R&D program of MSIP (Ministry of Science, ICT and Future Planning) / IITP (Institute for Information & communications Technology Promotion). 12221-14-1001, Next Generation Network Computing Platform Testbed.

5. References

[1] J. Moy, “OSPF Version 2,” RFC2328, 1998.

[2] R. Coltun, D. Ferguson, J. Moy, A. Lindem, “OSPF for IPv6,” RFC5340, 2008.

[3] V. Devarapalli et al., “Network Mobility (NEMO) Basic Support Protocol,” RFC3963, 2005.

[4] Paul Moceri, “Enabling Network Mobility: A Survey of NEMO,” http://www.cs.wustl.edu/~jain/cse574-06/ftp/network_mobility/, April 2006.

[5] Eranga Perera, Vijay Sivaraman, Aruna Seneviratne, “Survey on network mobility support,” ACM SIGMOBILE Mobile Computing and Communications Review, Volume 8 Issue 2, April 2004.

[6] Wollman, W.V., Barsoum, Y., "Overview of open shortest path first, version 2 (OSPF V2) routing in the tactical environment", Proceeding of IEEE Military Communications Conference (MILCOM), Vol.3, 925p~930p, 1995.

[7] Loay F. Hussein, Aisha-Hassan A. Hashim, Mohamed Hadi Habaebi, Wan Haslinah Hassan, "A Scheme (Diff NEMO) for Enhancing QoS in Network Mobility", Journal of Applied Sciences Vol. 15 No. 3 474p ~ 482p, 2015.

[8] Cordero Fuertes, J.A., Philipp, M., Baccelli, E., "Routing across wired and wireless mesh networks: Experimental compound internetworking with OSPF", Proceeding of International Wireless Communications and Mobile Computing Conference (IWCMC), 739p ~ 745p, 2012.

[9] Dhurandher, S.K., Obaidat, M.S., Diwakar, K., "A mechanism for reducing congestion while routing bulky data in Mobile Ad Hoc Networks", Proceeding of IEEE International Conference on Electronics, Circuits, and Systems (ICECS), 142p~145p, 2010.

[10] T. Clausen, P. Jacquet, Project Hipercom, INRIA, “Optimized Link State Routing Protocol (OLSR),” RFC3626, 2003.

Hand Gesture Interface using Light-dark changes in an Illuminance meter built in mobile devices

Kento Matsui[1], Hiroto Aida[2], Hikaru Ichikawa[1], Hiroki Murakami[1], and Mitsunori Miki[2]
[1]Graduate School of Science and Engineering, Doshisha University, Kyoto, Japan
[2]Department of Science and Engineering, Doshisha University, Kyoto, Japan

Abstract— *This study examines a hand gesture interface using light-dark change in illuminance meters built in smartphones and tablets. A hand gesture is to be detected from a change in illuminance values obtained by an illuminance meter and classified through decision tree learning. This realizes intuitive operations by hand gestures on a mobile device. While many of studies on hand gesture recognition use camera images, it result in large computational complexity required for gesture recognition due to image analysis. Privacy consideration is also required. By using light-dark changes in an illuminance meter, hand gestures can be recognized with less computational complexity and lighter applications can be realized. It also does not require privacy consideration. The result of the precision verification experiment showed that 5 types of hand gestures were recognized at the average recognition rate of 95%. Generic applicability to unknown users and unknown light environments was also confirmed.*

Keywords: hand gesture, natural user interface, illuminance meter

1. Introduction

NUI (natural user interface), which allows intuitive operation, has drawn attention in recent years. NUI is a user interface that allows operation by an intuitive action and realizes interaction with computer using an action such as touch or voice operation. Extensive research has been made on gesture interface above all, among topics on NUI, and many researchers have proposed various methods for gesture recognition [1][2][3]. There has also been a progress in development and commercialization of a device capable of gesture recognition, represented by Kinect [4] and Leap Motion [5]. Technology of gesture recognition has been positively utilized in a variety of areas including entertainment and medicine [6][7][8].

On the other hand, as mobile devices have rapidly spread in recent years, there has been increasing research using information from sensors of such devices. A diverse variety of sensors, including proximity sensor, accelerometer, and geomagnetic sensor, are embedded in smartphones and tablets. These sensors are used for various research and services including user behavior estimation and indoor location estimation.

There are studies on gesture interface also among research using mobile devices. Above all, there is extensive research on hand gesture interface for the recognition of hand movement. Various recognition methods have been proposed including a method for recognizing desk rubbing gestures by using accelerometers and microphones [10], a method for recognizing hand gestures using an RGB camera of a mobile device [11], etc. These methods, however, have issues such as requiring additional devices other than a mobile device or resulting in large computational complexity because of the use of camera images for gesture recognition.

This paper proposes hand gesture interface using light-dark changes in an illuminance meter built in a mobile device (hereinafter referred to as "built-in illuminance meter") or "HGI/LI" in short, and examines HGI/LI’s precision in hand gesture recognition. It also evaluates usability of HGI/LI by conducting an experiment with human subjects.

2. Hand gesture interface using mobile devices

Active researches have been conducted on hand gesture interface using mobile devices in recent years. SideSwipe [9] detects and recognizes hand gestures using GSM signal by a circuit board with four antennas attached to the back of a smartphone. This realizes recognition of hand gestures not only over the device but also those around it. SurfaceLink [10] recognizes gestures using an accelerometer, a vibration motor, a speaker, and a microphone built in smartphones. This realizes sharing and exchanging information among multiple devices on the same surface.

In these studies, however, additional devices other than mobile devices are required in order to recognize hand gestures. Therefore, if additional devices are expensive, the introduction cost of the system is high. Even if additional devices are inexpensive, a method using mobile devices only is preferred in light of advantages of using mobile devices which have become generic products, although it depends on the types of gestures recognized, the precision of recognition, and applications realized by a given method. HGI/LI proposed in this paper uses a single mobile device only in order to recognize hand gestures.

Song et al. [11] expanded the interaction with a mobile device by using an RGB camera built in a smartphone to recognize gestures and combining this with touch operation. Robust gesture recognition is realized by using an algorithm based on the random forest. Recognizing hand gestures using camera images, however, requires large computational complexity in analyzing images. The battery duration of a mobile device needs to be considered. Privacy consideration is also required in using camera images.

HGI/Li uses a built-in illuminance meter to recognize hand gestures with small computational complexity. This can realize lighter applications. Unlike a method using camera images, it does not require privacy consideration.

3. Hand gesture interface using light-dark changes in an illuminance meter built in mobile devices

3.1 Concept

A diverse variety of sensors, including proximity sensor and accelerometer, are installed on smartphones and tablets. An illuminance meter is also built in a mobile device for the purpose of adjusting the brightness of its display. This study proposes HGI/LI, which realizes a hand gesture interface using light-dark change of the built-in illuminance meter. HGI/LI realizes intuitive operation by hand gestures on a mobile device by obtaining illuminance information from an illuminance meter, extracting features from changes in illuminance values, and classifying gestures into five types using a classification model obtained through decision tree learning.

3.2 Type of hand gesture

HGI/LI recognizes five types of hand gestures: hide, roll, up, down, and slash. Fig. 1 - 5 is a picture of these hand gestures.

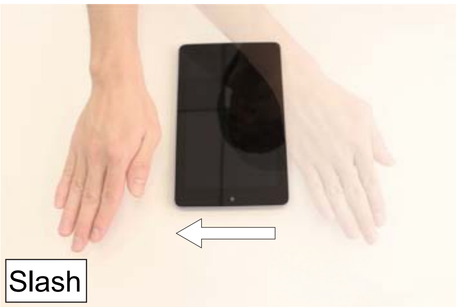


Fig. 1: Slash gesture

A "roll" gesture is a motion of turning your hand around over the built-in illuminance meter. A "slash" gesture is a motion of moving your hand in horizontally. An "up" gesture is a motion of moving your hand up while pushing it out. A "down" gesture is a motion of pulling your hand toward you while moving it down. A "hide" gesture is a motion of hiding the illuminance meter with your hand. It causes a large change in the illuminance value relative to other gestures, resulting in the illuminance value of almost 0.

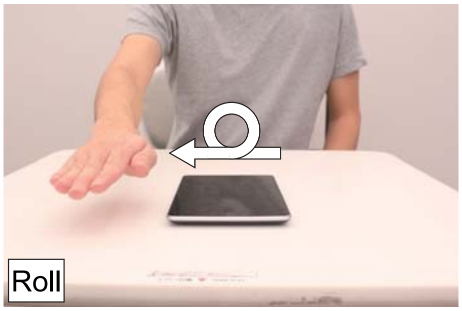


Fig. 2: Roll gesture

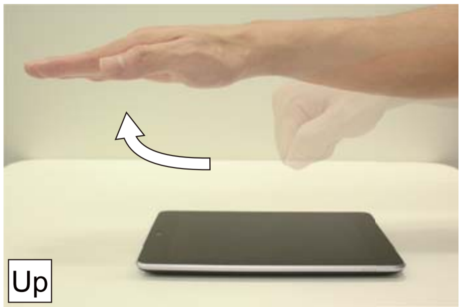


Fig. 3: Up gesture

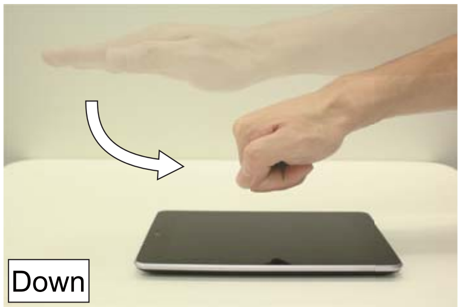


Fig. 4: Down gesture

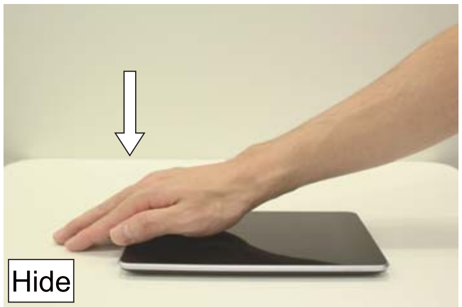


Fig. 5: Hide gesture

3.3 Hand gesture recognition algorithm

The hand gesture recognition algorithm is given below, and the explanation of each step follows.

1) Detect an illuminance change by the built-in illuminance meter and obtain illuminance values.
2) If an illuminance change greater than the threshold occurs, obtain the set of data from that point to the point where illuminance values stabilize at the original value again.
3) Extract features from the data set obtained and classify the corresponding motion as the operation mode toggling gesture or disturbance.
4) If it is classified as the operation mode toggling gesture, enter the operation mode and move to step (5). If it is classified as disturbance, go back to step (1).
5) Just as in step (1), detect an illuminance change by the built-in illuminance meter and obtain illuminance values.
6) Just as in step (2), if an illuminance change greater than the threshold occurs, obtain the set of data from that point to the point where illuminance values stabilize at the original value again.
7) Extract features from the data set obtained and classify the corresponding motion as the operation mode toggling gesture or other hand gestures.
8) If it is classified as the operation mode toggling gesture, exit the operation mode and go back to step (1). If it is classified as other hand gestures, execute the process assigned to each hand gesture.
9) Repeat steps (5) through (8) until the operation mode toggling gesture is recognized.

First of all, HGI/LI detects an illuminance change by the built-in illuminance meter and obtains an illuminance value. In obtaining illuminance values, values obtained by the sensor fluctuate even if the lighting luminance is kept constant. This issue is resolved by setting a threshold based on the result of a preliminary experiment. If a change in illuminance values is below the threshold, the current state is judged to be such that the lighting luminance is kept constant and that no hand gesture is made. If a change in illuminance values is greater than the threshold, a set of data is to be obtained from that point to the point where illuminance values stay constant again. Features are extracted from the data set obtained, and hand gesture recognition is performed by using extracted features and classification models obtained in advance through machine learning. Feature extraction and the hand gesture classification method in HGI/LI are elaborated in the next subsection.

Let us now explain steps (3), (8), and (9). In performing hand gesture recognition using light-dark change, it is necessary to consider disturbance, just as in other methods. An illuminance meter detects a light-dark change due to the tilt of the device and also affected by the shadow of a person or papers. In order to enhance the precision in hand gesture recognition, this interface has the operation mode. The operation mode corresponds to steps (5) through (9). Operations by other hand gestures are enabled after entering the operation mode using the operation mode toggling gesture. In this study, a "hide" gesture is adopted as the operation mode toggling gesture. Since performing an operation by a "hide" gesture causes a large change in illuminance values, resulting in the illuminance value of almost 0, it is considered possible to distinguish a "hide" gesture not only from other hand gestures but also from disturbance.

In HGI/LI, it is also necessary to consider a change in the light environment. A change in the light environment occurs when a user moves to a room whose light environment is different or when the lighting luminance is changed in an environment in which a controllable lighting system is used. A light environment herein refers to an illuminance environment. This issue can be dealt with by resetting the data set subject to detection when a change in illuminance values stabilizes at or below the threshold for a specified time or longer and by going back to step (1) from step (2) or to step (5) from step (6).

3.4 Feature extraction and classification of hand gesture

Performing an operation by hand gesture causes a change in illuminance values given by a built-in illuminance meter to generate a wave of illuminance values. This wave is obtained as a data set, from which features are extracted.

HGI/LI classifies hand gestures by extracting the total of four features: the number of waves of illuminance values and features D, S, and Tt given by equations (1) through (3).

$$D = \frac{A}{I} \tag{1}$$

$$S = \left|\frac{A}{T_s}\right| - \frac{A}{T_e} \tag{2}$$

$$T_t = T_s + T_e \tag{3}$$

A : amplitude, I : illuminance environment [lx]
T_s : time from start point to deepest point of wave [ms]
T_e : time from deepest point to end point of wave [ms]

Decision tree learning is used for classification. A decision tree is composed of root node, split nodes, and leaf nodes. Classification starts at the root node, and each split node classifies an input value into one of children nodes based

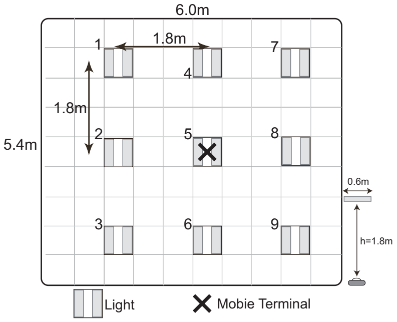


Fig. 6: Experimental environment

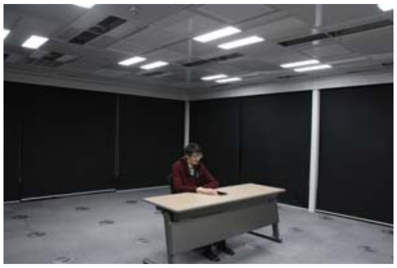


Fig. 7: Experimental situation on the recognition accuracy

on the result of learning. Classification is performed by repeating this until reaching leaf nodes.

While decision tree learning is not a method that has high classification precision, it has characteristics that it is highly accessible to human understanding and highly readable. It is one of most widely used learning methods. It also has such characteristics that, in comparison with other methods, it requires less computational complexity in performing classification and enables a faster classification. This study attempts to recognize hand gestures with less computational complexity by classifying them using a shallow decision tree based on four features mentioned above.

4. Hand gesture recognition accuracy verification

4.1 Experimental overview

A verification experiment was conducted on HGI/LI's hand gesture recognition precision. Fig. 6 shows an experiment environment, and Fig. 7, a scene of experiment. The experiment was conducted by placing a desk directly under the lamp in the center of nine LED lamps and a mobile device on the surface of the desk. The installation interval of lighting fixtures was 1.8 m, which is the same as that in a typical office. Subjects were 7 students aging from 23 to 24. The experiment was conducted after subjects were briefed for about 5 minutes about each gesture before the experiment.

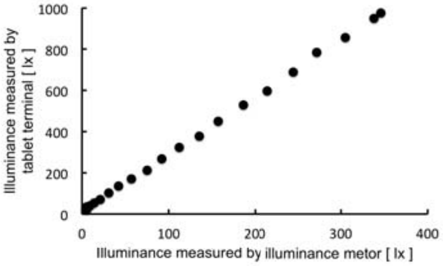


Fig. 8: Comparison of illuminance values mesured by mobile terminal and illuminance metor

In the experiment, subjects performed five types of gestures indicated in Chapter 3. Four patterns of illuminance environment were prepared, and the experiment was conducted by changing the illuminance on the desk surface to 1000 lx, 700 lx, 500 lx, and 300 lx. This was repeated 10 times in total. These 1400 data were collected to evaluate precision verification in order to avoid a bias in data resulting from learning or fatigue, instructions for gestures were randomly given and the illuminance environment was randomly changed in the experiment.

In conducting the recognition precision verification experiment, a performance verification experiment was conducted with a mobile device to be used in the former as a preliminary experiment.

4.2 Performance verification of built-in illuminance meter

As performance verification experiments of built-in illuminance meters, we conducted an experiment comparing values obtained by the built-in illuminance meter and illuminance values measured by an illuminance meter and one evaluating the response performance of the built-in illuminance meter. A mobile device used in this study is Nexus 7 (2012 model) tablet.

In the experiment comparing values obtained by the built-in illuminance meter and measured values of the illuminance meter, the illuminance meter and the mobile tablet were placed on the desk surface and a single lamp directly above them was turned on. A light fixture with dimming control in 256 levels was used in the experiment. Brightness at each step was measured by using the built-in illuminance meter and the illuminometer. As for an illuminometer, ANA-F11 made by Tokyo Koden was used. Fig. 8 shows the comparison of values obtained by the built-in illuminance meter and measured values of the illuminometer. The number of plots shown are reduced in order to make the result easier to see. As a result of the experiment, while values obtained by the built-in illuminance meter were greater than measured values of the illuminometer, the linearity of their relation was confirmed.

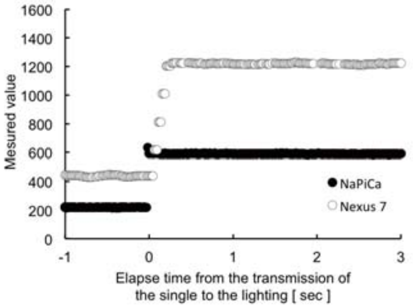


Fig. 9: Reaction performance of a built-in illuminance sensor

The response performance verification experiment of the built-in illuminance meter examined the interval at which the built-in illuminance meter obtains an illuminance value and time required for illuminance values to converge to the correct value. Just as in the experiment described above, this experiment was conducted by placing the mobile device on the desk surface and turning on only a lamp directly above.

In the experiment, time from the transmission of a light control signal to the lamp to the convergence of values obtained by the built-in illuminance meter to a constant value was measured by first turning on the lamp at 30% of the maximum lighting luminance and then raising luminance to 90% of the maximum lighting luminance. In order to measure a precise change in illuminance, illuminance values were measured, at the same time, by using NaPiCa illuminance meter made by Panasonic [12]. This meter can obtain illuminance values at an interval of approximately 1.2 ms.

The result of the experiment is shown in Fig. 9. The horizontal axis indicates the time elapsed from the time a light control signal was transmitted to the lamp, and the vertical axis indicates values obtained by the built-in illuminance meter. As a result of the experiment, it was found that the built-in illuminance meter obtained illuminance information at an interval of approximately 50 ms and that approximately 200 ms of time was required from the transmission of a light control signal to the lamp to the convergence of illuminance values to a constant value.

4.3 Experimental Results

Evaluation was done by four patterns: leave-one-out cross validation (LOOCV) using data for al subjects, LOOCV using data for each subject, leave-one-subject-out cross validation (LOSOCV), and leave-one illuminance-out cross validation (LOIOCV).

LOOCV using data for each subject, for which both test and training data are composed of data for one subject only to evaluate classification precision for each subject. LOSOCV groups data by subjects and uses data for one subject as test data and data for other subjects as training data. Generic applicability to unknown users is evaluated by using LOSOCV in evaluation. LOIOCV groups data by illuminance environments and uses data for one illuminance environment as test data and data for other illuminance environments as training data. Generic applicability to unknown environments is evaluated by using LOIOCV in evaluation.

Table 1 gives the result of LOOCV using data for all subjects, and Table 2, other results. Table 1 shows the recognition rate of all hand gestures by confusion matrix. "Per User" in Table 2 refers to the result of LOOCV using data for each subject.

Table 1: Confusion matrix for 5 recognized gestures(LOOCV)

	Hide	Roll	Up	Down	Slash
Hide[%]	96.8	0.0	2.1	1.1	0.0
Roll[%]	0.0	95.4	2.9	1.4	0.4
Up[%]	0.7	0.0	94.3	0.4	4.6
Down[%]	1.4	0.0	1.4	93.6	3.6
Slash[%]	0.0	0.0	3.2	0.4	96.4

Table 2: Results of Per User, LOSOCV, and LOIOCV

	Hide	Roll	Up	Down	Slash	Average
Per User[%]	97.5	98.6	91.8	98.2	96.4	96.5
LOSOCV[%]	95.4	96.1	90.0	96.8	91.8	94.0
LOIOCV[%]	96.1	98.2	94.3	97.9	95.4	96.4

Based on Table 1, it is found that HGI/LI shows high precision of 95.3% on average and recognizes hand gesture of each type at high precision of 93.6% or greater. In addition, based on Table 2, it is found that gestures were classified at precision of 91.8% or greater, 90% or greater, and 94.3% or greater according to evaluation by Per User, LOSOCV, and LOIOCV, respectively.

4.4 Discussion

Fig. 10 shows the decision tree created by using 1,400 data collected in this experiment. Let us dwell on what causes hand gesture recognition precision to decline in HGI/LI.

With the tablet used in this study, it takes approximately 200 ms to converge to a correct value. In addition, Fig. 10 shows that gestures are classified into "hide" gestures and other gestures by using the feature D. Consequently, when a gesture is executed so fast that it is finished before the illuminance value fully declines, a "hide" gesture may be wrongly recognized as another gesture.

Table 1 tells that the probability of wrongly recognizing "up" and "down" gestures as "slash" gestures is higher than those of other patterns of misrecognition. Based on Fig. 10, it is confirmed that HGI/LI classifies gestures into "up," "down," and "slash" gestures mostly by using the feature Tt. Therefore, this is considered as misrecognition caused by

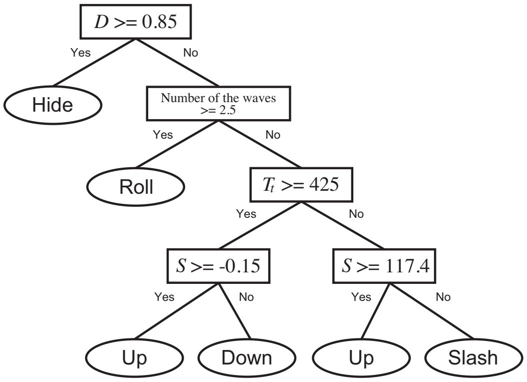


Fig. 10: Dicision tree with 1400 data collected in the experiment on recognition accuracy(D, S, and T_t is explained in Section 3.4 as Expression 1 - 3)

the fast execution speed of some "up" or "down" gestures. A "slash" gesture executed too carefully results in its slower speed, which is considered to cause its misrecognition for "up" or "down" gesture.

While LOSOCV resulted in a lower recognition rate than the rate evaluated by LOOCV using data for each subject, it still shows high precision at 94.0% on average. Based on this result, HGI/LI is considered to have high generic applicability to unknown users. By LOSOCV, however, the recognition rate was 80.5% on average over all gestures. In particular its recognition rate of "slash" gestures was low at only 60.0%.

On the other hand, LOOCV resulted in the recognition rate of 95.4% on average for all gestures and the recognition rate of 96.4% for "slash" gestures. Since high recognition rates are given by LOOCV, including data of the user who actually uses the mobile device in learning data is considered to ensure a high recognition rate.

As LOIOCV resulted in the high recognition rate of 96.4% on average, operations by hand gestures through HGI/LI are considered to be possible if the illuminance value of the environment in which HGI/LI is used ranges between 300 lx and 1000 lx.

5. Usability evaluation

5.1 Experiment overview

This section describes a subjective experiment conducted with 8 students aging from 22 to 24 to evaluate the usability of each hand gesture in HGI/LI. A questionnaire survey and an interview were conducted with subjects after the experiment to provide materials for reflection on HGI/LI.

In this experiment, a demonstration of about five minutes was first given to subjects. We then had subjects actually use HGI/LI and conducted a subjective questionnaire survey about each gesture. The questionnaire was composed of 5-point Lickert Scale questions and asked respondents to evaluate the ease of each gesture. The experiment environment is the same as that of the precision verification experiment described in the previous subsection. The illuminance environment is set to 700 lx. The experimental environment is shown in Fig. 11.

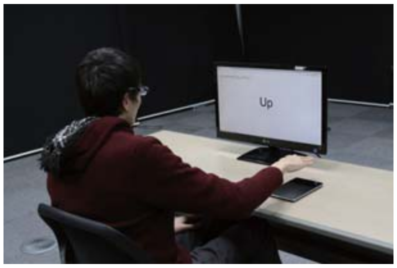


Fig. 11: Experimental situation on the usability evaluation

5.2 Application

In order to evaluate usability, we implemented HGI/LI and created an application displaying the result of classification of a gesture on a PC display. In this application, a "hide" gesture was adopted as an operation mode toggling gesture. We therefore conducted a preliminary experiment for classifying gestures into "hide" gestures and other disturbance before creating the application. The preliminary experiment was conducted in the experiment environment shown in Fig. 6.

In the preliminary experiment, disturbance data affecting gesture recognition were collected first with cooperation by three subjects. In this experiment, illuminance changes occurring upon the following motions were collected as disturbance data.

- Look into the mobile device
- Stand up and sit down
- Leave the seat
- Sit the seat
- Move along the table
- Move the paper above the mobile device
- Tilt the mobile device

We had three subjects repeat the above motions five times, changing the speed of each motion each time. Accordingly, 105 data were collected. Then, learning was done using "hide" gesture data collected in the illuminance environment of 700 lx, as described in the previous section, and these disturbance data to create a decision tree.

This application classifies gestures into operation mode toggling gestures and other disturbance data by using this decision tree. The application created also classifies gestures by using the decision tree created on the basis of 1,400 data shown in the previous experiment.

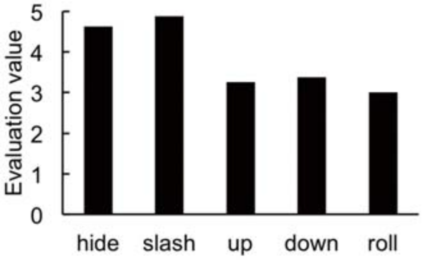


Fig. 12: Evaluation for easiness of each gestures

5.3 Results and Consideration

Fig. 12 gives the result of the questionnaire survey about the ease of each gesture. Fig. 12 shows that, while "hide" and "slash" gestures obtained high scores, the scores for other three kinds of gestures were below 4. On "up" and "down" gestures, there were remarks such as "It is a bit hard to move my elbow, wrist, and fingers at the same time," and "I find back-and-forth motions hard such as pulling my hard toward me, pushing my hand forward." On a "roll" gesture, there were remarks such as "Interaction time is long," and "A motion that includes not only horizontal but also vertical movements is bothersome." A "roll" gesture is a motion such that you move your hand in a circle across the built-in illuminance meter back-and-forth 1.5 times. This gesture is distinguished from other gestures by using the number of waves of illuminance values. Therefore, it is misrecognized for other gestures unless the user's hand moves across the built-in illuminance meter three times. Multiple subjects felt this gesture difficult. In the future, it is considered necessary to evaluate usability again after creating a specific application using HGI/LI and conducting an experiment that compares it with other interfaces for the same operations.

6. Conclusion

This study examined HGI/LI, hand gesture interface using light-dark changes in the illuminance meter built in a mobile device. While studies on gesture interface often uses an infrared camera or depth sensor to recognize gestures, doing so requires a dedicated device. As HGI/LI uses the illuminance meter built in a mobile device, which has rapidly spread in recent years, its introduction cost can be reduced. In addition, HGI/LI minimizes computational complexity and realizes a lighter application as it uses light-dark changes in an illuminance meter and perform classification only by using a shallow decision tree.

As a result of the experiment for verifying the hand gesture recognition precision, HGI/LI showed recognition rates of 93.6% or above for all gestures. It was found that hand gestures can be classified into five types using a shallow decision tree based on four types of features extracted from illuminance information. The generic applicability of HGI/LI to unknown users and illuminance environments was also confirmed.

References

[1] Cohn G. and Morris D. and Pate S. and Tan D.，"Humantenna: Using the Body as an Antenna for Real-time Whole-body Interaction,”，"In Proc CHI 2012，pp.1901-1919

[2] Gupta S. and Morris D. and Patel S. and Tan D.，"Soundwave: Using the Dopper Effect of Sense Gestures，"In Proc CHI 2012，vol.1，no.1322，pp.1911-1914

[3] Harrison C. and Tan D. and Morris D., "Skinput: Appropriating the Body as an Input Surface, "In Proc CHI 2010, pp.456-462

[4] Shotton J. and Fitzgibbon A. and Cook M. et al, "Real-time human pose recogniition in parts from single depth images, "In Proc CVPR 2011, pp.189-192

[5] Leap Motion, https://www.leapmotion.com/

[6] Lee J. and Olwal A. and Ishii H. and Boulanger C., "SpaceTop: Integrating 2D and Spatial 3D Interactions in a See-through Desktop Environment, "In Proc CHI 2013, pp.189-192

[7] Taylor S. and Keskin C. and Hilliges O. and Izadi S. and Helmes J., "Type-hover-swipe in 96 Bytes: A Motion Sensing Mechanical Keyboard, "In Proc CHI 2014, pp.1695-1704

[8] Juan W. and Helman S. and Yael E. and Michael G. and Craig F. and Mark S. and Jon H., "A Real-Time Hand Gesture Interfece for Medical Visualization Applications, "Applications of Soft Computing, pp.153-162

[9] Zhao C. and Chen K. and Aumi M. T. I. and Patel S. and Reynolds M. S., "SideSwipe : Detecting in-air Gestures Around Mobile Devices Using Actual GSM Signals, "In Proc UIST 2014, pp.527-534

[10] Goel M. and Lee B. and Aumi I. T. Md. and Patel S. and Borriello G. and Hibino S. and Begole J., "SurfaceLink: Using Inertial and Acoustic Sensing to Enable Multi-device Interaction on a Surface, "In Proc CHI 2014, pp.1387-1396

[11] Song J. and Sörös Gábor and Pece F. and Fanello S. R. and Izadi S. and Keskin C. and Hilliges O., "In-air Gestures Around Unmodified Mobile Devices, "In Proc UIST 2014, pp.319-329

[12] Panasonic:illuminance meter NaPiCa, http://www3.panasonic.biz/ac/download/control/sensor/illuminance/catalog/bltn_jpn_ams.pdf.

Infrastructured Mobility Management Approach for Future Internet ETArch Networks

Felipe Dantas Silva[1], Augusto Neto[2], Douglas Maciel[2], José Castillo-Lema[3], Flávio Silva[4]
[1]Federal Institute of Education, Science and Technology of Rio Grande do Norte (DIATINF/IFRN), Brazil
[2]Department of Informatics and Applied Mathematics (DIMAp), Federal University of Rio Grande do Norte, Brazil
[3]Polytechnic School (Poli-USP), University of Sao Paulo, Brazil
[4]Faculty of Computing (FACOM), Federal University of Uberlândia, Brazil
felipe.dsilva@ifrn.edu.br, augusto@dimap.ufrn.br, braz@lcc.ufrn.br, josecastillo@usp.br, flavio@facom.ufu.br

***Abstract*—Among the current proposals for Future Internet new architectures, the Entity Title Architecture (ETArch) stands out because of its innovative approach. This system is able to integrate new features through an information-centric network that makes use of the Software Defined Networking (SDN) paradigm, exceeding the capacity of current IP-based infrastructures. However, the mechanisms adopted by ETArch for the Quality of Service (QoS) and mobility control were not designed in an integrated manner, which means they are unable to keep users well connected to mobility demands. In light of this, this paper proposes extensions to the legacy ETArch Mobility Manager which, when integrated to the QoS Manager, is able to support the following operations: *(i)* network-initiated mobility control to allow improved resource allocation; *(ii)* quality-oriented access point selection for the maintenance of best-connected mobile nodes; *(iii)* mobility load balancing, to maximize admissions of mobile nodes in conditions of congestion; and *(iv)* IEEE 802.21 compliant infrastructured handover setup. The resulting mobility control ecosystem benefits the ETArch by allowing a maximized admission of mobile sessions experiencing congestion, by means of a maximized transport capacity. The evaluations were carried out on a testbed that considered real events, and provided evidence that the proposal outperforms legacy ETArch mobility control functionalities.**

I. Introduction

Among the current initiatives for the *Future Internet* (FI), the *Entity Title Architecture* (ETArch) [1], stands out as a promising architecture since it makes use of the content-oriented paradigm, and employs a new naming and addressing scheme based on the Title. It is a realization of the *Entity Title Model* [2], and envisages how the entities should be able to semantically specify their requirements and capabilities so that they can communicate with each other. ETArch can inherently support mobile group-communication based on the OpenFlow [3] substrate within the Workspace, which is a channel that is able to bring together two or more communicating participants.

The mobility control functions of ETArch are based on the IEEE 802.21 standard (MIH – *Media Independent Handover*) [4], and are mainly designed for an exchange of access points (PoA – *Point of Attachment*).

Despite its innovative approach, ETArch was not designed to take account of important aspects of Future Internet concepts. This in particular applies to the dismissal of control mechanisms that have the capacity to establish workspaces that can support a transport model that goes beyond the current *best effort* delivery of the Internet. In other words, ETArch does not support mechanisms that allocate sufficient bandwidth to accommodate high-demand sessions with *Quality of Service* (QoS) to obtain minimum rates of loss, delay and variations in delay overtime. In addition, the ETArch mobility management model is absolutely user-centric, which means that the user is responsible for making an explicit request for a move to another PoA.

Recent work by our research group [5] has improved the ETArch ecosystem with the QoS Manager, a new control component that allows applications to semantically express quality requirements (flow, tolerance to losses and delays, codecs, etc.). In addition, the QoS Manager is responsible for allocating network resources (for class bandwidth and workspaces) dynamically and systematically, as well as accommodating sessions in line with quality requirements, especially for those with high demands (such as video streaming and voice).

The QoS Manager orchestrates the admission control functions and resource allocation in intra-domain links of ETArch wired networks. It is based on the dynamic control of an oversized resources strategy [6], to allow the admission control to be able to accommodate multiple sessions in the same workspace. This only has signaling in specific nodes (edge nodes), unlike the classic per flow model, where all of the selected path must be reconfigured to meet the demands of the new session. The strategy consists of making oversized workspaces during the system bootstrap, and aiming to make available a significant amount of workspaces in advance with an oversized bandwidth at each interface of the network nodes in each workspace.

In this manner, the QoS Manager makes local decisions in advance about the available information (without classical instantaneous collection), and only configures the flow tables of the edge nodes in the selected workspace, either to aggregate (i.e. join the workspace) or disaggregate (i.e. leave the workspace for another network) the flow packets of the demanded session.

The QoS Manager is able to accommodate multiple sessions with *Quality of Experience* (QoE) that meet their QoS requirements, while at the same time, maintaining good levels of network performance and scalability.

The admission control is activated in two situations: *(i)* during the establishment of a new session; *(ii)* or in response to the explicit request of a MN handover. This non-transparent mobility strategy that relies on an explicit request for handover (without mobility prediction), in addition to the lack of knowledge of the QoS Manager in the system mobility patterns, enables it to accommodate mobile sessions until the limit of the resources of the oversized reserves. Under conditions of resource exhaustion, the QoS Manager rejects the admission of the mobile session. In fact, this admission control behavior of mobile sessions is natural.

However, we believe in an approach that involves maximizing the admission rates of mobile sessions in congested PoAs (and without any the likelihood of increasing the depleted resources) through integrating the QoS Manager functions with an improved Mobility Manager. This hypothesis is based on the ability to make the Mobility Manager capable of moving connected sessions to a highly desired PoA (selected as the best or only alternative for moving sessions) to another PoA. The purpose of this is to release resources and thus accept the mobile session, since, currently, the QoS Manager denies access to the sessions because it is impossible to release resources (which only occurs at the end of sessions). As a result of this release of resources in response to handover, the moving sessions can be accommodated in the desired PoA, and have continuity. Furthermore, this model justifies its application in resilience and load balancing scenarios in response to dynamic network anomalies (PoA failure).

For this reason, this paper proposes making an extension to the legacy ETArch Mobility Manager, called *Quality-oriented Mobility Management Approach* (QoMMA), to support network-initiated quality-oriented handover management. Moreover, the Mobility Manager operates together with the QoS Manager in order to deploy mobility-based load balancing, and allows the admission of sessions affected by mobility patterns to be maximized by means of moving sessions in the demanding PoA to others with a greater capacity for accommodation. The proposal makes contributions in the following areas: *(i)* network-initiated mobility prediction; *(ii)* quality-oriented PoA selection; *(iii)* mobility load balancing; *(iv)* IEEE 802.21 compliant infrastructured handover setup. The evaluation was carried out in a real testbed scenario consisting of OpenFlow/802.11 access points that consider real events.

The remainder of the document is structured as follows: Section II presents the background for this work, highlighting not only the supporting technologies, but also other related approaches. Section III provides an overview of the QoMMA proposal. Section IV outlines the basic operations of QoMMA. Section V shows the results of the evaluations in the control plane. Finally, Section VI offers some concluding remarks and makes suggestions for future work.

II. Background

The popularity of wireless networks requires the development of mobility control mechanisms to support the different traffic characteristics and needs of mobile users in various infrastructural conditions [7]. The increasing demand for real-time content and services require the wireless networks management systems to provide mechanisms that support different traffic features at different levels of quality [8]. In essence, the mobility management process consists of ensuring the mobile user is *Always Best Connected* (ABC), and is responsible for offering connectivity alternatives that best suit the user's needs.

For since many years, a number of strategies have been proposed as solutions to improve the mobility management and to support the growing requirements of mobile users. The main strategies adopted in designing mobility management solutions include the use of *Received Signal Strength* (RSS) monitoring, *Multiple Attribute Decision Making* (MADM) [9] methods and Fuzzy Logic [10]. These strategies are based on the principle of PoA selection as an alternative to connectivity, and in some cases, estimates the level of quality in the network as a condition for triggering the mobility procedures.

The work in [11] employs a combination of the AHP and TOPSIS MADM methods to form a decision-making mechanism, where the considered criteria are RSS, available bandwidth and the network load. Although it yields results based on simulations, the work has limited value, since no additions were suggested to the existing methods. [12] establishes and evaluates a mobility control framework based on an IEEE 802.21 standard. The evaluation was conducted in a physical SDN testbed consisting of one OpenFlow Controller and 802.11 Openflow-enabled switches. Although it clearly demonstrates the benefits of its performance compared with other related approaches, the solution does not take into account qualitative factors, and the decision process is guided solely by the RSS of the candidate networks. [13] set out a handover decision mechanism which is based on fuzzy logic, and uses RSS prediction (PRSS – *Predicted RSS*), available bandwidth and user preferences as input parameters. The strategy does not use the actual value of the RSS but depends on a prediction, which can lead to inaccuracy in the decision-making process. Furthermore, the mobility decision is executed by the MN, and is thus unsuitable for devices with energy constraints.

Many of the proposed solutions for dealing with mobility management lacks some features of the network in terms of efficiency and Quality of Service [7]. One of the longstanding challenges in the design of mobile systems is the provision of QoS guarantees that are required by the applications in a diverse networking infrastructure [14]. Furthermore, another critical problem is the complexity involved in managing all the mobility information regarding a large number of MNs as well as the signaling overhead that is needed to control their common mobility procedures. This is an issue that can be easily be overcome as a result of the flexibility provided by the SDN framework, where network functions, including mobility and quality of service, can be simply deployed (such as the software in the control plane without computational overhead and updates to the network devices) [15].

Although several studies have explored the quality-oriented mobility control field extensively, there are extremely few which have addressed this question within the framework of a Future Internet integrated architecture. In addition, few studies are explicitly concerned with the mobility control in an integrated architecture that makes use of the SDN paradigm.

In the next session, we provide an overview of the proposed solution, by describing the new features and their relationship with the others components of the ETArch framework.

III. Overview of the QoMMA Proposal

The QoMMA proposal is composed by three main components: Decision Maker, QAMC and E2BS. The proposed extensions were developed and integrated into the DTSA [1], which acts as the SDN Controller, and thus enables mobility procedures to be managed in the network. The following subsections detail its subcomponents.

A. Decision Maker

This is the central core element of the decision-making mechanism. It is responsible for mediating the different requests to the other sub-components, such as: *(i)* changing the status of the monitoring and data collection system (QAMC) (within predefined limits), increasing efficiency (in critical situations), in processing the data collected, such as the MN moving; *(ii)* mapping the CoS to which a particular session belongs and, as a result, determining the importance of the values (weights) of the attributes, through the MADM AHP method [16], and where necessary using the E2BS subcomponent; *(iii)* sending the information with the decision of the new network to the MIHF, in cases where the handover is needed.

B. QAMC

The *Quality Attribute Monitoring and Collector* (QAMC) is responsible for monitoring and collecting the parameters that trigger: *(i)* the occurrence or need for mobility, loss or reduction of RSS (which show that the MN is moving) and; *(ii)* network quality level, through the QoS parameters. The collected data will be used by the E2BS network selection mechanism, which is outlined in the following subsection.

The QAMC monitoring interval is adjusted to the system status, defined by the Decision Maker:

- **Regular** – Every 15 seconds, to obtain network quality parameters and every 5 seconds, to obtain RSS;
- **Alert** – Every 2 seconds, to obtain network quality parameters and every second, to obtain RSS.

The regular monitoring is the default mode. In this case, the collecting is performed every 15 seconds, to obtain the network quality parameters and every 5 seconds, to obtain the RSS between the PoA and the MN. If the RSS between a PoA and MN exceeds the threshold that has been previously configured, the system runs in alert mode, which leads to an imminent disconnection of the MN. Thus, the data collecting interval will be reduced, and this will allow the decision-making system to immediately identify alternatives (selection of a new PoA), if these limits are exceeded again, which indicates the sudden need for mobility.

C. E2BS

In our previous work [7] , we proposed the *Extended Elitism for Best Selection* (E2BS), a handover decision method inspired by the Elitist Selection Strategy [17], and combined with MADM features to enable efficient quality-oriented mobility decisions. Its main goal is to meet both the quality requirements of active mobile session flows and to match the current quality standards of neighbouring PoA candidates.

The elitist strategy employed by E2BS is based on a multi-attribute evaluation of the QoS candidate networks. In our model, the population is represented by a set of PoAs and their attributes. This technique is used to select the PoA which offers the best criteria for connection. Assessing the QoS offered by the various PoA to select the best one is carried out by measuring the similarity [18] between the attributes of the elite individual, represented by the reference PoA, and the other candidates. The reference PoA is considered to be the one that has the ideal values, (i.e. attributes like delay and jitter should have values close to zero).

E2BS was based on the MADM approach and designed to deal with the attribute importance (i.e. weight) of diverse applications by means of different traffic classes with distinct requirements [19].

In [7] we carried out a performance evaluation of E2BS which confirmed that the capacity of the proposed solution was superior to that of the alternative methods currently available.

IV. QoMMA basic operation

This section provides a detailed account of the interaction between ETArch features and the new proposed operations supported by the Mobility Manager that makes use of the QoMMA functionalities.

A. System bootstrap

The system bootstrap is designed to boot the system with oversized network resources. In this case, the PoAs are configured with over-reservation resources, and this information is recorded in the state table of DTSA. Since this information will be available in advance, the Mobility Manager will be able to make admission decisions in several sessions without any signaling events either for consultation or to set up a ground of resources in the PoA.

In case there is any change in the network topology caused by the entry of a new PoA, the system bootstrap mechanism is triggered for this device. In this way, the QoS Manager sends an OpenFlow message to the new PoA, and sets the CoS over-provisioning patterns, in a way that is compatible with the underlying QoS approach (for instance, by configuring the priorities for packet scheduling).

At this stage, with the support of QAMC, the Mobility Manager will be able to identify the conditions (available bandwidth per CoS, delay, jitter, loss, RSS etc.) of each registered PoA. This information will be used by the E2BS to give priority to the candidate PoA classification.

B. Mobile session setup

This process is triggered whenever: *(i)* the DTSA receives a request from a MN to be attached to a PoA or when; *(ii)* the Mobility Manager detects the need for the mobility of a MN owing to the loss or reduction of RSS, which is mainly caused by its movement.

If the first case occurs, the requester MN must register itself at DTSA, by stating the communication requirements (required bandwidth, delay/jitter/loss tolerance etc.). If this process was triggered because of the need for mobility (which is identified

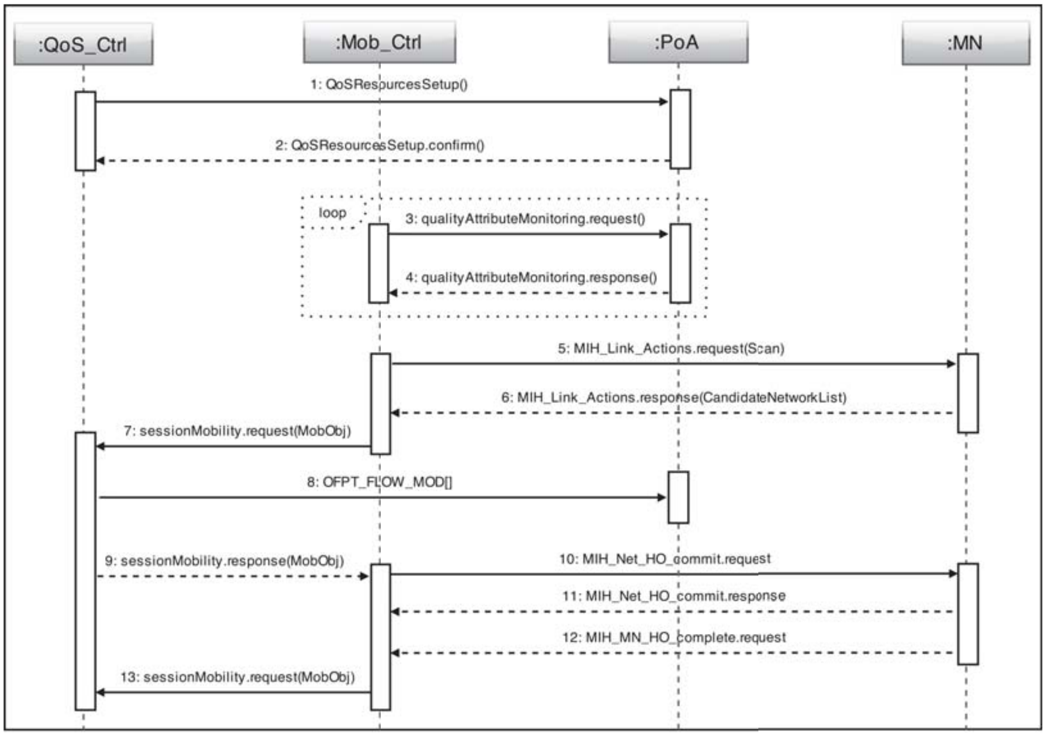

:QoS_Ctrl
:Mob_Ctrl
:PoA
:MN
1: QoSResourcesSetup()
2: QoSResourcesSetup.confirm()
loop
3: qualityAttributeMonitoring.request()
4: qualityAttributeMonitoring.response()
5: MIH_Link_Actions.request(Scan)
6: MIH_Link_Actions.response(CandidateNetworkList)
7: sessionMobility.request(MobObj)
8: OFPT_FLOW_MOD[]
9: sessionMobility.response(MobObj)
10: MIH_Net_HO_commit.request
11: MIH_Net_HO_commit.response
12: MIH_MN_HO_complete.request
13: sessionMobility.request(MobObj)

Figure 1: Generic mobility scenario

by the Mobility Manager), this information will be available in advance at the DTSA state table (in this case, the MN is already registered in the DTSA).

On the basis of this information, the QoS Manager will use the admission control mechanism to check whether the candidate PoA has capacity to accommodate the requester MN session at the desired CoS. If not, new over-reservation patterns will be applied to meet the MN request.

The new over-reservation patterns are carried out by making readjustments to the limits of each traffic class until the *Maximum Reservation Threshold* ($MRth$) or, if there is availability, through the provision of available resources in other classes to a congested class.

Figure 1 shows a generic scenario, from the system bootstrap to the requester MN handover setup in a new PoA, that displays the events and their respective signaling messages:

1) The process starts with the over-reservation PoA configuration.
2) After processing the bootstrap settings, the PoA sends a confirmation message to the QoS Manager.
3) This process is undertaken by the QAMC, and is performed by requesting information about parameters that identify the occurrence of mobility (monitoring of RSS) and of the network quality level (through SNMP and OpenFlow queue queries).
4) The query answer, consisting of a continuous process of monitoring and collecting, is sent to the Mobility Manager so that it can be used by decision methods, when necessary.
5) In the occurrence of an event, e.g. a MN is about to lose connection with the current PoA, the Mobility Manager starts the handover process by sending a notice through a *MIH_Link_Actions.request(Scan)* message, and then the MN detects candidate networks in its coverage area.
6) Through a *MIH_Link_Actions.response* message, the MN sends to the Mobility Manager a list of candidate networks. This enables it to sort the viable networks, by priority, using the E2BS method.
7) The Mobility Manager then informs the QoS Manager about the candidate networks, in order of priority, so that admission checks can be performed. If any of them has compatible resources with the MN needs, its attachment will be allowed.
8) Once the MN's admission to the network has been authorized, the QoS Manager provides the necessary resources to the target PoA.
9) The QoS Manager then notifies the Mobility Manager about the admission of the MN, so that it can setup the handover to the new network.
10) Through a *MIH_Net_HO_commit.request* message, the Mobility Manager instructs the MN to perform the handover for the selected network.
11) The MN performs the association procedure for the new PoA and informs the Mobility Manager about this operation, by sending a *MIH_Net_HO_Commit.response* message.
12) At the end of the handover procedure, the MN notifies the Mobility Manager by sending a *MIH_MN_HO_complete.request* message.
13) At this stage, the Mobility Manager knows that the MN is no longer associated with the old network and requests the QoS Manager to release all the associated resouces, in the old PoA.

C. Mobility load balancing

If the admission possibilities provided by the over-reservation mechanism are not sufficient to accommodate new mobile sessions, and shows a lack of resources in the PoA, the Mobility Manager will release resources through a mobility load balancing operation, that reduces the effects of this scarcity, and as a result, the rejection of new mobile session requests.

This process consists of moving already associated MNs in the required PoA to another feasible PoA that is within its coverage area and provide available resources. Through this operation, it is possible to maximize admissions to the network, by always keeping the MNs well connected.

On receiving a request from a MN that wishes to be associated with a PoA where the CoS does not have sufficient resources to carry out the over-reservation procedures, the Mobility Manager will identify other MNs that are already connected to this PoA that can be moved. Hence, there will be a release of sufficient resources for the admission of the requester mobile session.

Algorithm 1: Mobility load balancing description

```
Retrieve $QoS_{req}$ of the $MN_r$ attachment in $PoA_t$;
for each $MN_c(i)$ in $PoA_t$ do
    Order available networks ($PoA_c$) in $MN_c(i)$ range by priority (using E2BS);
    for each $PoA_c(j)$ in $MN_c(i)$ range do
        Perform admission control verifications in $PoA_c(j)$;
        if $PoA_c(j)$ is able to acommodate $MN_c(i)$ $QoS_{req}$ then
            Prepare required resources for $MN_c$ in $PoA_c(j)$ (OpenFlow);
            Move $MN_c(i)$ to $PoA_c(j)$ (using 802.21);
            Release all $MN_c(i)$ associated resources in $PoA_t$ (using OpenFlow);
            Prepare required resources for $MN_r$ in $PoA_t$ (OpenFlow);
            Allow $MN_r$ attachment in $PoA_t$;
            break;
Reject the $MN_r$ attachment;
```

The process for selecting the candidate MN to handover consists, initially, of identifying the MNs where the mobility results in a likelihood of a higher admission of the requester MN. This means that an individual analysis will be conducted of the alternative forms of connectivity for candidate MN that comply with certain criteria, such as: *(i)* low priority CoS; *(ii)* largest amount of reserved resources; *(iii)* equivalent reserved resources to that required by the requester MN, among others. The analysis of alternative forms of connectivity is carried out by giving priority to candidate networks in each MN coverage area, through the E2BS decision method. Each available network will be checked by the admission control process, and this will identify whether it is able to accommodate the candidate MN mobile session in the respective CoS. If so, the required resources for the candidate MN mobile session admission will be provided and then it will be transferred to the new PoA. Finally, all the reserved resources associated with the transferred MN are released and will be available for the mobile session of the requester MN.

The mobility load balancing operation is described in Algorithm 1.

Where:

- QoS_{req}: QoS requirements;
- MN_r: Requester MN;
- PoA_t: Target PoA;
- MN_c: Connected MN;
- PoA_c: Candidate PoA.

If the mobility load balancing operation cannot release the necessary resources to accommodate the requester mobile session, it will be rejected.

V. Performance evaluation

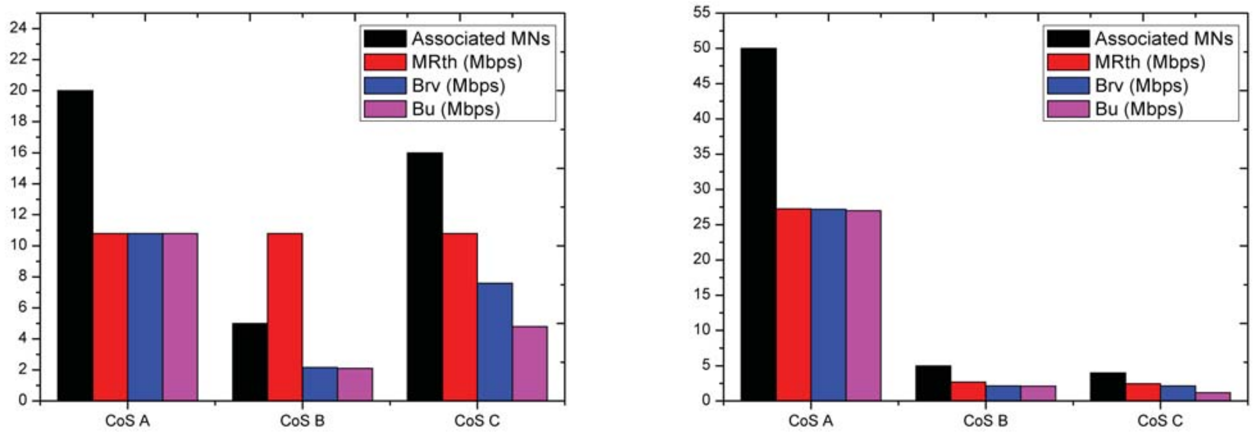
In seeking to evaluate the feasibility of our proposal, we extended the ETArch Mobility Manager implementation with the QoMMA architecture in accordance with the guidelines outlined in Section III. The aim of this evaluation was to compare the performance of the ETArch admission control strategy (without QoMMA) and QoMMA-enabled Mobility Manager, with load balancing functionalities by means of the network admission capacity of mobile sessions.

A. Evaluation Scenario

The experiments were carried out in a real testbed composed by three TP-Link TLWR1043ND routers embedding EDOBRA Switch Configuration [20], to support both IEEE 802.21 and QoS-aided OpenFlow v1.0 (queuing control) facilities. The wireless configuration of EDOBRA switches were set at in 802.11g mode. A network server hosts the DTSA OpenFlow Controller by implementing ETArch features with the facilities provided by the new Mobility Manager extensions. The testbed described above was used to perform the evaluation in the control plane, and a wide range of mobile sessions requests were considered with varying constant bitrate requirements of 450, 350 and 250 kbps [21], linked to three CoS (A, B and C), respectively.

In this scenario, we initialized each CoS over-reservation with 20% of total bandwidth, i.e., 10.8 Mbps. After the system bootstrap, all the session requests were triggered to the same AP, namely AP1. Figure 2a shows the information about each CoS of AP1 before the readjustment between the CoS was carried out.

In this case, the CoS A carried out mobile session admissions until the Maximum Reservation Threshold ($MRth$) capacity (initially configured at 20% of the total bandwidth) and took account of both the *Reserved Bandwidth* (Brv) and the *Used Bandwidth* (Bu). At this point, AP1 accomodates 20 MNs in CoS A, 5 MNs in CoS B and 16 MNs in CoS C. The graph in Figure 2a shows that the Brv of the CoS C is not aligned with the respective Bu, because following the Cisco guidelines for implementing QoS provisioning [22], the QoS Manager reserves 20% beyond the actual bitrate required. Before it could perform new mobile session admissions in CoS A, the Mobility Manager had to ask the QoS Manager to make some readjustments between the other CoS. Figure 2b shows the AP1 CoS A state after the readjustments as a result of which it was possible to admit 30 new requester MNs.



(a) AP1 CoS utilization before readjustment

(b) Readjustment between CoS in AP1

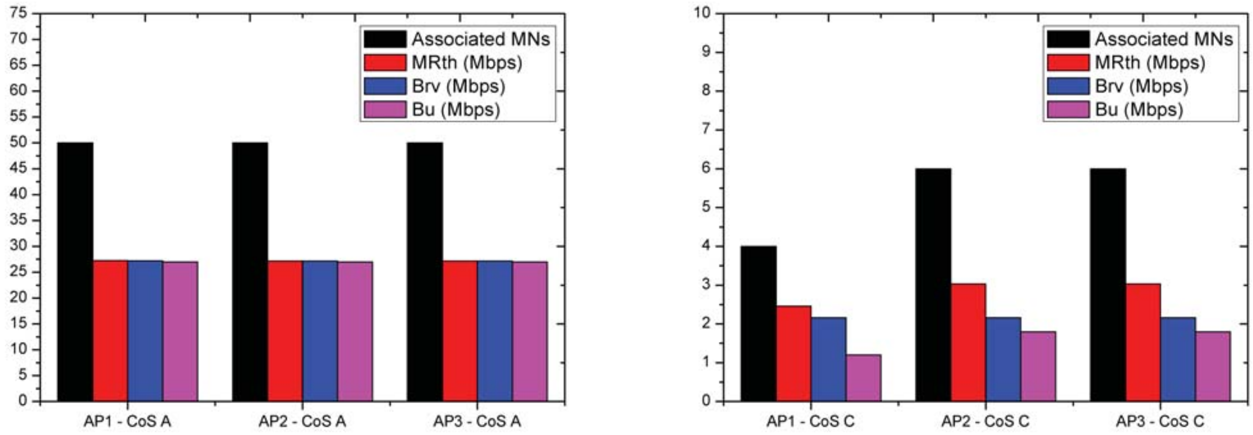
Figure 2: AP1 state before load balancing

At this point the AP1 is no longer capable of accommodating new mobile session requests in CoS A until the mobility load balancing procedure has been executed, and new resources released.

As can be seen in Figure 3a, new mobile session requests were admited through the mobility load balancing procedure until it reached the full capacity of the available resources of the network devices. The sessions accommodated in the CoS A of AP1 before, were transferred to AP2 and AP3, and resources in AP1 released, so that new mobile sessions could be received. The same ocurred with the sessions accommodated in the CoS C of AP1. Before the load balancing process (as displayed in Figure 2a, there were 16 sessions in CoS C. Figure 3b shows the scenario after this process, where AP1, AP2 and AP3 accomodates 4, 6 and 6 mobile sessions, respectively.

The results of Figure 4 reveal the maximization of the admissions of the mobile sessions which could be obtained from the facilities provided by the QoMMA proposal. It is well-know that after the QoMMA mobility procedures, it was possible to reconfigure the network, and thus, to some extent, avoid the rejection of new mobile sessions. In total, there were made 172 requests for association to AP1, and 107 of them were rejected by the approach without QoMMA and only 1 by QoMMA.

The numerical analysis confirms this behavior, and shows



(a) CoS A utilization

(b) CoS C utilization

Figure 3: CoS utilization after load balancing

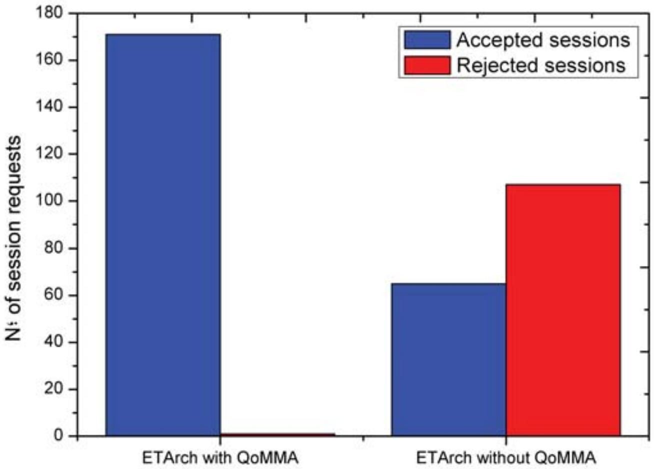


Figure 4: Admission rate with QoMMA and without QoMMA

that QoMMA increased the mobile session admission optimization at a rate of approximately 163%, for this scenario, compared with the previous admission control strategy.

VI. Conclusion and Future Works

In this paper, there has been an investigation of the Quality-oriented Mobility Management Approach (QoMMA) as an additions to ETArch, to support quality-oriented mobility procedures in the network (as in the case of mobility prediction operations). The proposed extensions follow a dynamic control of an always best connected principle, that aimed at keeping the MNs with higher QoS guarantees. It allows a dynamic and preemptive reconfiguration of the network by providing a better use of resources and the maximization of mobile session admissions. The results of the evaluation confirm these benefits while, at the same time, keeping best-connected mobile nodes. The next stage of this work is to evaluate the extensions of the proposal in a data plane and also estimate the benefits of the application perspective through different benchmarks. The objective is to confirm all the QoMMA capabilities in terms of QoS and QoE.

Acknowledgment

This work has been developed under the SMART project (MCTI/CNPQ/Universal 14/2014, grant agreement n.457051/2014-0). Authors also thank to CAPES, CNPq and IFRN. Experiments were carried out at NPITI/UPLab of Instituto Metropole Digital (IMD).

References

[1] F. Silva, M. Goncalves, J. Pereira, R. Pasquini, P. Rosa, and S. Kofuji, “On the analysis of multicast traffic over the entity title architecture,” in *Networks (ICON), 2012 18th IEEE International Conference on*, pp. 30–35, 2012.

[2] J. H. de Souza Pereira, F. O. Silva, E. L. Filho, S. T. Kofuji, and P. F. Rosa, “Title model ontology for future internet networks.,” in *Future Internet Assembly* (J. Domingue, A. Galis, A. Gavras, T. Zahariadis, D. Lambert, F. Cleary, P. Daras, S. Krco, H. Müller, M.-S. Li, H. Schaffers, V. Lotz, F. Alvarez, B. Stiller, S. Karnouskos, S. Avessta, and M. Nilsson, eds.), vol. 6656 of *Lecture Notes in Computer Science*, pp. 103–116, Springer, 2011.

[3] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, “Openflow: enabling innovation in campus networks,” *SIGCOMM Comput. Commun. Rev.*, vol. 38, pp. 69–74, Mar. 2008.

[4] D. Corujo, C. Guimãraes, B. Santos, and R. L. Aguiar, “Using an open-source ieee 802.21 implementation for network-based localized mobility management.,” *IEEE Communications Magazine*, vol. 49, no. 9, pp. 114–123, 2011.

[5] J. Castillo-Lema, F. Silva, F. Silva, P. R. Rosa, C. G. Guimarães, D. Corujo, and R. L. Aguiar, “Evolving future internet clean-slate entity title architecture with quality-oriented control plane extensions,” in *Advanced International Conference on Telecommunications (AICT 2014)*, pp. 161–167, July 2014.

[6] A. Neto, E. Cerqueira, M. Curado, P. Mendes, and E. Monteiro, “Scalable multimedia group communications through the over-provisioning of network resources,” in *Management of Converged Multimedia Networks and Services, 11th IFIP/IEEE International Conference on Management of Multimedia and Mobile Networks and Services, MMNS 2008, Samos Island, Greece, September 22-26, 2008. Proceedings*.

[7] F. Silva, J. Castillo-Lema, A. Neto, F. Silva, P. Rosa, D. Corujo, C. Guimaraes, and R. Aguiar, “Entity title architecture extensions towards advanced quality-oriented mobility control capabilities,” in *Computers and Communication (ISCC), 2014 IEEE Symposium on*, pp. 1–6, June 2014.

[8] J. Sen, “Mobility and handoff management in wireless networks,” in *Trends in Telecommunications Technologies*, 2010.

[9] P. K. Yoon, C.-L. Hwang, and K. Yoon, *Multiple Attribute Decision Making: An Introduction (Quantitative Applications in the Social Sciences)*. Sage Pubn Inc, Mar. 1995.

[10] M. Zekri, B. Jouaber, and D. Zeghlache, “A review on mobility management and vertical handover solutions over heterogeneous wireless networks,” *Comput. Commun.*, vol. 35, pp. 2055–2068, Oct. 2012.

[11] M. Sharma, “Multi network handover decision using extended multi criteria decision making technique,” *IU-Journal of Electrical & Electronics Engineering*, vol. 14, no. 1, 2014.

[12] C. Guimaraes, D. Corujo, R. Aguiar, F. Silva, and P. Frosi, “Empowering software defined wireless networks through media independent handover management,” in *Global Communications Conference (GLOBECOM), 2013 IEEE*, pp. 2204–2209, Dec 2013.

[13] M. Sharma and D. K. Khola, “Fuzzy logic based handover decision system,” *Wireless Communication*, vol. 4, no. 13, 2012.

[14] J. Chan and A. Seneviratne, “A practical user mobility prediction algorithm for supporting adaptive qos in wireless networks,” in *Networks, 1999. (ICON ’99) Proceedings. IEEE International Conference on*, pp. 104–111, Sept 1999.

[15] K.-H. Lee, “Mobility management framework in software defined networks,” *International Journal of Software Engineering and Its Applications (IJSEIA)*, vol. 8, no. 8, 2014.

[16] S. Dhar, R. Bera, and A. Ray, “Design and simulation of vertical handover algorithm for vehicular communication,” *International Journal of Engineering Science and Technology*, vol. 2, no. 10, pp. 5509–5525, 2010.

[17] A. P. Engelbrecht, *Computational Intelligence: An Introduction*. Wiley Publishing, 2nd ed., 2007.

[18] Z. Tang, Y. Zhu, G. Wei, and J. Zhu, “An elitist selection adaptive genetic algorithm for resource allocation in multiuser packet-based ofdm systems.,” *JCM*, vol. 3, no. 3, pp. 27–32, 2008.

[19] 3GPP, *QoS Concepts and Architecture: TS 23.107, 3rd Generation Partneship Project (3GPP)*. 2009.

[20] EDOBRA, “Edobra switch os - odtone openwrt,” 2013. Available at: https://github.com/ATNoG/odtone-openwrt. Acessed 4 December 2014.

[21] Microsoft, “Lync server 2013 network bandwidth requirements for media traffic,” 2013. Available at: https://technet.microsoft.com/en-us/library/jj688118.aspx. Acessed 14 May 2015.

[22] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017,” 2013. Available at: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html. Acessed 14 December 2014.

A Classification and Comparison Between Clustering Algorithms for Wireless Networks

Ula'a A. Al-Haddad, Ghadah Aldabbagh
Computer Science Department, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract - *Clustering is a vital research topic for wireless networks because clustering makes it possible to guarantee essential levels of system implementation, such as throughput, interference and delay, in the presence of proliferation of smart-phone and other mobile devices. A large variety of approaches for wireless clustering have been presented, whereby dissimilar approaches usually focus on different performance metrics. This paper presents a classification and an analysis of clustering approaches and algorithms, mainly based on whether nodes belong to one cluster or several clusters. This paper aims at providing an overview of general approaches used in different wireless techniques and provides descriptions for widely used algorithms and comparison between well-known clustering schemes.*

Keywords: Clustering, Networks, Wireless, Cellular Networks, Base Station (BS), Cluster Header (CH).

1 Introduction

Nowadays, the growth in wireless communication technologies and the increase in smart phone usage and personal computing have gained universal attention.

Many people have the benefit of and rely on networking applications due to the great publicity of Internet services. However, this expansion has created the demand for Internet to be available anytime and anywhere, and hence it cannot satisfy people's demand for networking communication. Clustering is a key technique used to increase the lifetime of network by reducing energy and power consumption also it increases network scalability and capacity [1]. Mobile ad hoc networks (MANETs) are a kind of network with no fixed infrastructure which allows mobile nodes to establish a opportunistic network for immediate using in disaster and urgent cases [2].

It has been proved [3, 4] that a flat structure, that is, a structure with no grouping or hierarchy, encounters capacity and scalability problems with increased network size, especially in the presence of simultaneous node mobility. Moreover, a flat structure either based on reactive or proactive routing schemes and communication overhead of link based proactive routing protocols are typically $O(n^2)$, where n is the total number of mobile nodes in a network [4]. Consequently, for achieving a better performance in a large-scale wireless or cellular networks, a hierarchical architecture is usually helpful.

Since a cluster structure is a very basic hierarchy structure, in this paper we present a proposal of a taxonomy that accounts for several proposed clustering algorithms used in cellular networks. The rest of this paper is organized as follows: section II presents an overview of clustering; section III shows a comparison between clustering algorithms, section IV has a discussion on applicability of the different approaches to wireless and cellular network scenarios and finally we conclude this study in section IV.

2 What is Clustering?

Clustering is a general technique defined in [5] as a "division of data into groups of similar objects. Each group, called a cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups". Clustering can be considered one of the most important unsupervised problems; as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. Another definition of clustering could be "the process of organizing objects into groups whose members are somehow similar". A clustering algorithm is therefore a partitioning process, which divides data into clusters whose objects are similar to each other and dissimilar to the objects belonging to other clusters, according to specific criteria. Clustering algorithms have many uses, for instance they have been used in the medical field to identify cancer occurrence in certain populations[6, 7].

Clustering algorithms for wireless networks might differ, depending on the application and network architecture. Fig. 1 shows an example of a cluster structure. In this graph, the nodes are divided into a number of virtual groups, based on specific rules. The nodes could have different roles, such as cluster heads (CHs), cluster members or cluster gateways. A cluster head usually represents local cluster members. A cluster gateway is a node that can access neighbor clusters and may be used to transfer information between clusters. A cluster member is usually an ordinary node which receives a service from the cluster head.

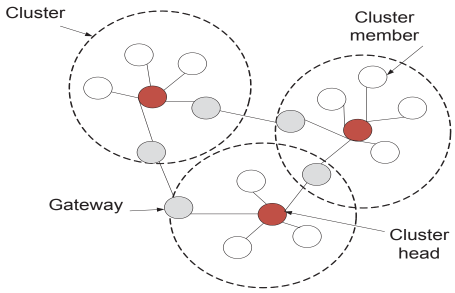


Fig 1: *Clusters structure illustration*

3 Why Clustering in Wireless Networks?

It has been shown that cluster structure is an efficient way for topology control [8]. In clustering scheme, nodes can be partitioned into a number of small groups called clusters in order to support data aggregation through effective network organization. Clustering results in a two-layers hierarchy in which cluster heads (CHs) form the higher layer while member nodes form the lower layer. The CHs aggregate the data that came from member nodes and send them to the central base. (CHs) lose more energy compared to member nodes, since they frequently transmit data over longer distances. Therefore, network may be re-clustered occasionally in order to select energy-plentiful nodes to serve as CHs, as a result distributing the load equally on all the nodes. In addition achieving power efficiency, clustering reduces packet collisions and channel disputation, consequential in better network throughput under high capacity [1]. Also, in hierarchical routing protocols whole network nodes is divided into several clusters and cluster-head is the only node that can communicate to Base station that's reduces the routing overhead of normal nodes by transmit only to cluster-head [9]. Figure 1 illustrates data flow in a clustered network. There are at least three benefits we can get from clustering [10, 11]:

1) A cluster structure increases the system capacity by having a more efficient spatial reuse of resources, such as transmission frequency (two clusters may set up the same frequency if they are not adjacent clusters [9]). Also, it allows for a better way of coordinating transmission events and saving resources used for retransmission. This results in reduced transmission collision.

2) In a cluster structure the generation and distribution of routing information can be limited to cluster heads and cluster gateways, because they can form a virtual backbone for inter-cluster routing [12, 13].

3) A cluster structure makes networks, such as ad hoc networks, look more stable to each mobile node [10]. In this way, when a mobile node moves to a different cluster, only mobile nodes that belong to the involved clusters need to update their information. This means that local changes do not necessarily need to be updated in the whole network [14, 15].

Therefore, in dense areas and in the presence of a large number of mobile nodes and high mobility clustering, it is important for a network to achieve good scalability and increasing its capacity. Clustering algorithms should meet all or some of these important requirements: scalability, optimality according to particular criteria, usability, handling of different types of attributes, ability to deal with noise, and adaptability to changes, among others. In [16] proposed new technique QTHN that converts the entire dense wireless network into hexagonal clusters via two layer network communication using clustering approach. In a cluster, a cluster head is selected then acts as a master node known as a Hotspot (HS), which is connected directly to the Base station (BS). This proposed QTHN aims to improve QoS using LTE and White Spaces in a wireless dense area. Study in [17] has proposed Distributed dynamic load balancing (DDLB) cellular-based TVWS and LTE technique, whereas by simply switching a cellular-based device's frequency when necessary to allow operate on both bands. In [18] also proposed algorithm iteratively clusters the nodes into hotspots and slaves and allocated resources to maximize spectrum utility using tethering over white spaces (WS) without need to deploy new infrastructure. That allows cellular systems to evolve hierarchically in dense areas as necessary. In [19], authors proposed a new dynamic protocol for clustering Dynamic Clustering Protocol (DCP) considering the possible changes happening in a cellular network. This will reduce the required time and signaling and enhance service quality for the cluster users. In [20] had the same case of dense area networks it compared the access types of femtocells and presented the related challenges in terms of mobility management , femtocell interference and offloading. Study in [21] authors extend existing clustering configuration to consider mobile users' requirements and network events by studying the corresponding handover scenarios and signaling schemes which will reduce the complexity of the clustering algorithm from where of time and signaling.

4 Classification Clustering Approaches

Clustering approaches may be classified according to different criteria. In this paper, we classify the clustering approaches depending on whether a certain node belongs to one cluster or several clusters.

4.1 Exclusive Clustering

In this type of clustering algorithms, data is grouped in an exclusive way. Therefore, if a certain node belongs to a particular cluster, then it could not be included in any other cluster. This is called “hard clustering”. An example is the K-means algorithm [15]. Fig. 2 illustrate this kind of clustering.

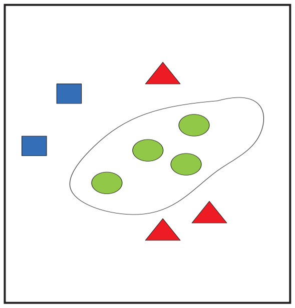


Fig 2: *Exclusive clustering.*

4.2 Overlapping Clustering

The overlapping clustering (Fig. 3) uses fuzzy sets to cluster data, so that each node may belong to two or more clusters with different degrees of membership. An example is the Fuzzy C-means algorithm [22]. This type of algorithms is also called “soft clustering”.

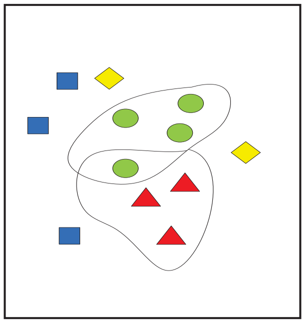


Fig 3: *Overlapping clustering.*

4.3 Hierarchical Clustering

With this kind of clustering approach, a node may belong to several clusters and clusters are built in a hierarchical way. This is a special case of an overlapping clustering approach that deserves a separate analysis. In this clustering approach two or more clusters may form a new cluster up in the hierarchy. The top cluster in the hierarchy is the whole node set and the bottom clusters may be defined as unit sets with each of the nodes.

Hierarchical clustering algorithms can be divided into bottom-up and top-down clustering algorithms. Bottom-up algorithms are based on the union of two or more clusters in order to form an upper-level cluster. The initial condition is performed by defining every node as a cluster. Top-down algorithms work by splitting clusters into K-means is a well-known clustering method. Nodes are partitioned into k groups, where k has been initially chosen. The cluster members remain as close as possible to each other but as far as possible from members of other clusters. A cluster is built around a central node that is called as “centroid”.

A centroid in wireless networks could be a node whose coordinates can be calculated as the average value of each of the coordinates of all nodes assigned to the cluster. After smaller ones, starting from the top cluster. Clusters resulting from a splitting process belong to the following level down in the hierarchy. Fig. 4 illustrates an example of a hierarchy clustering approach.

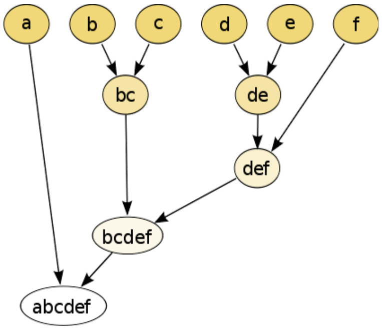


Fig 4: *Hierarchical clustering*

5 Clustering Algorithms

In this section three representative algorithms, one for each of the clustering approaches, are described and compared. These algorithms are K-means, Fuzzy C-means and a standard bottom-up hierarchical algorithm.

5.1 K-Means Algorithm

calculating the centroid for each group then cluster membership is determined by assigning each node to the group with the nearest centroid. This concept minimizes the overall dispersion within a cluster by iterative reallocation of cluster members.

K-Means clustering is an efficient algorithm for large data sets with both numeric and categorical [23] attributes.

Algorithmic steps for K-Means clustering: [24]

1) **Initialize K**. Choose a number of clusters, K.
2) **Centroids selecting**- Randomly select the centroids in the given dataset. They are taken as the initial starting values.
   $(c_1, c_2, \ldots. c_k)$
3) **Classification**: Assign each point in the node set to the cluster whose centroid is nearest to it. Compute the distance between the centroids and points using the Euclidean Distance equation.

$$d_{ij} = \left\| x_i - c_k \right\|^2$$

4) **Centroid calculation**: After each node in the node set is assigned to a cluster, recalculate the new k centroids and update the centroids.
5) **Convergence criteria**. The process will stop when no point changes its cluster or until the centroids no longer move. Otherwise, go to step (3).

5.2 Fuzzy C-Means Algorithm

A well-known fuzzy clustering algorithm is Fuzzy C-Means (FCM). The central idea in fuzzy clustering is that there is a non-unique partitioning of data in a collection of clusters. FCM is a clustering algorithm that is applied to a wide variety of problems related to feature analysis, clustering and classifier design. FCM is widely applied in sectors such as image analysis, agricultural engineering, astronomy, geology and chemistry, among others. It can also be applied for wireless network node clustering [25].

In FCM applied to node clustering, a node set is grouped into n clusters with every node related to each cluster. A node will have a high degree of belonging to a cluster if it is near of the cluster center and a low degree of belonging to clusters whose center is far away from it [26, 27].

Algorithmic steps for Fuzzy C-Means clustering: [20]

Given a finite set of nodes, the algorithm returns a list of **c** cluster centers:

$$v = \{v_1, \ldots v_c\}$$

and a partition matrix U

$$\mu = \mu_{ij} \in [0,1], i = 1,\ldots,\mathrm{n}, \mathrm{j} = 1,\ldots,\mathrm{c}$$

where each element μ_{ij} is the degree to which element belongs to cluster c_j . x_i

1) *Calculate the V center vector.*

$$v_{ij} = \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_{kj}}{\sum_{k=1}^{n} (\mu_{ij})^m}$$

2) *Calculate the distance matrix D[c,n].*

$$D_{ij} = \left(\sum_{j=1}^{m} (x_{kj} - v_{ij})^2 \right)^{1/2}$$

3) *Update the partition matrix for the r^{th} step, $U^{(R)}$ as*

$$\mu_{ij}^{\;r-1} = \left(1 / \sum_{j=1}^{c} (d_{ik}^{r} / d_{jk}^{r})^{2/m-1} \right)$$

Where c is the number of clusters (2<=c<n), and m is the level of fuzziness. that is, a larger m results in smaller memberships, μ_{ij} *and hence, fuzzier clusters.* $\mu_{(0)}$ is an initial partition matrix, given as input.

The algorithm will stop when $\| \mu(k+1) - \mu$ (k) $\| < \delta$ otherwise it will return to step 1 by updating the cluster centers and also the membership grades for each node [28].

5.3 Hierarchical Algorithms

Hierarchical clustering techniques create a nested sequence of partitions either from unit node sets at the bottom or from the set of all nodes at the top. In contrast with partitioning algorithms such as K-means, hierarchical algorithms either combine or divide clusters and build the hierarchy by merging or splitting the clusters initially defined. The result of a hierarchical clustering algorithm can be graphically illustrated as a tree called dendogram, which depicts the merging or splitting process and the middle clusters.

Then, the two fundamental approaches to generate a hierarchical clustering are:

a) **Agglomerative (bottom-up)**: Begin with clusters containing a single node and merge the most similar pair of clusters at each step. This approach is called Hierarchical Agglomerative Clustering (HAC).

b) **Divisive (top-down)**: Begin with one cluster containing all nodes and splits clusters at each step until only clusters of individual nodes remain. In this case, we need to decide which cluster to split and how to perform the split in each step**.**

Agglomerative techniques are more commonly used than divisive techniques.

Algorithmic steps for Hierarchical Agglomerative Clustering [5] :

1. Compute the similarity matrix containing the distance between each pair of patterns (clusters).

Parameters	K-means	Fuzzy C- means	HAC
Approach	Exclusive (partitions)	Overlapping	Hierarchical
Efficiency	Fairer	Slower	The slowest
Complexity	O (nkdi) Where: n is the number of data points d-dimensional vectors k the number of clusters i the number of iterations needed until convergence	O (ndc^2i) Where: i number FCM over entire dataset. n number of data points. c number of clusters d number of dimensions	O(n^2 log n) where: 'n' is the number of data points Agglomerative: O(n3) Divisive: O(2n)
Application	- Image retrieval algorithms. - General-purpose even cluster size. - Flat geometry.	- Segmentation of magnetic resonance imaging (MRI). - Analysis of network traffic. - Fourier –transform infrared spedtroscopy (FTIR).	- Ray casting. - Some bioinformatics applications.
Performance	Traditional and Limited use.	Can be used in variety of clusters and can handle uncertainty.	- Flexibility due to a level of granularity. - Easily of conduct of any forms of distance or similarity. - More applicability and versatile.
Size of data	Huge/small	Huge/small	Huge/small
No. of clusters	Large/small	Large/small	Large/small
Capability of tackling high dimensional data [29]	No	No	No

Table 1 explains how these three algorithms are compared.

2. Merge the most similar pairs of clusters. Update the similarity matrix to reflect the pairwise similarity between the new cluster and the original clusters.

3. Repeat steps 2 and 3 until only a single cluster remains.

6 Application on Wireless Networks

From the point of view of their applicability on wireless network node clustering, the described algorithms have different advantages and disadvantage With regard to the specific issues described in section 1.a., all described schemes may potentially behave in a different way.

The first issue, reuse of resources such as transmission frequency, may lead us to think that clustering algorithms which define overlapping clusters may be more aware of cluster gateways that share frequency attributes with the clusters they belong to.

With respect to the second issue mentioned in section 1.a., the possibility of defining deputy nodes for inter-cluster communication makes us think that algorithms which are based in centroid definitions are better. Hierarchical algorithms present a disadvantage in this case.

The third issue, which is dealing with mobility, does not seem to be a factor for deciding in favor of any of the schemes. In any case a change in the belonging attributes may be enough for moving a node between clusters.

However, the tree-like nature of hierarchical clustering approaches may have the advantage of communication efficiency in terms of hop count, as a tree may be an optimized graph structure with minimal paths between nodes, as long as they are balanced.

7 Conclusion

This paper presents a study of some common clustering approaches and algorithms, classified according to whether a node can be included in a single cluster or several clusters. A comparison of three common algorithms of clustering, namely K-means Clustering, Fuzzy C-Means Clustering and Hierarchical Agglomerative Clustering, is presented. We discussed the similarities and differences between these schemes, as well as their features, such as node overlapping, algorithms and application scenarios. A discussion on this clustering approach in terms of suitability to wireless networks was finally presented.

Acknowledgment

This paper was funded by King Abdulaziz City for Science and Technology (KACST), titled: "High Density Wireless Access- Coverage for Crowded Events and Locations" (under grant numbered: 12-INF2743-03). The author, therefore, acknowledges with gratitude KACST's technical and financial support. Also, the author would like to thank Dr. Nikos Dimitriou and Dr. Tahir Bakhsh for their support and guidance.

References

[1] R. Mitra and D. Nandy, "A survey on clustering techniques for wireless sensor network," in *International Journal of Research in Computer Science*. vol. 2, 2012, pp. 51-57.

[2] C. E. Perkins, *Ad hoc networking*: Addison-Wesley Professional, 2008.

[3] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *Information Theory, IEEE Transactions on,* vol. 46, pp. 388-404, 2000.

[4] K. Xu, X. Hong, and M. Gerla, "An ad hoc network with mobile backbones," in *Communications, 2002. ICC 2002. IEEE International Conference on*, 2002, pp. 3138-3143.

[5] O. A. Abbas, "Comparisons Between Data Clustering Algorithms," *Int. Arab J. Inf. Technol.,* vol. 5, pp. 320-325, 2008.

[6] R. T. Ng, J. r. Sander, and M. C. Sleumer, "Hierarchical cluster analysis of SAGE data for cancer profiling," in *BIOKDD*, 2001, pp. 65-72.

[7] E. Malo, R. Salas, M. n. CatalÃ¡n, and P. LÃ³pez, "A mixed data clustering algorithm to identify population patterns of cancer mortality in Hijuelas-Chile," in *Artificial Intelligence in Medicine*: Springer, 2007, pp. 190-194.

[8] R. Rajaraman, "Topology control and routing in ad hoc networks: A survey," *ACM SIGACT News,* vol. 33, pp. 60-73, 2002.

[9] S. E. L. Khediri, N. Nasri, A. Wei, and A. Kachouri, "A New Approach for Clustering in Wireless Sensors Networks Based on LEACH," *Procedia Computer Science,* vol. 32, pp. 1180-1185, 2014.

[10] A. B. McDonald and T. F. Znati, "A mobility-based framework for adaptive clustering in wireless ad hoc networks," *Selected Areas in Communications, IEEE Journal on,* vol. 17, pp. 1466-1487, 1999.

[11] T.-C. Hou and T.-J. Tsai, "A access-based clustering protocol for multihop wireless ad hoc networks," *Selected Areas in Communications, IEEE Journal on,* vol. 19, pp. 1201-1210, 2001.

[12] U. C. Kozat, G. Kondylis, B. Ryu, and M. K. Marina, "Virtual dynamic backbone for mobile ad hoc networks," in *Communications, 2001. ICC 2001. IEEE International Conference on*, 2001, pp. 250-255.

[13] M. R. Pearlman and Z. J. Haas, "Determining the optimal configuration for the zone routing protocol," *Selected Areas in Communications, IEEE Journal on,* vol. 17, pp. 1395-1414, 1999.

[14] A. Iwata, C.-C. Chiang, G. Pei, M. Gerla, and T.-W. Chen, "Scalable routing strategies for ad hoc wireless networks," *Selected Areas in Communications, IEEE Journal on,* vol. 17, pp. 1369-1379, 1999.

[15] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Applied statistics,* pp. 100-108, 1979.

[16] G. Aldabbagh, S. T. Bakhsh, N. Akkari, S. Tahir, H. Tabrizi, and J. Cioffi, "QoS-Aware Tethering in a Heterogeneous Wireless Network using LTE and TV White Spaces," *Computer Networks,* vol. 81, pp. 136-146, 2015.

[17] G. Aldabbagh, S. T. Bakhsh, N. Akkari, S. Tahir, S. Khan, and J. Cioffi, "Distributed dynamic load balancing in a heterogeneous network using LTE and TV white spaces," *Wireless Networks,* pp. 1-12, 2015.

[18] H. Tabrizi, G. Farhadi, J. Cioffi, and G. Aldabagh, "Coordinated Tethering over White-Spaces," 2014.

[19] N. Akkari, G. Aldabbagh, M. Nahas, and J. Cioffi, "Dynamic Clustering Protocol for coordinated tethering over cellular networks," *Journal of Network and Computer Applications,* vol. 42, pp. 92-101, 2014.

[20] A. Khalifah, N. Akkari, and G. Aldabbagh, "Dense areas femtocell deployment: Access types and challenges," in *e-Technologies and Networks for Development (ICeND), 2014 Third International Conference on*, 2014, pp. 64-69.

[21] N. Akkari, G. Aldabbagh, M. Nahas, B. Bawazeer, J. Cioffi, and H. Tabrizi, "Coordinated tethering over cellular networks: Handover scenarios and

signaling," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, pp. 2170-2174.

[22] Y. Yong, Z. Chongxun, and L. Pan, "A novel fuzzy c-means clustering algorithm for image thresholding," *Measurement Science Review,* vol. 4, pp. 11-19, 2004.

[23] J. Heer and E. H. Chi, "Identification of web user traffic composition using multi-modal clustering and information scent," in *Proc. of the Workshop on Web Mining, SIAM Conference on Data Mining*, 2001, pp. 51-58.

[24] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, pp. 881-892, 2002.

[25] S. Ghosh and S. K. Dubey, "Comparative analysis of k-means and fuzzy c-means algorithms," *International Journal of Advanced Computer Science and Applications (IJACSA),* vol. 4, 2013.

[26] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 34, pp. 1907-1916, 2004.

[27] H. Lin, "Method of image segmentation on high-resolution image and classification for land covers," in *Natural Computation, 2008. ICNC'08. Fourth International Conference on*, 2008, pp. 563-566.

[28] V. S. Rao and D. S. Vidyavathi, "Comparative Investigations and Performance Analysis of FCM and MFPCM Algorithms on Iris data," *Indian Journal of Computer Science and Engineering,* vol. 1, pp. 145-151, 2010.

[29] R. Xu and D. Wunsch, "Survey of clustering algorithms," *Neural Networks, IEEE Transactions on,* vol. 16, pp. 645-678, 2005.

Urban Mobility over Internet of Things to Smart Cities

ICWN
Carlos Henrique Rodrigues de Oliveira, Rogerio Moreira Lima Silva, Leonardo Henrique Gonsioroski Furtado da Silva
Computer Engineering - Technological Sciences Center (CCT)
State University of Maranhão (UEMA) - São Luís, Brazil
{carloshenrique, rogeriomls, gonsioroski@uema.br}

***Abstract*—This paper proposes a manner of rational use of cars in Smart Cities focused in preserving the environment and better quality of life now and to the future that represents an economical and cheaper way to increase the flow of cars on the roads compared to major infrastructure projects.**

Keywords—Urban Mobility; Smart Cities; PIR; sensor; IoT; WiMAX.

I. Introduction

Nowadays is raising the number of cars in the streets of many countries due to the several reasons but in the case of Brazil: (1) the rate of new cars incoming the streets is much higher than old cars removed of circulation, (2) it is a emergent country and the purchasing power is higher in the social classes C and D which are the majority, (3) the policy of government to control new/old cars in the streets is weak, (4) the government incentives the production of cheaper popular cars, (5) the policy of government to control cars that circulate in the streets does not work.

As the focus of this paper is to facilitate the urban mobility to better quality of life and environment issues it is worth to mention that the circulation of a lot of vehicles compromises the flow of traffic, contributes negatively to environment increasing the pollutant emissions, besides favoring accidents caused by dropping the level of traffic safety.

The next sections are organized as follows. In Section II it is presented a proposed of change. Section III presents considerations about an awareness campaign. Section IV contains a scheme of changed. A scheme of control is presented in section V. Section VI highlights the benefits. Section VII addresses the subject PIR, a passive infrared electronic sensor. Car control network topology is presented in Section VIII. Coverage prediction is presented in Section IX. Section X presents the exception of the rule. Section XI presents the car ride. Finally in Section XII are presented the conclusions.

II. Proposed of change

A. Behavior Change

This is the kind of situation that requires first a reflection if the change should not start for us. Perhaps the answer to a simple question can bring important information: how many times in a week we drive alone without any other passenger?

If the answer was five (5), this represents almost 72% of weekdays.

It will probably be a big inconvenience seek another way of transportation either public transportation or get a ride, but if we taking into account the rate cost/benefit of this action probably many people will agree that the effort pays off due to the environment issues and the quality of life.

B. Rotation

The Capital of the State of São Paulo in Brazil adopted vehicles rotation (applied for cars and trucks except buses, ambulances, fire trucks and motorcycles) that limits the access permission of the main public roads according to the final of vehicle identifications, finals 1 and 2 is forbidden to access to major public roads on Mondays, finals 3 and 4 is forbidden on Tuesdays, finals 5 and 6 is forbidden on Wednesdays, finals 7 and 8 is forbidden on Thursdays and finals 9 and 0 is forbidden on Fridays. They are released on holidays and weekends.

But what we have seen over the years was the traffic jam again especially because many families ended up buying another vehicle to avoid the prohibited day.

III. Awareness campaign

The first measure to be adopted is a public campaign to inform and clarify the population of the need to change behavior of car users and the benefits this change brings to society informing that it will be given a deadline for people to practice and adapt to these changes.

The second measure to be adopted is the notification issued by the transit regulatory agencies for users using their cars with only one passenger.

The third measure to be taken is to apply traffic violation ticket issued by the transit regulatory agencies for users using their cars with only one passenger.

IV. Scheme of changed

The main change in behavior is to people do not use their cars with only one passenger inside. But what is the minimum number of passenger required for the car owner does not receive traffic violation ticket? The answer to this question will depend on the result of the relieved traffic flow and Carbon Dioxide (CO_2) emission reduction obtained with the change requiring at least two passengers in the same car.

In case of reducing the amount of cars on public roads is not enough to reduce CO_2, increases the flow of cars and improving the quality of life of the population, the minimum amount to be adopted should increase gradually. In the limit this measure is extreme and requires the transformation of the individual transportation concept, but it is the price we have to pay if we want to have both at the same time mean of transportation and quality of life.

V. Scheme of control

After the third measure of the awareness campaign and during a certain period of time, the transit regulatory agencies will have to analyze the traffic statistics to know how many percent of all cars are used with at least two passengers. This quantity of passengers is the first measure adopted in the scheme of control.

Based in the total of cars on the streets before the third measure of the awareness campaign, it will be possible to determine how many percent of cars does not go to the street with only one passenger due to the possibility to be mulcted. The value of mulct is according to the number of passengers inside the car, only one passenger is higher than with two and so on.

VI. Benefits

Let's suppose: a) 100 cars are used with only one passenger inside; b) the measure adopted in the scheme of control resulted in at least two passengers per car and c) same path from house to work of these 100 people. These means:

i. Half the cars out of the streets;

ii. Probably half the CO_2 will not be launched in the atmosphere;

iii. The flow of traffic to and from work will decrease considerably;

iv. Probably it will be the half of time to arrive and return to and from work (crash car, traffic light broken etc, are not considered);

v. The quality of life will increase.

VII. PIR

The scheme of control will be based on counting the number of passengers inside the cars and PIR (Passive infrared) is appropriate to it.

A passive infrared sensor measures infrared light emitted from objects that generate heat, and therefore infrared radiation, in its field of view. Crystalline material at the center of a rectangle on the face of the sensor detects the infrared radiation. The sensor is actually split into two halves so as to detect not the radiation itself, but the change in condition that occurs when a target enters its field.

The term passive in this instance refers to the fact that PIR devices do not generate or radiate any energy for detection purposes. They work entirely by detecting the energy given off by other objects. It is worth to notice that PIR sensors do not detect or measure "heat" per se; instead they detect the infrared radiation emitted from an object which is different from but often associated/correlated with the object's temperature.

A. Construction

Infrared radiation enters through the front of the sensor, known as the 'sensor face'. At the core of a PIR sensor is a solid state sensor or set of sensors, made from *pyroelectric* materials which generate energy when exposed to heat. The sensor is often manufactured as part of an integrated circuit (IC) [1].

VIII. Car control network topology

The car control network topology is presented in Figure 1.

The electronics in the PIR will control a small relay. This relay completes the circuit across a pair of electrical contacts connected to a detection input zone of the two cameras. The system will be designed such that if two passengers or more are detected, the relay contact is closed—a 'normally closed' (NC) relay. If only one passenger is detected, the relay opens, triggering the cameras.

In a pole there will be a central control to register both the counting and the photos took from the two cameras. Camera1 will be responsible to take photo not only the passengers in the car rear but also the ID of the car. Camera2 will be responsible to take photo the passengers in the side view.

The central control will be fully weatherized die-cast aluminum enclosed and have a four-port switch and a CPE (Customer-Premises Equipment) radio to transmit the counting and photos to the counting Base in 250 MHz multipoint-multipoint (mesh) system. The choice of the frequency and the Effective Isotropic Radiated Power (EIRP) depend on the regulatory agency of each country.

Counting data management is responsible to analyze the data from the cars in irregular condition and send this information to the transit department of the city via Internet to take appropriate actions.

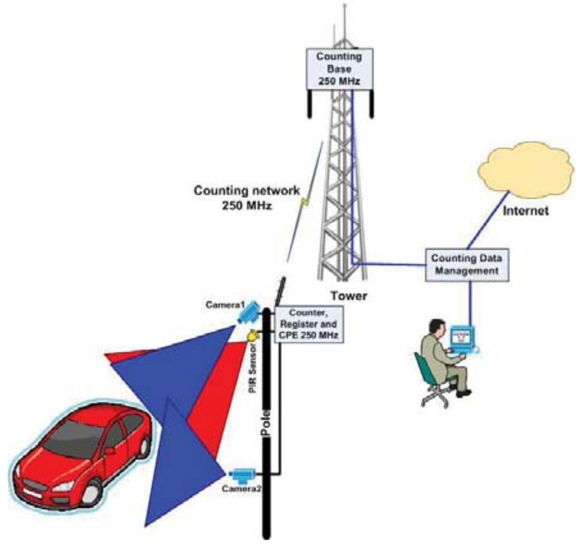


Figure 1. Car control network topology

IX. COVERAGE PREDICTION

According to the Resolution 555 of the Regulatory Agency in Brazil (Anatel) it is possible to operate in 250 MHz with EIRP more than 25 dBm to FDD (Frequency Division Duplexing) in point-multipoint and point-to-point modes or with EIRP until 25 dBm to TDD (Time Division Duplexing) in multipoint-multipoint mode. All of three modes with channel bandwidth (BW) of 25 kHz or 1.25 MHz with channel aggregation possibility resulting in until 2 channels of 6.25 MHz.

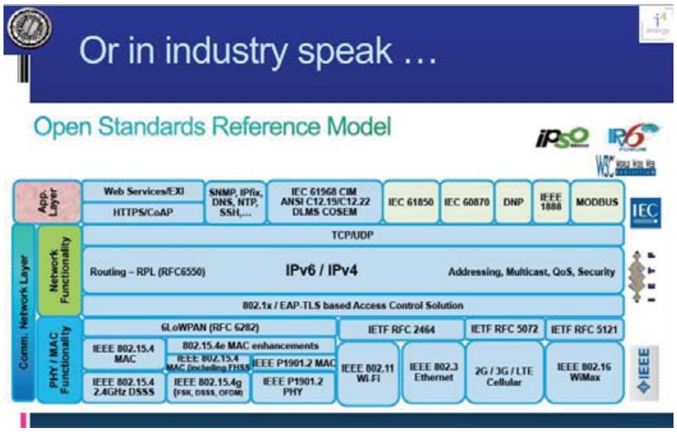
Also, according to the Anatel’s Resolution 555: a) the guard band is 17.5 MHz corresponding to 7.3 %; b) the duplex distance is 22.5 MHz corresponding to 9.4 % and c) 2 channels of 5 MHz or 4 channels of 2.5 MHz or 8 channels of 1.25 MHz give the possibility to cover a higher geographic area reusing the frequencies (considering appropriated reuse distance to mitigate co-channel interference) with potential of higher capacity due to the higher BW.

These features represent a great opportunity to operate in an optimized way with the main characteristics in terms of communication that are capacity and coverage with minimized interference because the frequency is licensed.

A hybrid solution it is a good approach with access network in multipoint-multipoint mode with IoT radios and backhaul network in point-multipoint mode with WiMAX or LTE radios.

A. *Internet of Things (IoT)*

Some suppliers in the industry already offer platform to operate based on IoT open standards reference model presented in Figure **2** that is applicable in this work.



Or in industry speak …

Open Standards Reference Model

App. Layer
Web Services/EXI
HTTPS/CoAP
SNMP, IPfix, DNS, NTP, SSH,…
IEC 61968 CIM ANSI C12.19/C12.22 DLMS COSEM
IEC 61850
IEC 60870
DNP
IEEE 1888
MODBUS

Comm. Network Layer
Network Functionality
TCP/UDP
Routing – RPL (RFC6550)
IPv6 / IPv4
Addressing, Multicast, QoS, Security
802.1x / EAP-TLS based Access Control Solution

PHY / MAC Functionality
6LoWPAN (RFC 6282)
IETF RFC 2464
IETF RFC 5072
IETF RFC 5121
IEEE 802.15.4 MAC
802.15.4e MAC enhancements
IEEE 802.15.4 MAC (including FHSS)
IEEE P1901.2 MAC
IEEE 802.15.4 2.4GHz DSSS
IEEE 802.15.4g (FSK, DSSS, OFDM)
IEEE P1901.2 PHY
IEEE 802.11 Wi-Fi
IEEE 802.3 Ethernet
2G / 3G / LTE Cellular
IEEE 802.16 WiMax

Figure 2. IoT open standards reference model [2]

1) *Car control access network*

Based on Anatel’s Resolution 555 and IoT chipset data sheet [3], there is a list of parameters as input to the access network coverage prediction as follows:

- Transmission power: 20 dBm;
- Transmission antenna gain: 6 dBi;
- Cable and connectors losses: 1 dBi;
- Receiver sensitivity (BER < 0.1%, 500 kbps, 250 kHz BW, GSFK): -97 dBm;
- Installations: Base Station (BS): 20 m in tower; CPE: 7 m in energy pole;
- Frequency: 250 MHz;
- Fast and slow fading margin: 10 dB;
- Point-to-multipoint communication mode;
- Tool of coverage prediction: Radio Mobile [5];
- Topography and morphology data base resolution: 100 m;
- Uplink and downlink balanced.

Figure 3 shows the coverage prediction results considering dense urban environment in a very difficult area to the RF (Radio Frequency) propagation that is the central area of São Paulo Capital in Brazil.

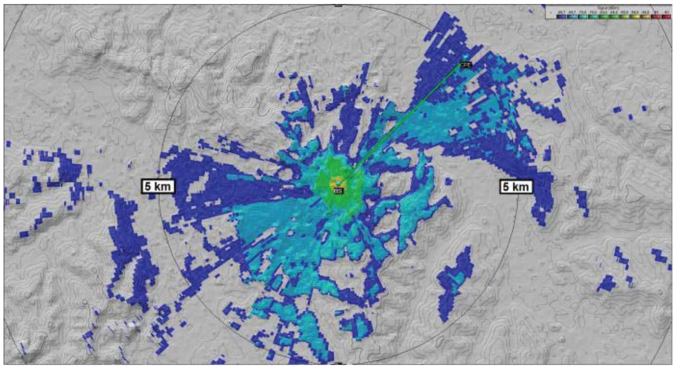


Figure 3. Access network coverage prediction

The coverage map in Figure 3 shows that RF propagation is very limited according to the environment. In some azimuths the signal (from BS to CPE and vice versa) reaches more than 5 km but in others less than 500 m. This is a typical scenario where the mesh network is appropriated due to the massive presence of obstacles and the possibility of signal repetition.

B. *WiMAX (Worldwide Interoperability for Microwave Access)*

Added to the WiMAX Forum® System Profile Requirements for Smart Grid Applications announced in January 24 2013 [4], the Anatel’s Resolution 555 represents a good opportunity to WiMAX Forum consider a new profile to Smart Cities in 250 MHz especially because Brazil is an emergent country with huge territorial extension and opportunities, it is possible to operate with channel bandwidth initiating in 1.25 MHz aligned with IEEE 802.16e WiMAX standard and due to the fact that other countries may be under the same regulatory conditions.

1) *Car control backhaul network*

Based on Anatel’s Resolution 555 and typical data of WiMAX radios, there is a list of parameters as input to the backhaul network coverage prediction as follows:

- Transmission power: 37 dBm;
- Transmission antenna gain: 6 dBi;
- Cable and connectors losses: 1 dBi;
- Receiver sensitivity (BER 10^{-5}, 8 Mbps, 5 MHz BW, 16QAM-1/2): - 89.7 dBm;
- Installations: Base Station (BS): 20 m in Tower; CPE: 7 m in energy pole;
- Frequency: 250 MHz;
- Fast and slow fading margin: 10 dB;
- Point-to-multipoint communication mode;
- Tool of coverage prediction: Radio Mobile [5];
- Topography and morphology data base resolution: 100 m;
- Uplink and downlink balanced.

Figure 4 shows the coverage prediction results considering dense urban environment in a very difficult area to the RF (Radio Frequency) propagation that is the central area of São Paulo Capital in Brazil.

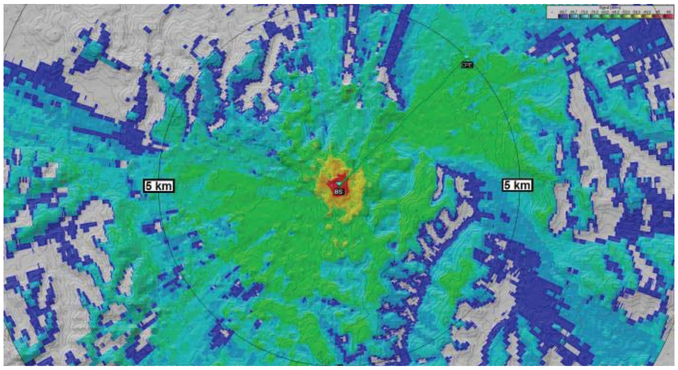


Figure 4. Backhaul network coverage prediction

The coverage map in Figure 4 shows that RF propagation is very limited according to the environment. In some azimuths the signal (from BS to CPE and vice versa) reaches more than 10 km with better coverage than shown in Figure 3 (and higher capacity as well due to the higher BW) appropriated to the car control backhaul network share this resource among the traffic of several CPEs and other traffic sources as video monitoring.

X. EXCEPTION

The exception of the car control rule is applied to the taxi cabs. The reason of this is first because most of the time the taxi cabs are occupied with more than one passenger and second it is a way in gain livelihood, driver is a profession in many countries.

Counting data management shown in Figure 1 is responsible to analyze the ID of the taxi cabs to avoid issuing traffic violation ticket in an irregular way.

A. *Trick*

Even someone tries to use plastic doll to simulate a passenger inside the car, PIR sensor will not sense the heat energy in the form of radiation. The trick will turn against who made. The “crime” will not pay.

XI. CAR RIDE

Currently some applications are freely available on the Internet and can help to facilitate the sharing of the car ride. The car ride is the car of the day to be used in a range of cars available for a group of passengers. The logistic is facilitated when the passengers live near or work close to each other or in the same company.

Some apps allow passengers to share a ride. The Sidecar [6] is one of them. Just checking which are the place and destination and the registered user as nearest driver is alerted.

XII. DEDICATED DETECTION ALGORITHM

In the cases of the presence of domestic animals inside the car, probably they will be detected but should not be counted as passengers.

PIR sensors have quite unique sensing model, thus, it will be necessary to develop dedicated detection algorithm [7] to deal with these cases.

XIII. CONCLUSIONS

Change is necessary and the first change is in our own behavior. It is easier and more practical to use a car thinking individually but if we consider the practical results with a lot of cars congesting the streets, damaging the environment conditions and decreasing our quality of life, the price to pay is low compared with the benefits that this initiative of car control can bring to our and the next generations.

It worths to notice that this represents an economical and cheaper way to increase the flow of cars on the roads compared to major infrastructure projects.

REFERENCES

[1] Piero Zappi, Elisabetta Farella, and Luca Benini, "Tracking motion direction and distance with Pyroelectric InfraRed Sensors," *IEEE Sensor Journal Class Files*, 2008

[2] http://postscapes.com/internet-of-things-protocols

[3] http://www.silabs.com/Support%20Documents/TechnicalDocs/Si4464-63-61-60.pdf

[4] http://www.wimaxforum.org/press-release/wimax-forum-publishes-smart-grid-utility-requirements-document-for-wigrid-networks

[5] http://www.cplus.org/rmw/english1.html

[6] http://www.side.cr/

[7] Z. Zhang, X. Gao, J. Biswas, J.K. Wu, Moving targets detection and localization in passive infrared sensor networks, in: Proceedings of the 10th International Conference on Information Fusion, Quebec, 2007, pp. 1-6.

A Distributed Parallel Reconstruction Method for Mobile Terminals

A. Z. Liu[1], B. Co-T.T. Meng[2], and C. Co-X. Liu[2], Co-Y.Y. Jiang[3]

College of computer science &engineering, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, P. R. China

Abstract - *With the rapid development of mobile Internet technology and real-time reconstruction technology, based on the multi-view of mobile terminal oriented 3D display provides remote interactive model reconstruction. The core of the algorithm is based on multiple images as input, and by calculating it generates sparse 3D point cloud through the expansion of that could generate 3D point cloud, and through the surface reconstruction it can accomplish the 3D model. But the high computing complexity and large data handling scale affect the quality and real-time performance of 3D demonstration. Against the above shortcomings, a distributed parallel reconstruction method for mobile terminals is proposed, which reconstructs model both in the server and in the client. The server uses levels of detail technology to control the scene's complexity and generates initial reconstruction frames. The client uses image-based reconstruction technology to re-render the image, which can improve reconstruction quality. Experiments show that the method improves reconstruction speed, reduces the transmitted data size, and improves the image quality.*

Keyword: Distributed reconstruction; Levels of Detail; Image based reconstruction; CUDA

1. Introduction

With the continuous development of virtual reality technology, the networked 3D demonstration shows highly realistic stereoscopic image by using real-time reconstruction and interaction of 3D models, which is totally different with traditional information communication methods based on text or image. However, it is a huge amount of work that reconstructing the 3D model with the associated 3D modeling software manually. At the same time, the cost of scanning equipment is expensive, so it is a hot research focus in the field of computer vision that how easy to obtain 3D model of the object from the real world[1]. Meanwhile, the further development of mobile intelligent terminal results in the transition of user's terminal equipment from traditional personal computers to mobile phones, tablets, etc., which provides preconditions for the expansion of 3D display in the field of mobile applications.

Faced with increasingly reconstruction quality and real-time interactive requirement from mobile users, there are some shortcomings of traditional geometry-based reconstruction techniques for complex three-dimensional models[2,3]. On the one hand, for the scene composed by complex three-dimensional model, the larger data complexity causes more graphic hardware requirements of mobile terminal, which impacts the reconstruction real-time capacity.

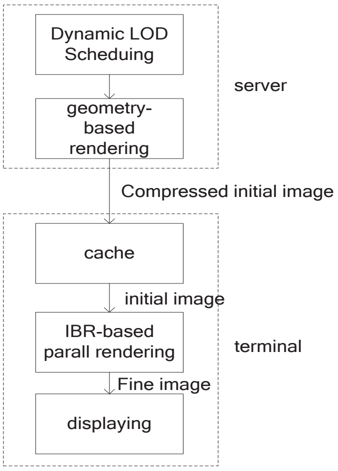


Fig.1 Overall schematic diagram of the distributed parallel reconstruction method

On the other hand, the mobile network bandwidth may not meet the speed requirement of the stereoscopic image display and interaction. With the development of massively parallel processing technology, the use of distributed parallel processing technology is an effective way to solve this problem.

In this paper a distributed parallel reconstruction method for mobile terminal is proposed to realize distributed reconstruction both in server side and terminal side. CUDA-based parallel reconstruction data processing is used in the method which achieves the goal of real-time complex three-dimensional model displaying on mobile terminals and reduction of network bandwidth consumption. The overall process of the method is shown in Figure 1, in which the server side generates LOD model by CUDA parallel computing and dynamic form the corresponding initial reconstruction image, also image-based reconstruction is used to realize the compression and transmission of the image, while the terminal side re-renders the image to improve the image fineness by CUDA-based IBR techniques.

2. Server side primary reconstruction

To improve the server's ability of reconstruction complex three-dimensional model, a primary reconstruction strategy on the server side is proposed to ensure the real-time reconstruction capacity. The server pre-generates LOD simplified model of different resolutions by CUDA parallel computing. And in the reconstruction process, the appropriate LOD model is reconstructed based on the distance between the viewpoint and the model as well as the reconstruction frame rate, to improve the overall computing efficiency.

2.1 Parallel simplification of LOD model based on CUDA

CUDA is known as a parallel computing platform and programming model implemented by GPU. This paper uses CUDA to simplify model by means of edge collapse calculation. Edge collapse refers to choose two connected vertices and replace them with a single vertex, and all the vertexes connected to the two vertices will re-connect to the new vertex to maintain the triangle grid appearance. In this paper, we select one of the two connected vertices as the new vertex. Garland Quadratic Error Metrics (QEM) is used to compute the edge collapsing cost, which can efficiently preserve the features on the surface.

In the simplification process, LOD model generation introduces the concept of vertices importance as the trade-off of the sequence of edge collapse operation. Vertex importance reflects the geometry importance degree of triangle mesh. It is discussed in [4] that in the neighborhood of the vertex in the mesh, if the steeper the vertex is, the greater the impact on the mesh geometry is; while the sparser the vertex is, the bigger the vertex importance value is as well. So the vertex importance can be summarized as follows:

$$Q(v_j) = K_j \times \bar{l} = \frac{\sum_{v_i \in neiverts(v_j)} c_{ji}}{m} \times \bar{l} \qquad (1)$$

Call c_{ji} the cosine of the angle between the normal vectors v_j and v_i . v_i is the element in the vertices collection in which the vertices are connected with v_j , m is the total number of the set, and $\bar{l}$ refers to the average length of the associated edges. So K_s , the average cosine value of the vector angles, can be regarded as the curvature of v_j , and therefore $Q(v_j)$ can reflect the geometry importance of v_j .

We use Octree to divide the original mesh into several independent sub meshes. A generation method of LOD model is proposed by using CUDA to parallel simplify those sub meshes based on edge collapse operation. The process is shown in Figure 2.

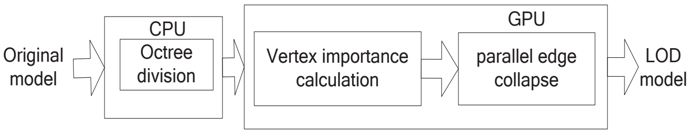


Fig.2 LOD parallel simplification process

In order to facilitate the subsequent edge collapse calculation, we storage all the vertex info in the sub mesh into an array *vtable*, which is defined as follows:

```
struct vernode
{
  float x,y,z;    //vertex three-dimensional coordinates
  long ver_importance;
  int foldingindex;    // the folding vertex index
} ;
vernode Vtable[vernum];
```

As is shown in Figure 2, the original model is partitioned by octree spatial decomposition on CPU at first, which regards the entire model bounding box as the octree root node. The bounding box is split into 8 sub cubes, in accordance with intermediate section in the direction of the three-dimensional coordinates. Correspondingly, the model is divided into 8 sub meshes.. Recursively decompose the sub cubes until the amount of triangle facets in the sub mesh is less than a given threshold. Finally the model is organized as an octree structure.

In GPU computing stage, parallel threads are created to calculate each vertex importance according to (1). After the vertex importance calculation is finished, the vertex is sorted by its importance in each sub mesh and the sequence number is stored in *vtable*. During parallel edge collapse operation, the sub meshes is simplified concurrently. For each sub mesh, edge collapse is implemented by vertex importance ascending sequence. For each edge collapse operation, the edge (v_i, v_j) that contains vertex v_i which owns the lowest importance is replaced by the vertex v_j . For the edges that all contains vertex v_i , we calculate the edge collapse cost by QEM to select the appropriate one. Recursively simplify the sub mesh until the number of remaining vertices meets the requirement.

The parallel simplification algorithm of LOD model is as follows:

Algorithm 1GPU parallel LOD generating

```
Input:
mesh:the original model triangle mesh
Output:
simplifiedmesh:the simplified mesh
submeshes←CPU_OctreeMeshDivde(mesh) ;/*submeshes:array of divided submesh */
MemcpySync(submeshes,host->device);/*Load submeshes to GPU device memory*/
for i=0 to $NUM_{submeshes}$-1parallel do
    Ver_importance←verCalculating_kernel(submeshes);
    __syncthreads();      /*threads synchronize */
    Vtable←importanceSort_kernel(Ver_importance);
    __syncthreads();      /*threads synchronize */
    simplifiedmesh←edgeCollapse_kernel(submeshes,Vtable);
end for
MemcpySync(simplifiedmesh,device->host);
```

2.2 Dynamic LOD Scheduling stagey

While the server reconstruction, it should select the appropriate resolution level of LOD model based on the distance between the viewpoint and model in advance [5, 6]. Shown in Figure 3, while the distance is d1,d2,…,dn, the corresponding LOD level is LOD1,LOD2,…,LODn. The server adjusts LOD model according to the distance in order to keep its reconstruction efficiency. Also the last frame reconstruction time cost is took into consideration. If the frame rate is lower than the real-time requirement, a more simplified LOD model is selected until the frame rate meets the requirement. After the frame is reconstructed, the server will transmit the result in form of image to the terminal for the further reconstruction and displaying.

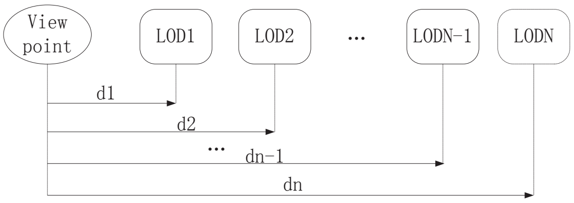


Fig.3 Viewpoint distance-dependent LOD strategy

3. Terminal side secondary reconstruction

To further reduce the server computing load and the amount of data transmitted between the server and the terminal, the terminal uses CUDA to refine initial image from the server and utilizes the difference between geometry-based reconstructed image and reconstructed image via IBR[7].

3.1 IBR-based compression and secondary reconstruction algorithm

As shown in Figure 4, call L the generated view of the model corresponding to position V of the viewpoint. So $L(x)$ $(x \in Q = [0,W] \times [0,H])$ can be seen as a two-dimensional array of pixels, where every pixel $X = [x, y, z]$ refers to the 3D position in the reference system of V . When the viewpoint changes into V' , the generated view is $L'(x')$ $(x' \in Q' = [0,W'] \times [0,H'])$. In this situation, the pixels X' in the reference system of V' can be divided into two categories. One can be seen as the transformed pixels in view V , the other are those new pixels first into the screen space. For the former one, pixels can be calculated by the three-dimensional transformation theory as follow:

$$X' = [x', y', z(x', y')]^T = T(X) = T(x, y, z(x, y)) \quad (2)$$

T is a suitable 3D projective transformation obtainable by matrix T_g in homogeneous coordinate. Denote $\hat{L}'(x')$ the view with respect to V' via IBR procedure, in order to distinguish with $L'(x')$ the view reconstructed from the 3D model with respect to V' .

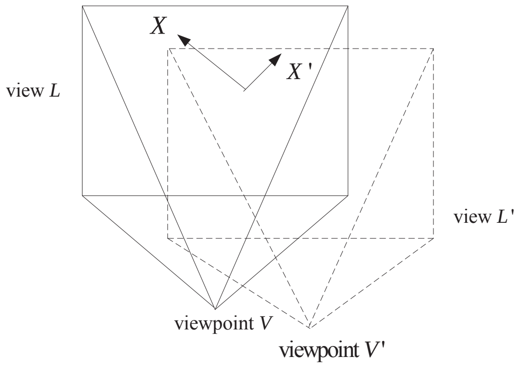


Fig.4 Relationship of pixels between view V and V'

For the pixels in $\hat{L}'(x')$, one may obtain:

$$x' = t(x), \quad x \in Q \tag{3}$$

The two pixels set $I_1 = Q' \cap t(Q)$ and $I_2 = Q' - I_1$ in view $\hat{L}'(x')$ can be written as:

$$\hat{L}'(x') = \begin{cases} L(t^{-1}(x')) & x' \in I_1 \\ 0 & x' \in I_2 \end{cases} \tag{4}$$

So the only difference between $L'(x')$ and $\hat{L}'(x')$ are the pixels in I_2. To reduce the amount of data transferred between the server and the terminal, the server only need to transmit the correction data (pixels in I_2) to the terminal instead of the whole view $L'(x')$, if the terminal can calculate $\hat{L}'(x')$ itself. $E(x')$ can be written as:

$$E(x') = L'(x')\chi(x'), \chi(x') = \begin{cases} 1 & x' \in I_1 \\ 0 & x' \in I_2 \end{cases} \tag{5}$$

During the secondary reconstruction process in the terminal, the server pre-renders fine reference view L_{high} with respect to V based on high-resolution LOD model selected by the dynamic LOD stagey. For views L' with respect to subsequent viewpoints V', the server replaces with a relative low-resolution LOD model. The view L' generate by the terminal with respect to V' via IBR procedure can be also written as:

$$\hat{L}_{high}'(x) = \begin{cases} L_{high}(t^{-1}(x)) & x \in I_1' \\ 0 & x \in I_2' \end{cases} \tag{6}$$

As can be seen, both L' and $\hat{L}_{high}'$ are the descriptions of the same original model. The difference is that $\hat{L}_{high}'$ does not contain the new pixels which are put into the screen space by the viewpoint transformation, but for pixels in I_2', the view $\hat{L}_{high}'$ contains more model details of L'.So the terminal can re-render the view as follow:

$$L'(x) = \begin{cases} \hat{L}_{high}'(x) & x \in I_1' \\ L'(x) & x \in I_2' \end{cases} = \begin{cases} L_{high}(t^{-1}(x)) & x \in I_1' \\ \hat{L}'(x) + E(x) & x \in I_2' \end{cases} \tag{7}$$

3.2 Parallel secondary reconstruction in the terminal based on CUDA

Since the parallel secondary reconstruction in the terminal is based on image-based reconstruction, the standard practice is to reset the reference frame of every p frames to reduce the prediction error[8]. A scheme of principle for reconstruction p views $(L_1, L_2, ..., L_p)$ of a 3D model is as follows:

(1) In reconstruction preprocessing stage, the server uses GPU to implement parallel LOD model generation.

(2) L_1 is set as the reference frame view, and the server reconstructs a relative high-resolution simplified model selected by dynamic LOD stagey and sends the fine view L_1 to the terminal.

(3) The server replaces the model with a lower-resolution one, computes and sends L_2 to the terminal.

(4) For frame view $L_i, 2 < i \le p$

(a) At both, server and terminal compute

$$\hat{L}_i(x) = \begin{cases} L_2(t^{-1}(x)) & x \in I_1 \\ 0 & x \in I_2 \end{cases}$$

(b) At server's side,

compute $E_i(x) = L_i(x)\chi_i(x)$ and send $E_i(x)$ to the terminal

(c)At terminal's side, compute

$$L_i(x) = \hat{L}_i(x) + E_i(x);$$

(5) At terminal's side, update

$$L_i'(x) = \begin{cases} L_1(t^{-1}(x)) & x \in I_1' \\ L_i(x) & x \in I_2' \end{cases}$$

CUDA-based parallel reconstruction in the terminal is supposed to send L_1 and L_2 to GPU memory at first, and then generate parallel GPU threads for reconstruction the corresponding pixels calculation[9], which is described as follows:

Algorithm 2GPU parallel secondary reconstruction

Input:

frameImage:the ordinary frame viewL_i from the server side ;

refImage: the reference frame viewL_1;

IBRimage: the ordinary frame viewL_2;

```
size: the size of frameImage;
Output:
Sec_frameImage: the secondary reconstructed view
MemcpySync(refImage,host->device);/*Load refImage to GPU device memory*/
MemcpySync(IBRimage,host->device);/*Load IBRimage to GPU device memory*/
for i=0 tosize parallel do
    local_Image←IBRcalculating_kernel(IBRimage);/* calculating the IBR view on GPU*/
end for
frameImage←receive();    /*get frameImage difference $E_i(x)$ from the network */
MemcpySync(frameImage,host->device);
for i=0 tosize parallel do
    local_Image←imageCorrection_kernel (IBRimage, frameImage);
    Sec_frameImage←secondReconstruction_kernel(local_Image, refImage);
end for
MemcpySync(Sec_frameImage,device->host);
```

4. Performance analysis

For the tests reported in this paper, the server node was equipped with CPU: Intel(R) Xeon CPU E5504 2.00GHz, GPU: Nvidia tesla S2050 and the terminal application was run on an Nvidia Tegra TK1 equipped with Nvidia Kepler GPU by OpenSceneGraph platform. Figure 5 is the example of the server side initial frame view and the terminal side secondary frame view.

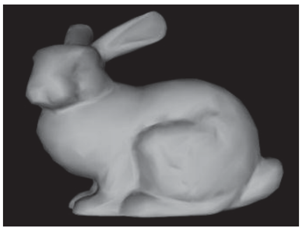


Fig.5.1 (a) Primary reconstruction in server

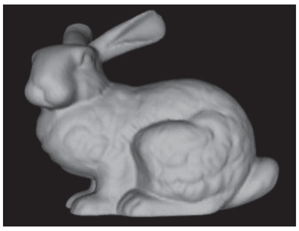


Fig.5.2 (b) Secondary reconstruction in terminal

To compare the calculating efficiency, GPU and CPU are used to simplify the model and image-based secondary reconstruction. The time cost is shown in Figure 6. We can see that with the increasing size of data processing, GPU-based parallel computing can effectively reduce the time consuming and improve the real-time capacity.

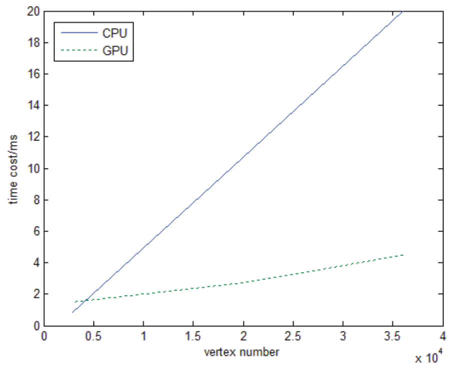


Fig.6.1 (a) time of LOD generating

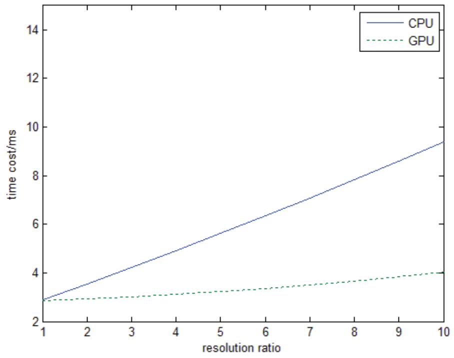


Fig.6.2 (b) time of Secondary reconstruction

When the view resolution is 320*240, the frame view data size within a certain period of time is shown in Figure 7, where each cylindrical size represents the total volume of the frame view, and the marked region means the size of image component. From the result, we can see the IBR-based image compression can effectively reduce the data volume and the bandwidth requirements.

Experiments show that the distributed parallel reconstruction method can guarantee the view reconstruction quality, improve real-time capacity as well as reduce the data transmission.

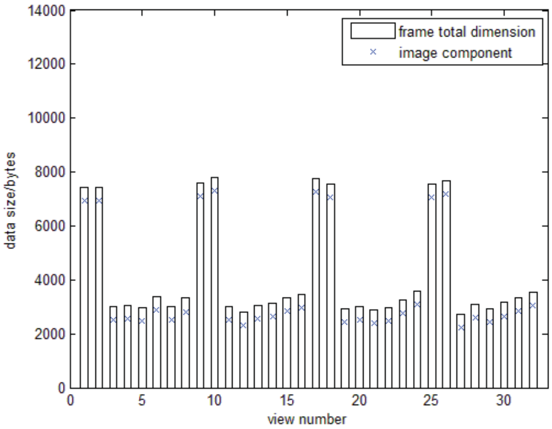


Fig.7 Server side frame data transmission amount

5. Conclusion

This paper describes a distributed parallel reconstruction method for mobile terminals, which implements the distributed reconstruction on both server and terminal, as well as parallel reconstruction data processing by GPU. The server dynamically schedules the LOD model to realize real-time reconstruction, combined with the image-based reconstruction to compress the frame view. While the terminal further re-renders the view to improve the view fineness. Experiments show that this method can improve the overall reconstruction efficiency, reduce the bandwidth requirements and improve the reconstruction speed.

6. References

[1] Zhong M L, The Image-based High-resolution Representation Technique of Three Dimensional Model[D]. Zhejiang University,2006.

[2] Li M S. Research on Virtual View Rendering Method Based on Depth Image[D].Shandong University, 2012.

[3] Zheng L P, Chen B,Wang P W, et al.Remote Visualization Based on Distributed Rendering Framwork[J].Journal of Computer Research and Development,2012,49(7) :1438-1449.

[4] Li W Q,Hong Y X,Wu H Z.Real-time Algorithm for Feature-based LOD Models Generation[J].Journal of system simulation, 2005, 17 (2) :429-431.

[5] Li J, Wu H Y, Yang C W, et al. Visualizing Dynamic Geosciences Phenomena Using an Octree-based View-dependent LOD Strategy within Virtual Globes[J].Computers & geosciences, 2011, 37(9) : 1295-1302.

[6] Zhang G, Liu X M. Similar Curvature Edge Collapse Simplification Based on Octree[j]. Application research of computers, 2010,05 : 1955-1958.

[7] Gao L J. Research on Virtual View Rendering Method Based on Depth Image In 3DTV System[D]. Shandong University, 2013.

[8] Bernardini R, Cortelazzo G M, Tormen G. IBR-based Compression for Remote Visualization[C].3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on.IEEE,2002 : 513-519.

[9] Liu Z, Hao D N and Mei X D. Based on the parallel rendering algorithm in CUDA[J]. Journal of Image and Graphics,2013,18(11):1457~1461.

SESSION

NETWORKS AND SYSTEMS: NOVEL APPLICATIONS AND ALGORITHMS, SYSTEMS + SENSOR TECHNOLOGIES

Chair(s)

TBA

PAPR Reduction of Coded OFDM Signals

Pedro Bento*†, Marko Beko§¶, João Nunes*†, Marco Gomes*†, Rui Dinis*‡ and Vitor Silva*†
*Instituto de Telecomunicações (IT), Portugal
†Department of Electrical and Computer Engeneering, University of Coimbra, 3030-290 Coimbra, Portugal
‡FCT-UNL, 2829-516 Caparica, Portugal
§Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal
¶UNINOVA – Campus FCT/UNL, Caparica, Portugal

***Abstract*—A wide range of techniques have been proposed to reduce the Peak-to-Average Power Ratio (PAPR) of Orthogonal Frequency Division Multiplexing (OFDM) signals with different trade-offs between PAPR reduction and Bit Error Rate (BER) performance degradation. Usually these techniques are studied in an uncoded context and over an ideal Additive White Gaussian Noise (AWGN) channel, and a compromise between PAPR and BER degradation is achieved depending on the adopted constellation.**

In this paper, we go further and compare several PAPR reduction techniques for Quadrature Amplitude Modulation (QAM) constellations under a typical coded OFDM scenario with time-dispersive channels. Our results show that low complexity techniques as clipping can have a good trade-off between PAPR reduction and BER degradation close to the one of the best techniques analyzed, even in a scenario with multipath time-dispersive channel and for relatively large QAM constellations. Once that is shown that clipping is a good choice for PAPR reduction, it opens a subject in the study of PAPR reduction techniques in order to improve its behavior. We also consider the PAPR reduction as a constrained optimization problem which, although too complex to solve, provides an approximate bound on the achievable PAPR.

***Index Terms*—OFDM, PAPR, Clipping and Filtering (CF), BER**

I. Introduction

A key feature of a communication system is to make efficient use of available power for data transmission. One of the most spectrally-efficient modulations is the Orthogonal Frequency Division Multiplexing (OFDM) scheme, which is widely used in most broadband wireless systems [1], [2] and currently being recommended for future 5G systems [3], [4]. However, it is widely recognized that one of the main drawbacks of OFDM signals is their high envelope fluctuations and so a high Peak-to-Average Power Ratio (PAPR) which lead to amplification difficulties [1]. A wide variety of techniques was proposed to reduce the envelope fluctuations of OFDM signals. These include multiple signal representations such as with Partial Transmit Sequences (PTS) and Selective Mapping (SLM) techniques [5]–[8], clipping techniques [7]–[11], Tone Reservation (TR) techniques [7], [8], [12]–[14], among many other techniques.

PTS and SLM techniques have the advantage of not leading to performance degradation, since the transmitted signals are not distorted (actually, these techniques might require some side information, and transmitting it might lead to a slight spectral and power efficiency decrease). However, the computational complexity increases significantly when we want to reduce substantially the envelope fluctuations of the transmitted signals. On the other hand, clipping techniques are very simple and flexible, although the signal distortion might lead to significant performance degradation. TR techniques can have limited performance degradation since only the reserved tones are modified (in that case, the performance degradation is essentially due to the power spent on the reserved tones), unless we also modify data subcarriers. They can be particularly interesting for large constellations. However, since only a fraction of the subcarriers is effectively used for data transmission we have some degradation in the spectral efficiency, which is higher if we want to have signals with very low envelope fluctuations. Moreover, it is not always easy to obtain the optimum symbols for the reserved tones.

The achievable PAPR with a given technique is hard to obtain, although we can formulate the PAPR reduction as an optimization problem and employ powerful math tools to solve it [13], [15], for example by minimizing the PAPR conditioned to a given Error Vector Magnitude (EVM) or minimizing the EVM conditioned to a given PAPR target. Although this usually leads to non-practical PAPR reducing methods (we need to solve a complex optimization problem for each transmitted block), this has the advantage of providing some reference on the achievable PAPR.

The study of PAPR reducing techniques that lead to performance degradation is usually performed for an uncoded transmission in an ideal Additive White Gaussian Noise (AWGN) channel, which makes sense since the PAPR is essentially a transmitter problem, regardless of the channel. In fact, for small constellations, e.g., Binary Phase Shift Keying (BPSK) or Quadrature Phase Shift Keying (QPSK), the signal distortion usually is not high enough to lead to significant Bit Error Rate (BER) degradation. However, when we consider larger constellations like 16-Quadrature Amplitude Modulation (QAM) or 64-QAM this is no longer true. In that case, we should consider both the channel effects and the adopted channel coding scheme.

Some previous works [16], [17] have considered coded

This work was supported in part by the projects UID/EEA/50008/2013 IT (pluriannual funding and GLANCES project), EnAcoMIMICo EXPL/EEI-TEL/2408/2013, ADIN (PTDC/EEI-TEL/2990/2012), DISRUPTIVE EXCL/EEI-ELC/0261/2012, GALNC EXPL/EEI-TEL/1582/2013 and CoPWIN PTDC/EEI-TEL/1417/2012.

OFDM however, they only have considered a single PAPR reducing technique over an ideal AWGN channel. Therefore, in this paper we compare several PAPR reducing techniques, when multipath time-dispersive channels are considered and appropriate channel coding schemes are employed. We focus our study in the most well-know PAPR reducing techniques, such as SLM, PTS and Clipping and Filtering (CF), as well as the constrained optimized PAPR reduction as to obtain a reference on the achievable PAPR.

II. PAPR Reduction Overview

OFDM signals have high envelope fluctuations which brings some difficulties in amplification, therefore it is necessary to keep the envelope fluctuations of the signal below a certain level. In order to do that, a measure of the envelope fluctuations has to be used and the most widely used is PAPR.

The PAPR is defined as the ratio of the peak power of the signal to its average power [8]. Mathematically, the PAPR of m^{th} OFDM symbol is written in time domain as:

$$PAPR\left(x_m\left[n\right]\right) = \frac{\max\limits_{0\leq n\leq \ell N-1} \left|x_m\left[n\right]\right|^2}{E\left[\left|x_m\left[n\right]\right|^2\right]} \ , \tag{1}$$

where $E\left[.\right]$ is the expectation operator, N the number of subcarriers and ℓ is the oversampling factor that it must be at least 4 in order to ensure that the difference between continuous time and discrete time PAPR is negligible as it is shown in [18]. Then if PAPR is limited to a certain level, envelope fluctuations will be limited too and it is possible to achieve better results in amplification. Therefore lots of research has been made to reduce PAPR and in this section we will briefly present the concepts of some PAPR reduction techniques: CF, PTS, SLM, TR and constrained optimized PAPR.

Let $\mathbf{X} \in \mathbb{C}^N$ be an original OFDM frequency-domain symbol and $\widetilde{\mathbf{X}} \in \mathbb{C}^N$ the PAPR optimized frequency-domain symbol using N subcarriers. The original, $\mathbf{x}$, and the optimized, $\widetilde{\mathbf{x}}$, OFDM time-domain symbols are obtained by Inverse Fast Fourier Transform (IFFT) with ℓ-times oversampling, i.e., $\mathbf{x} = \text{IFFT}_{\ell N}(\mathbf{X}) = \mathbf{AX}$ and $\widetilde{\mathbf{x}} = \text{IFFT}_{\ell N}(\widetilde{\mathbf{X}}) = \mathbf{A}\widetilde{\mathbf{X}}$, where the matrix $\mathbf{A} \in \mathbb{C}^{\ell N\times N}$ is the first N columns of the corresponding Inverse Discrete Fourier Transform (IDFT) matrix.

The OFDM subcarriers are usually divided into three disjoint sets: data subcarriers, free subcarriers and pilot subcarriers, with cardinalities d_{sub}, f_{sub} and p_{sub}, respectively, so that $d_{sub} + f_{sub} + p_{sub} = N$, where N is the total number of subcarriers. For simplicity, pilot subcarriers will not be considered here, although the results can be easily generalized to systems with pilot subcarriers. In addition, in order to identify on a symbol $\widetilde{\mathbf{X}}$ the data subcarriers, it will be use a diagonal matrix $\mathbf{S} \in \mathbb{R}^{N\times N}$ with $\mathbf{S}_{kk} = 1$ when the k^{th} subcarrier is reserved for data transmission and $\mathbf{S}_{kk} = 0$ otherwise, i.e. data subcarriers can be obtained by the product $\mathbf{S}\widetilde{\mathbf{X}}$.

A. Clipping and Filtering

CF [7]–[11] is the simplest way to reduce PAPR. This technique clips every sample of the signal above a certain defined level. Therefore, the clipped version of $\mathbf{x}$ can be expressed as

$$\widetilde{x}_k = \begin{cases} x_k & \text{if } |x_k| < A_{CL} \\ \frac{x_k}{|x_k|}.A_{CL} & \text{otherwise} \end{cases} \ , \tag{2}$$

where A_{CL} is the amplitude of clipping level. However, the Clipping Ratio (CR) defined as the amplitude of clipping level, A_{CL}, normalized by the Root Mean Square (RMS) value of OFDM signal, $\sigma = \sqrt{||\mathbf{x}||^2}$ is more suitable to use since it adapts the clipping level from symbol to symbol.

Although CF is a simple technique, it causes signal in-band distortion which results in a degradation in BER performance and it also causes out-of-band radiation. In order to reduce the out-of-band radiation, filtering is used but, unfortunately, this leads to a peak regrowth and the obtained signal may exceed the desired clipping level [8]. Therefore, an iterative process was proposed in [10], [19], [20], designed by Repeated Clipping and Filtering (RCF), that may require a few number of iterations that should be done in order to obtain the desired PAPR reduction. However, this process increases the computational complexity.

B. Selective Mapping

SLM technique [5], [7], [8] is a symbol scrambling technique based on the fact that an OFDM symbol can be scrambled by a certain number of different sequences and, then, the symbol with the lowest PAPR is chosen to transmit. In more detail, the transmitter generates U phase sequences (vectors) expressed as

$$\mathbf{P}^u = \left[p_0^u, p_1^u, ..., p_{N-1}^u\right]^T, u = 1, 2, ..., U, \tag{3}$$

with the same length of the original OFDM symbol $\mathbf{X}$, where each element of the phase vector, p_k^u, $k = 0, 1, 2, ..., N-1$, is randomly selected from a finite set of phase factors, e.g. $\{-1, 1, -j, j\}$. Each of these vectors, $\mathbf{P}^u$, is then point-wised multiplied by $\mathbf{X}$ and the resultant symbol with the lowest PAPR, expressed as

$$\widetilde{\mathbf{X}} = \left[X_0 p_0^u, X_1 p_1^u, ..., X_{N-1} p_{N-1}^u\right]^T, \tag{4}$$

is chosen. To ensure that the unmodified symbol is in the set of choices, $\mathbf{P}^1$ is the all-one vector.

When SLM is performed, the U phase sequences are stored at both the transmitter and the receiver. To perfectly recover the signal at the receiver, the transmitter must send side information to the receiver telling which phase sequence is used. If side information is incorrectly detected, the whole data information will be lost. Although side information is so important, its use leads to losses in spectral and power efficiency reducing the data transmission rate, mainly, because it needs a strong protection [7]. However, in [21] it is proposed a SLM technique without explicit side information.

For each OFDM symbol, it is necessary to execute U IFFT operations and $\lceil \log_2 U\rceil$ bits must be passed as side information (where $\lceil y \rceil$ denotes the lowest integer greater than y). Therefore, this technique may not be feasible, due to its computational complexity, which increases when N is large, and mainly, when U is increased in order to achieve a substantial PAPR reduction.

C. Partial Transmit Sequence

PTS technique [5]–[8] as SLM is a scrambling technique. The difference between them is that the first only scrambles groups of subcarriers, while the latter scrambles independently all subcarriers [1]. Therefore, SLM produces signals that are asymptotically independent, while signals generated by PTS are interdependent [22]. With this property, PTS can avoid some of the complexity of the several full IFFT operations, which results in an advantage over SLM [5].

PTS technique partitions each OFDM symbol into V disjoint subblocks with equal size as follows:

$$\mathbf{X}_v = \left[X_{v,1}, X_{v,2}, ..., X_{v,N-1}\right]^T, v = 1, ..., V, \tag{5}$$

and

$$\mathbf{X} = \sum_{v=1}^{V} \mathbf{X}_v. \tag{6}$$

This partition can be one of three kinds: adjacent, interleaved, and pseudo-random. Among these, the latter has been found the one that provides the best performance [23]. In either partition method the subcarriers in each subblock are independently rotated by a phase factor, b^v, in order to minimize the PAPR of the combined signal. The phase factor is selected from a finite set of phase factors, e.g. $\{-1, 1, -j, j\}$, with length W. Hence, the optimized time-domain OFDM symbol is

$$\widetilde{\mathbf{x}} = \sum_{v=1}^{V} b^v \mathbf{x}^v. \tag{7}$$

The PAPR reduction given by this technique depends on the number of subblocks, V, the number of allowed phase factors, W, and the subblock partitioning. However, the search complexity of this technique increases with the number of subblocks, and most importantly, it increases exponentially with the number of phase factors. Therefore, their selection is usually limited to a set with a finite number of elements [8]. Furthermore, this technique has the inconvenient of requiring the transmission of side information to the receiver telling which are the phase factors used for a correct decoding of the transmitted symbols, which reduces spectral and power efficiency as in the case of SLM. Thus, in PTS is required V IFFT operations and $\lceil \log_2 W^{(V-1)} \rceil$ bits of side information for each OFDM symbol.

D. Tone Reservation

In TR technique [7], [8], [12]–[14], some of the subcarriers, called frees, are not used to transmit information data. Once these subcarriers are free, they can take values that minimize PAPR, without introducing distortion in data subcarriers since all subcarriers are orthogonal. In some applications, there are subcarriers with SNR too low for sending information, therefore, they can be set as free subcarriers and used for PAPR reduction [8].

Let $\mathbf{C}$ denote the frequency-domain vector that contains the information of free subcarriers and that will be added to data vector, $\mathbf{X}$, in order to reduce the PAPR. For definition of the TR technique, $\mathbf{X}$ and $\mathbf{C}$ lie in disjoint frequency subspaces, resulting in the following signal

$$\widetilde{\mathbf{x}} = IFFT\left\{\mathbf{X} + \mathbf{C}\right\}. \tag{8}$$

The free subcarrier values are found solving a convex optimization problem (see Section II-E) and their locations are established *a priori* between the transmitter and the receiver.

As data subcarriers don't suffer distortion, this technique has no BER degradation, however this performance must have a penalty added once the free subcarriers require additional power and reduce the spectral efficiency. This penalty is given, in decibels, by

$$\mu = 10log_{10}\left(\frac{\left\|\mathbf{S}\widetilde{\mathbf{X}}\right\|^2}{\left\|\widetilde{\mathbf{X}}\right\|^2}\right), \tag{9}$$

where $\left\|\mathbf{S}\widetilde{\mathbf{X}}\right\|^2$ is the power of data subcarriers and $\left\|\widetilde{\mathbf{X}}\right\|^2$ is the total power of the OFDM symbol.

E. PAPR Reduction as Constrained Optimization Problem

Convex optimization has recently emerged as an efficient tool for reducing the PAPR of OFDM signals [13], [15]. This can be explained partially by the fact that convex optimization methods can efficiently compute global solutions to large scale problems in polynomial time. Furthermore, convex optimization approaches show advantages over the classical RCF approach [10]; see [13], [15] for more details.

An efficient way to reduce the PAPR is by distorting the OFDM constellation [24], [25]. The level of distortion is measured by the EVM and should be kept at a minimum, since a larger EVM value leads to a BER performance degradation. A single OFDM symbol's EVM is mathematically defined as,

$$\text{EVM} = \frac{||\mathbf{S}\left(\widetilde{\mathbf{X}} - \mathbf{X}\right)||^2}{||\mathbf{S}\mathbf{X}||^2}. \tag{10}$$

The PAPR can be further reduced by assigning a portion of energy to the free subcarriers [24], [25], i.e. a TR technique is performed. In this case, it must be taken into account the FCPO that measures the value of free subcarriers power and it is given by,

$$\text{FCPO} = \frac{||\left(\mathbf{I}_N - \mathbf{S}\right)\widetilde{\mathbf{X}}||^2}{||\mathbf{S}\widetilde{\mathbf{X}}||^2}. \tag{11}$$

The FCPO should be kept small since it measures the fraction of power "wasted" in the free subcarriers, which are not used to carry information. In this technique, we will focus on minimizing EVM for given PAPR and FCPO thresholds, i.e.,

$$\underset{\widetilde{\mathbf{X}} \in \mathbb{C}^{\mathbf{N}}}{\text{minimize}} \quad \text{EVM} \tag{12}$$

subject to

$$\text{PAPR} \leq g_1, \tag{13}$$

$$\text{FCPO} \leq g_2, \tag{14}$$

where g_1 and g_2 are PAPR and FCPO thresholds, respectively. It is straightforward to see that the optimization problem (12)–(14) is equivalent to

$$\underset{\widetilde{\mathbf{X}} \in \mathbb{C}^{\mathbf{N}}}{\text{minimize}} \quad ||\mathbf{S}\left(\widetilde{\mathbf{X}} - \mathbf{X}\right)||^2 \tag{15}$$

subject to

$$\widetilde{\mathbf{X}}^H \left(\mathbf{M}_i - \mathbf{T}_{g_1}\right) \widetilde{\mathbf{X}} \leq 0,\ i = 1, ..., \ell N, \tag{16}$$

$$\widetilde{\mathbf{X}}^H \left(\mathbf{I}_N - \mathbf{S}_{g_2}\right) \widetilde{\mathbf{X}} \leq 0, \tag{17}$$

where $\mathbf{M}_i = \ell N \mathbf{A}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{A}$, $\mathbf{T}_{g_1} = g_1 \mathbf{A}^H \mathbf{A}$, $\mathbf{S}_{g_2} = (g_2+1)\mathbf{S}$ and $\mathbf{e}_i$ represents the i^{th} column of the identity matrix $\mathbf{I}_{\ell N}$ [15]. The EVM optimization framework (15)–(17) results in a nonconvex optimization problem since the matrices $\mathbf{I}_N - \mathbf{S}_{g_2}$ and $(\mathbf{M}_i - \mathbf{T}_{g_1})$, for $i = 1, ..., \ell N$, are indefinite; in other words, all the constraints are nonconvex [26]. As most nonconvex problems, our problem is NP-hard and, thus, difficult to solve [26].

Alternatively, we can introduce an optional constraint that will keep the EVM below some preset threshold. This corresponds to a more challenging and realistic scenario where PAPR, FCPO and EVM are simultaneously constrained. In that case, the EVM optimization can be formulated as

$$\underset{\epsilon \in \mathbb{R}, \widetilde{\mathbf{X}} \in \mathbb{C}^{\mathbf{N}}}{\text{minimize}} \quad \epsilon \tag{18}$$

subject to

$$\text{EVM} \leq \epsilon\ \text{EVM}_{\max}, \tag{19}$$

$$(13),\ (14),\ \epsilon \leq 1, \tag{20}$$

where $\text{EVM}_{\max}$ is the maximum allowed EVM. Note that the optimization problem (12)–(14) is different from the one in (18)–(20), since the search space for $\widetilde{\mathbf{X}}$ in the former is larger than in the latter.

Although addressing the PAPR reduction of OFDM symbols as an optimization problem is a non-practical method, this has the advantage of providing some reference on the achievable PAPR, when the problem is properly formulated. Results obtained from formulation (18)–(20) will be used as comparison reference of the CF, SLM and PTS techniques previously described.

III. Performance Evaluation

In this section we perform a comparison of several PAPR reduction techniques on a coded OFDM transmission scenario when time-dispersive channels are considered, which to the best of the author's knowledge was not done before. We considered an OFDM signal with $N = 64$ subcarries and oversampling factor $\ell = 4$ under 16-QAM and 64-QAM constellation using Gray mapping rule. OFDM signal pass through different channels: AWGN and a time-dispersive channel with 32 paths. The coding scheme used was a $(1664, 840)$ LDPC code (coding rate near $1/2$), and the bits are interleaved inside each codeword. At the receiver, perfect channel estimation is assumed.

Results presented concern to the following techniques: CF, SLM, PTS and OPT, where OPT denotes the constrained optimized PAPR reduction technique. OPT was developed in two different ways, with free subcarriers, i.e. TR (OPT-TR), and without them (OPT). The analysis takes into account the following parameters: PAPR reduction, BER and the trade-off between them.

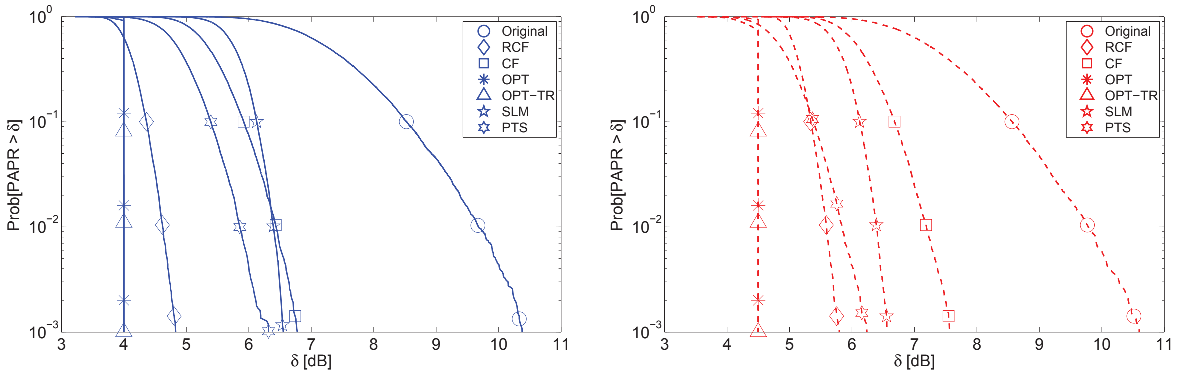
The CF technique was performed with 1 and 10 iterations, with the latter being denoted from now on as RCF, and the acronym CF referring to a single iteration. The CR was chosen to guarantee a good performance trade-off between PAPR reduction and BER, and was set to 1.4 and 1.7 (in linear units) for 16-QAM and 64-QAM, respectively.

In the case of SLM technique we multiplied each OFDM symbol by 32 different phase factor vectors with length N and values randomly chosen from the set $\{-1, 1, -j, j\}$. In PTS technique we partitioned each OFDM symbol into 4 time-domain sequences using adjacent partition and rotating them by phase factors $\{-1, 1, -j, j\}$.

In order to evaluate PAPR of each technique, the CCDF curves of each one are presented in Figs. 1a and 1b. The OPT techniques were set to achieve a PAPR of 4dB for 16-QAM and 4.5dB for 64-QAM and they will be used as reference, while the remaining techniques were tuned taking into account the trade-off between PAPR reduction and BER. At a clipping probability of 10^{-3} all techniques perform a considerable reduction in PAPR, however, as we will see in Figs. 3a and 3b, this results, in most cases, in a decrease in BER performance. Although this decrease is unbearable in OFDM uncoded as we can see in Fig. 2a, it can be bearable, even in a dispersive channel with multipath, if we use OFDM coded as we see in Fig. 2b.

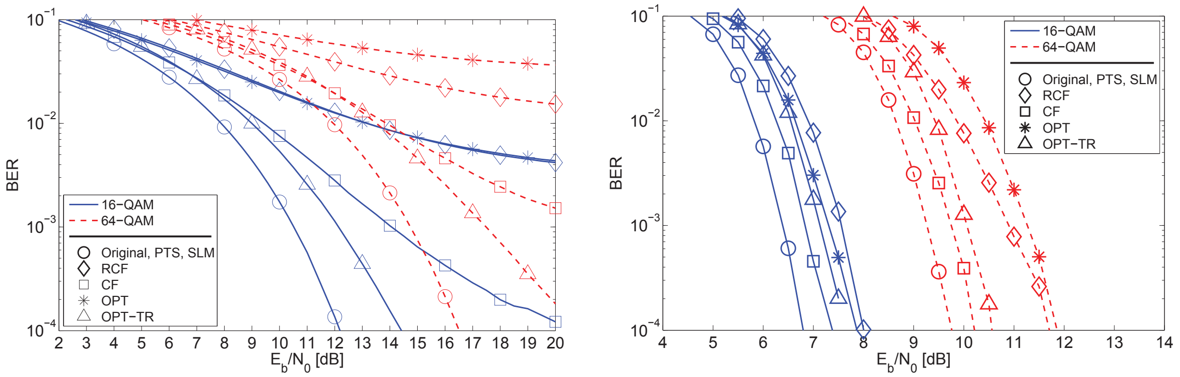
In fact, by analyzing Fig. 2b we can observe that all the PAPR reduction techniques perform close enough from the original OFDM transmission with unconstrained PAPR, i.e. no more than 1.2dB and 2dB for 16-QAM and 64-QAM respectively, thus with a power penalty much lower than the corresponding power gain achieved on restricting the PAPR. This shows that the use of an adequate coding scheme leads a great performance improvement even in a dispersive channel. It also shows that all techniques can have a similar performance when coding is used.

Thus, although PAPR reduction can lead to a decrease in BER performance which is a undesirable effect, this can be accepted in cases where this decrease is lower than the improvement in PAPR. Therefore, a BER curve shifted by the required amplifier's backoff given the PAPR level at a certain clipping probability is more informative, showing the trade-off between PAPR reduction and BER performance. Figs. 3a and 3b illustrate this trade-off for a time-dispersive channel, which is the most suitable real world situation and, $E_b^{peak} = E_b + PAPR(dB)$ with PAPR level chose to a clipping probability of 10^{-3} (see Figs. 1a and 1b).



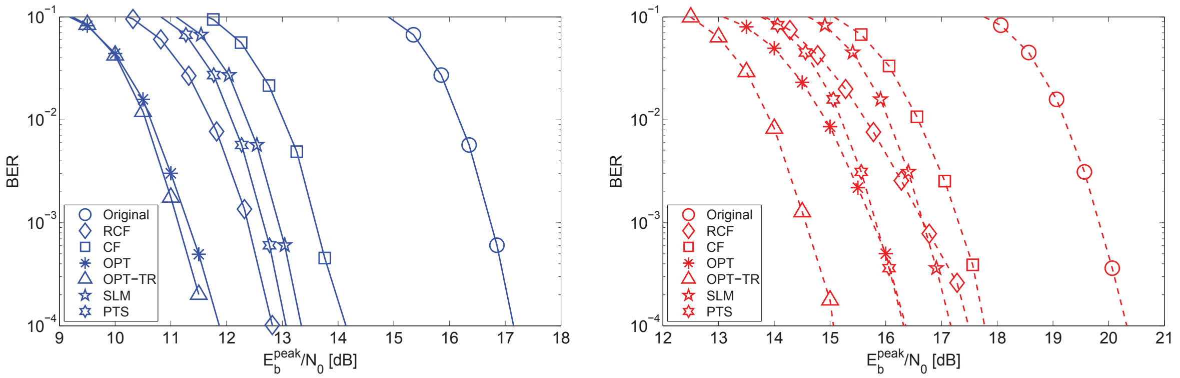
(a) 16-QAM (b) 64-QAM

Fig. 1: CCDF for different PAPR reduction techniques applied to OFDM



(a) Uncoded OFDM over an AWGN channel (b) Coded OFDM over a time-dispersive channel

Fig. 2: BER performance using several PAPR reduction methods.



(a) 16-QAM (b) 64-QAM

Fig. 3: BER results shifted by amplifier's backoff for coded OFDM over a time-dispersive channel using several PAPR reduction methods.

Analyzing Figs. 3a and 3b we can observe that techniques with higher complexity based on constrained optimization have the best results. However, in opposite direction, RCF technique has also good performance too with a very low complexity. Considering that we want to perform PAPR reduction in real-time, RCF with coded OFDM and the right parameters can have considerable good results even in a dispersive channel and, as so, it is a good candidate technique. The use in real-time applications rises another question for RCF, the latency of doing such iterative method. For this reason CF had been tested too and the results are very satisfactory mainly in 64-QAM modulation.

IV. Conclusion

In this paper, we go further than previous studies in PAPR reduction techniques and compare them for Quadrature Amplitude Modulation (QAM) constellations under a typical coded OFDM scenario with time-dispersive channels and not only for an ideal AWGN channel. Our performance results show that when OFDM is performed with an appropriate coding scheme, the most suitable technique for PAPR reduction, considering the implementation complexity, the response in a real-time situation and the trade-off between PAPR reduction and BER degradation, is clipping and filtering, with only 1 to 2.5 dB from the reference, even for relatively large QAM constellations.

Therefore, this work opens a subject in the study of PAPR reduction techniques. Once that is shown that CF is a good choice for PAPR reduction, it should be further studied in the future. To improve the performance of CF, coding and equalization schemes to tackle the distortion of clipping, and not only the channel effects, must be studied. If the effect of distortion of clipping can be reduced in the received signal, a better BER performance could be achieved and CF could have a trade-off between PAPR reduction and BER degradation even closer to the other techniques presented here. Also, the number the iterations of RCF may be studied in order to find an optimum value.

References

[1] R. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Artech House Publishers, 2000.

[2] U. S. Jha and R. Prasad, *OFDM Towards Fixed and Mobile Broadband Wireless Access*. Artech House Publishers, 2007.

[3] G. Wunder, P. Jung *et al.*, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 97–105, February 2014.

[4] P. Demestichas, A. Georgakopoulos *et al.*, "5G on the Horizon: Key Challenges for the Radio-Access Network," *Vehicular Technology Magazine, IEEE*, vol. 8, no. 3, pp. 47–53, Sept 2013.

[5] S. Müller, R. Bäuml, R. F. Fischer, and J. B. Huber, "OFDM with Reduced Peak-to-Average Power Ratio by Multiple Signal Representation," *Annals of Telecommunications*, vol. 52, no. 1-2, pp. 58–67, February 1997.

[6] L. Cimini and N. Sollenberger, "Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences," *Communications Letters, IEEE*, vol. 4, no. 3, pp. 86–88, March 2000.

[7] Y. Rahmatallah and S. Mohan, "Peak-To-Average Power Ratio Reduction in OFDM Systems: A Survey And Taxonomy," *Communications Surveys Tutorials, IEEE*, vol. 15, no. 4, pp. 1567–1592, Fourth 2013.

[8] S. H. Han and J. H. Lee, "An Overview of Peak-to-Average Power Ratio Reduction Techniques for Multicarrier Transmission," *Wireless Communications, IEEE*, vol. 12, no. 2, pp. 56–65, April 2005.

[9] R. Dinis and A. Gusmão, "On the performance evaluation of OFDM transmission using clipping techniques," in *Vehicular Technology Conference, 1999. VTC 1999 - Fall. IEEE VTS 50th*, vol. 5, 1999, pp. 2923–2928.

[10] J. Armstrong, "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," *Electronics Letters*, vol. 38, no. 5, pp. 246–247, Feb 2002.

[11] R. Dinis and A. Gusmão, "A class of nonlinear signal-processing schemes for bandwidth-efficient OFDM transmission with low envelope fluctuation," *Communications, IEEE Transactions on*, vol. 52, no. 11, pp. 2009–2018, Nov 2004.

[12] L. Wang and C. Tellambura, "Analysis of Clipping Noise and Tone-Reservation Algorithms for Peak Reduction in OFDM Systems," *Vehicular Technology, IEEE Transactions on*, vol. 57, no. 3, pp. 1675–1694, May 2008.

[13] J. Tellado, "Peak to average power reduction for multicarrier modulation," Ph.D. dissertation, Stanford Univ., Stanford, CA,, 2000.

[14] B. S. Krongold and D. Jones, "An active-set approach for OFDM PAR reduction via tone reservation," *Signal Processing, IEEE Transactions on*, vol. 52, no. 2, pp. 495–509, Feb 2004.

[15] Y. C. Wang, J. L. Wang, K. C. Yi, and B. Tian, "PAPR Reduction of OFDM Signals With Minimized EVM via Semidefinite Relaxation," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 9, pp. 4662–4667, Nov 2011.

[16] G. Yue and X. Wang, "A hybrid PAPR reduction scheme for coded OFDM," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 10, pp. 2712–2722, Oct 2006.

[17] L. Li, D. Qu, and T. Jiang, "Partition Optimization in LDPC-Coded OFDM Systems With PTS PAPR Reduction," *Vehicular Technology, IEEE Transactions on*, vol. 63, no. 8, pp. 4108–4113, Oct 2014.

[18] C. Tellambura, "Computation of the continuous-time PAR of an OFDM signal with BPSK subcarriers," *Communications Letters, IEEE*, vol. 5, no. 5, pp. 185–187, May 2001.

[19] Y. Wang and Z. Luo, "Optimized Iterative Clipping and Filtering for PAPR Reduction of OFDM Signals," *Communications, IEEE Transactions on*, vol. 59, no. 1, pp. 33–37, January 2011.

[20] X. Zhu, W. Pan, H. Li, and Y. Tang, "Simplified Approach to Optimized Iterative Clipping and Filtering for PAPR Reduction of OFDM Signals," *Communications, IEEE Transactions on*, vol. 61, no. 5, pp. 1891–1901, May 2013.

[21] M. Breiling, S. Muller-Weinfurtner, and J. Huber, "SLM peak-power reduction without explicit side information," *Communications Letters, IEEE*, vol. 5, no. 6, pp. 239–241, June 2001.

[22] T. Jiang and Y. Wu, "An Overview: Peak-to-Average Power Ratio Reduction Techniques for OFDM Signals," *Broadcasting, IEEE Transactions on*, vol. 54, no. 2, pp. 257–268, June 2008.

[23] S. Muller and J. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electronics Letters*, vol. 33, no. 5, pp. 368–369, Feb 1997.

[24] Y. C. Wang and K. C. Yi, "Convex Optimization Method for Quasi-Constant Peak-to-Average Power Ratio of OFDM Signals," *Signal Processing Letters, IEEE*, vol. 16, no. 6, pp. 509–512, June 2009.

[25] A. Aggarwal and T. Meng, "Minimizing the Peak-to-Average Power Ratio of OFDM Signals Using Convex Optimization," *Signal Processing, IEEE Transactions on*, vol. 54, no. 8, pp. 3099–3110, Aug 2006.

[26] S. Boyd and L. Vandenberghe., *Convex Optimization*. Cambridge University Press, 2004.

An Open-Source Based Speech Recognition Android Application for Helping Handicapped Students Writing Programs

Tong Lai Yu, and Santhrushna Gande
School of Computer Science and Engineering
California State University,
San Bernardino
tyu@csusb.edu, 004580341@coyote.csusb.edu

Ronald Yu
Department of Computer Science
University of Southern California,
Los Angeles
ronaldyu@usc.edu

WORLDCOMP‘15:ICWN15, July 27-30, 2015, Las Vegas, Nevada, USA.

Keywords: Speech Recognition, Android STT, Open Source, Learning Aid

Abstract

We describe in this paper how to use open-source speech recognition technologies to design and implement an Android application that helps students with physical disabilities write programs in classrooms. Google Voice Recognition (GVR)[13], which is a free and open Android tool, is utilized to convert the speech of a user to text. To fully utilize GVR, the Android phone has to be connected to the Internet.

In a typical setup, a handicapped student sits in front of a workstation with a large computer screen like the one shown in Figure 6. The student speaks to an Android phone, which converts speech to text using GVR. The text is then sent to the workstation through Wi-Fi for parsing and analysis. The processed text, which is a code segment of a program, is displayed on the workstation screen. Since the Android main activity thread cannot handle too many activities, we save the text of the speech in a buffer and use another thread, named communication thread, to send it to the workstation using the standard socket API. The producer-consumer paradigm is employed to synchronize the generation of text data by the main activity thread and the sending of the data by the communication thread[15, 18, 21].

The server program that runs on the workstation is written in C/C++. The main thread listens at a port. When it detects data, it creates two threads to handle the data. One thread created reads in the data, parses them into words, and puts the words in a circular queue. The other thread, named processing thread, simultaneously retrieves words from the queue, processes them to generate a code segment, saves the code in a file and displays it on the screen. A condition variable[7, 12] is used to synchronize the tasks between these two threads.

The keywords and symbols of the programming language that the student is using, which are saved in a file are loaded into a table. The processing thread uses a hashing and mapping scheme to obtain the proper keywords and symbols from the table; as humans often speak with inconsistency, several different words may map to the same keyword. For example, when one tries to say the word import, they may say it slightly different from the standard pronunciation and the recognizer generates the word important. The scheme will map important to the same location as import to retrieve the correct keyword.

1. Introduction

In recent years the number of mobile applications has been growing with tremendous speed. Mobile devices have become ubiquitous and in the last few years, Android, an open-source software stack for running mobile devices, has become the dominant platform of many mobile devices such as tablets and smart phones[8].

Open-source software has been playing a critical role in recent technology developments. A lot of breakthroughs in technology applications such as Watson’s Jeopardy win[4] and the phenomenal 3D movie Avatar[3] are based on open-source software. It is a significant task to explore the usage of available open-source or free tools to develop software applications for research or for commercial use[22]. We report in this paper the design and development of an Android application, based on free or open-source tools, which helps physically handicapped students write programs.

In a large university such as California State University at San Bernardino (CSUSB), which has more than 30,000 students, there is always a small but significant fraction of students who have certain physical disabilities that make typing difficult for them. When these students take a programming course, they require special services from the institution, as they may write or type too slowly to follow the pace of the lecture. Very often, the institution offers an assistant to help a handicapped student in class, typing code segments in a workstation for the student and executing them to see how the code works. These students have to overcome tremendous physical and mental barriers to finish a degree in science or engineering that often require some programming classes. This has been the situation in CSUSB for many years. The Android mobile application presented here would alleviate the disadvantages of these students by providing tools that allow them to write programs in a workstation effectively with very little or no typing. Using this application, a student sits in front of a workstation with large screen display and speaks to an Android phone to dictate a program, which is displayed on the workstation's monitor. (A mobile phone's display is too small for any beginning student to learn programming on it.)

The application consists of two components: a client and a server. The client runs in an Android device, which connects to the Internet via Wi-Fi. It accepts speech from the student and converts it to text using Google Voice Recognition (GVR)[13], a free tool with open API. The text is sent to the server, which runs in a workstation using the free open-source Linux operating system and is provided to students to write programs in a classroom.

The server is written in C/C++. It reads in the incoming text, and makes analysis of it to form a syntactically correct program segment of a specified programming language such as Java. It then displays the code segment on the screen and saves it to a file.

In rare cases, when a workstation is not available, the text is processed by another Java server program residing in the Android device. The Java server program uses a different technique to map ambiguous speech text to more specific keywords or symbols from a pre-constructed file; it saves the code segment in a file. In this paper, we mainly describe the normal operation of the application, which displays programs on a workstation screen.

Speech recognition (SR) by machine, which translates spoken words into text has been a goal of research for more than six decades. It is also known as *automatic speech recognition* (ASR), *computer speech recognition*, or simply *speech to text* (STT). The research in speech recognition by machine involves a lot of disciplines, including signal processing, acoustics, pattern recognition, communication and information theory, linguistics, physiology, computer science and psychology. Figure 1 shows a general block diagram of a task-oriented speech recognition system.

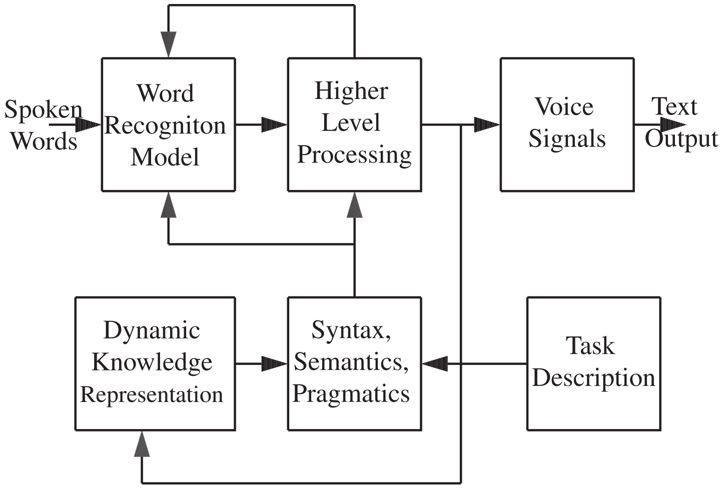


Figure 1 A Typical Speech Recognition System

Nowadays, speech recognition (SR) mobile products are ubiquitous. There are many third party SR apps that support Android. We have chosen Google Voice Recognition (GVR)[13], which is preinstalled in many Android devices, as our recognition engine. GVR makes use of neural network algorithms to convert human audio speech to text and works for a number of major languages but we use English as our example in our description.

A neural network consists of many processors working in parallel, mimicking a virtual brain. The usage of parallel processors allows for more computing power and better operation in real-time, but what truly makes a neural network distinct is its ability to adapt and learn based on previous data. A neural network does not use one specific algorithm to achieve its task; instead it learns by the example of other data. Though GVR may work in some Android phones offline, it normally accesses through Internet its large database for voice recognition attempted by previous users. It also looks at previous Google search queries so that the voice recognition engine can guess which phrases are more commonly used than others. This way, even if the user does not speak a certain word clearly, GVR can use the context of the rest of the spoken phrase or sentence to extrapolate what the user is most likely trying to say.

In general, a neural network can learn from two major categories of learning methods–supervised or self-organized. In supervised training, an external teacher provides labeled data and the desired output. Meanwhile, self-organization network takes unlabeled data and finds groups and patterns in the data by itself. GVR learns from its own database through the self-organization method.

2. Android Threads Synchronization

The Android client involves a few tasks, including interfacing to the user, accepting text from the GVR engine, and communicating with the server. The GVR itself also has to connect to the Internet through Wi-Fi to interact with the Google cloud database. The execution time for each task is never a constant. In particular, the bandwidth of a Wi-Fi communication can fluctuate widely, depending on the traffic of the environment. So in the application, we use two threads to handle the tasks independently so that they won't interfere with each other. The *main activity thread* interacts with the user and calls the GVR engine to convert any spoken words to text and saves it as strings in a shared queue. The other thread, the *communication thread*, reads the strings from the queue and sends it to the server, which resides in a workstation. This is shown in the block diagram of Figure 2.

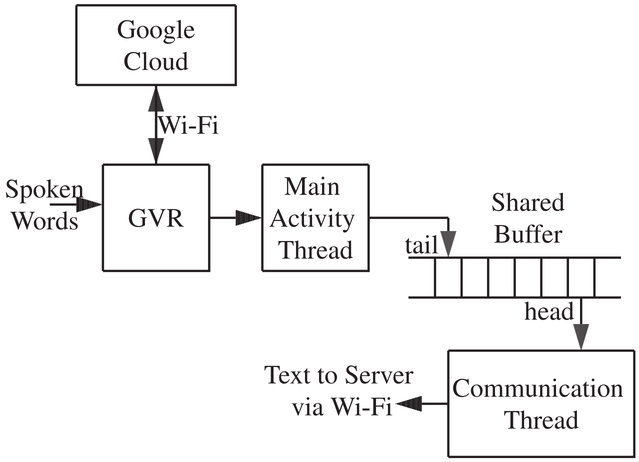
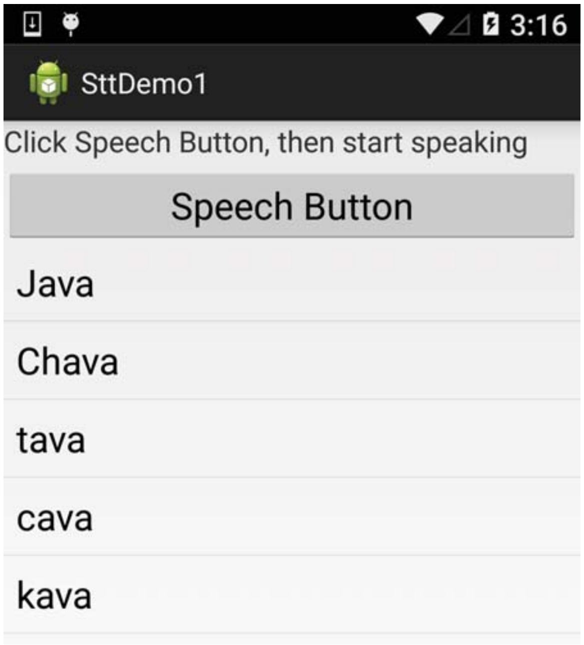


Figure 2 Android Client

To ensure that the two threads will not interfere with each other's task, we employ the producer-consumer paradigm[15, 18], a well-studied synchronization problem in Computer Science, to synchronize the tasks between them. A classical producer-consumer problem has two threads (one called the producer, the other the consumer) sharing a common bounded buffer. The producer inserts data into the buffer, and the consumer takes the data out. In our case, the buffer is a queue where strings are entered at the tail and are read at the head. Physically, the queue is a circular queue. Logically, one can imagine it to be a linear infinite queue[21]. The head and tail pointers are always advancing (incrementing) to the right. (To access a buffer location, the pointer is always taken the mod of the physical queue length, e.g $tail \ \% \ queue\_length$.) If the head pointer catches up with the tail pointer (i.e. $head = tail$), the queue is empty, and the consumer must wait. If the difference between the *head* and the *tail* is equal to the length of the buffer, the queue is full, and the producer must wait. In the application, the *main activity thread* is the producer and the *communication thread* is the consumer. In this way, the main thread can interact with the GVR engine and the user while the *communication thread* is sending data at the background.

Figure 3 UI of GVR



3:16

SttDemo1

Click Speech Button, then start speaking

Speech Button

Java

Chava

tava

cava

kava

Figure 4 Android Client Interface

A user pushes an image button presented by the client program to start GVR. The user then speaks to the phone, which is presenting the GVR interface as shown in Figure 3. When a user speaks a sentence or a word with ambiguity, GVR may suggest up to 5 choices. Based on our experience, the first suggested one is most likely the one we want. For simplicity, the application just sends the first choice, and discards the rest. If necessary, the user can issue a *discard* command to the server (see description below), which discards the previous sentence, and the user can repeat the speech. Figure 4 shows the Android interface of the application and the five words, *Java, Chava, tava, cava*, and *kava*, suggested by GVR when one of the authors spoke the word '*java*'.

3. The Workstation Server Threads

The text converted from speech by the Android client is sent to a server program running in a workstation placed in front of the student. The server is a multi-threaded program implemented in C/C++. Instead of using the POSIX threads, we have used the open-source cross-platform SDL threads[14], well-known by its robust characteristics, in our implementation. The SDL threads are significantly simpler than the POSIX threads but have enough features that satisfy all the requirements of our application. The C/C++ standard template library (STL) is used to facilitate the implementation.

We use the socket API function **read**() to read in the data as a stream of bytes from the network. The function **read**() is a blocking command, inhibiting the thread to proceed while it is waiting for data to come. Therefore, we create two threads to read and process the data. One thread, the *reading thread*, reads in the text using **read**(), obtains a word, puts it in a string buffer, which is a STL *deque* (double-sided queue), and continues to read in more text. The other thread, the *processing thread*, retrieves a word from the buffer and processes it. The two threads work independently and will not interfere with each other's activities.

The synchronization between them is done using a *condition variable*[12], which can help solve problems that could be complicated to solve using *semaphores*[7]. Supported by both POSIX and SDL, a condition variable is a queue of threads (or processes) waiting for some sort of notifications. A condition variable queue can only be accessed with two methods associated with its queue, typically called **wait** and **signal**. Threads wait for a guard [9] statement to become true to enter the queue and threads that change the guard from false to true could wake up the waiting threads. In practice, it always works with a mutual exclusion variable. The following code segment shows how such a variable is utilized to synchronize between the *reading thread* and the *processing thread* with some minor details omitted. In the code, the variable *mutex*, representing mutual exclusion, is a binary semaphore for locking and unlocking a code section, and the variable *strQueue* is the the condition variable; the routine **read_data**() reads a word, puts it in the character array *a*, and returns the number of characters read.

```
#include <SDL/SDL_thread.h>
#include <deque>
deque<string> strArray;//string buffer
SDL_mutex *mutex;
SDL_cond  *strQueue;
.....
Reading_thread:
  char a[200];
  while ( read_data( a ) > 0 ) {
     string s ( a );
     SDL_LockMutex ( mutex );
        strArray.push_back ( s );
     SDL_CondSignal ( strQueue );
     SDL_UnlockMutex ( mutex );
  }

Processing_thread:
  SDL_LockMutex ( mutex );
  while((size = strArray.size()) == 0)
     SDL_CondWait ( strQueue, mutex );
  string s = strArray.front();
  strArray.pop_front ();
  SDL_UnlockMutex ( mutex );
  .....
```

Note that in this example, the accessing of the string buffer *strArray* is guarded by the statement

```
 strArray.size() == 0
```

The command **SDL_CondWait**() sends the thread to sleep and releases the lock *mutex*. When it is awaken by the other thread, it will try to acquire the lock *mutex* again.

This code is significantly simpler than the circular buffer technique we used in the previous section of the Android client. The main disadvantage of this method is that it only allows one thread to access the string buffer at one time while the circular buffer allows both the producer and the consumer threads to access the buffer simultaneously as long as the *head* and *tail* do not point to the same slot. Since the server program is written in C/C++ and runs in a workstation, which has much more computing power and resource than that of a mobile phone, the technique could read and process data seamlessly and would not cause jitters in presenting the data.

4. Searching by Hashing

We use a hashing scheme to lookup keywords and commands in our applications. Hashing is a very fast searching method, with time complexity O(1). Its main disadvantage is that a table with preset size is needed to store the keys and associated data. The required table could be huge if the key space is large. However, the number of keywords in common computer languages such as Java and C/C++ is relatively very small as compared to a natural language. Therefore, in our application, hashing is an ideal candidate for looking up keywords or commands.

In our scheme, we save all the possible speech text for the keywords, which include the language keywords, symbols, and any made-up sentences, and load them into a table. Figure 5 shows a sample segment of this table. The first column shows the indices that the corresponding speech words will map to; the second column shows the number of words that will map to the index and the third column is the keywords that will be retrieved.

Idx	Count	Keyword	Speech text
1	4	import	import Import important impact
2	6	java	java Java Chava tava cava kava
3	4	public	public Public puppet poppet
4	5	b	b B bee Bee be
5	4	.	dot Dot thot doct
6	7	=	equal Equal eco Eco eagle Eagle Ecol

Figure 5 Hashing Table (Idx=Index)

For example, when the application receives any of the words, *equal, Equal, eco, Eco, eagle, Eagle*, or *Ecol*, the symbol "=" is retrieved. We also need another similar table that stores some commands we manually created. For example, we make the word *variable* a command. If users want to create in the program a variable called *beeb*, they have to first say *variable*, and then spell out 'b', 'e', 'e', 'b'. The command *upper* capitalizes a word or a number of words. In the program, each command is handled differently so that the application knows how to process each command accordingly.

Normally, when the application receives a word, it first searches the command table. If the app cannot find the word, it is not a command. The app then searches the keyword table. If it still cannot find the word, the speaker has to say the word again. In the worst situation, the user can always use the command *variable* to create a keyword by spelling its letters out. One can also use certain commands to discard the previous word or line.

A keyword shown in the third column of the table does not need to be a real keyword or symbol of the computer language. It can be any text that would help the user in the development process. For example, it could be a statement like,

```
public static void main(String[] args)
{
}
```

When a user says "main" or "Main", the whole statement is retrieved. Similarly, the set of keywords can contain some other commonly used statements such as "for (int i = 0; i < N; i++) " or "System.out.printf(". When saving such a keyword (which contains white spaces) in a file, one may use a special symbol like the '@' character as a marker, to mark the beginning and the end of the keyword, and removes the markers when loading the keyword into a table. Alternatively, one may handle keywords with multiple spaces separately, saving them in a separate file and in a different format.

The hashing scheme handles these cases just in the same way it handles a simple short keyword. In fact, the set of keywords can be tailored for any particular programming class. For example, the instructor can provide some programming templates to students as part of the keywords. A student retrieves the templates by speaking one word to the phone.

The following code, which has omitted some minor details, shows how this hashing scheme can be implemented using C/C++, assuming that the file pointer *fpi* points to the text file that contains the data in the format shown in Figure 5.

```
#include <ext/hash_map>
using  namespace __gnu_cxx;
const int MaxSize = 1024;
string keyWords[MaxSize];
hash_map<int, int>Map;
hash<const char *>H;
.....
int k = 1, n;
char buf[100];
int count;
//build the mapping table
while ( true ) {
  if (fscanf(fpi,"%d %s",
            &count,buf)==EOF) break;
  keyWords[k] = string ( buf );
  for ( int i = 0; i < count; i++ ) {
    fscanf ( fpi, "%s", buf );
    n = H ( buf ); //hash a word
    Map[n] = k;    //map to index,
                   // starting from 1
  }
  k++;
}
```

In the code, the hash function **H** hashes a string (an array of characters) to an integer. The hash map, **Map** maps the integer to an index of the keyword table, starting from the value 1. If an integer i has not been mapped to an integer, the value **Map**[i] is 0. From the code we see that all the words in the same row will map to the same index *k* and the corresponding keyword is given by *keyWords[k]*. We have to use namespace *__gnu_cxx* because the standard C/C++ libraries have not implemented this hashing scheme. It is supported by the external open-source GNU libraries.

Note that by using this hashing scheme, one can easily modify the hashing table of Figure 5, which is saved in a file, to allow the user to speak in another language other than English. As long as the speech text points to the correct keyword, it always produces the same final program.

5. Results and Discussions

Figure 6 shows a typical classroom for a programming class at CSUSB. Linux workstations with large screen and Internet connection are provided to students to write programs in a class session. As shown at the lower left corner of the figure, the mobile phone lying on the table is an Android phone, which is provided to students who need it to write programs. However, it has a relatively very small screen, inappropriate for direct program development. The student must make use of the Linux workstation to help him or her to do the tasks.

The room would become too noisy if all students used speech to write their programs. However, if only one or two students who are physically handicapped use speech to write their programs, the environment works well. We have carried out experiments on this situation and found that other students are not bothered by the speech input. Moreover, the lecturer is too far from the cell phone and won't interfere with the student's speech input to his or her mobile phone.

In conclusion, we have developed an Android application, using open source or free tools, that assists students who are physically handicapped to write programs in a programming class. The producer-consumer paradigm is used to synchronize tasks in the client, which runs in an Android phone. On the other hand, a condition variable is used to synchronize tasks in the server, which runs in a Linux workstation. A hashing scheme is used to simplify and retrieve keywords of a language, which we have only considered for the Java programming language. Though there are still many minor details that need to be improved and fine-tuned, the application works well in a typical programming classroom. In the future, we shall extend the application to include other common languages such as C/C++ and Perl. Moreover, the data files can be easily modified to let users speak in another language such as Chinese, Korean or Japanese but still generate the same program.

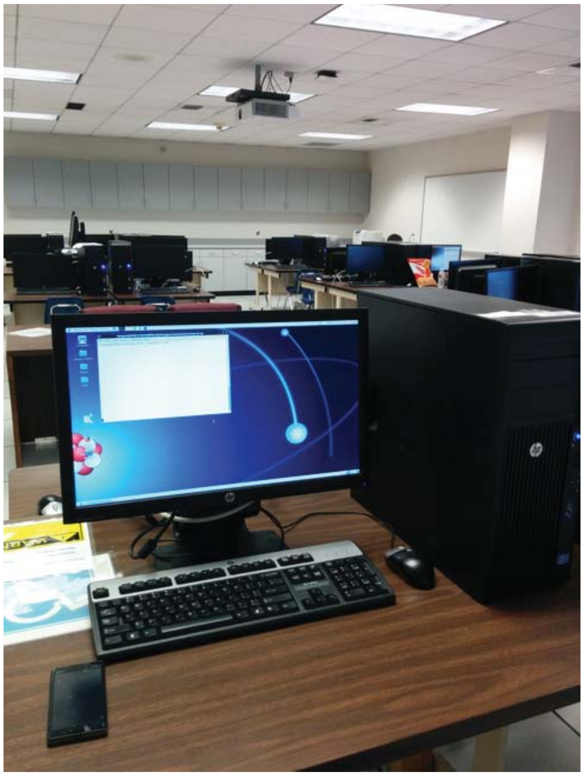


Figure 6 A Classroom for Programming Courses

References

[1] Android Open Source Project: TextToSpeech, *http://developer.android.com/reference/android/speech/*

[2] Android Developer Reference, *http://developer.android.com/reference/java/util/zip/package-summary.html*

[3] Jun Auza, *The Technology Behind Avatar (Movie)*, *http://www.junauza.com/2010/01/technology-behind-avatar-movie.html*

[4] Charles Babcock, *Watson's Jeopardy Win A Victory For Mankind*, Information Week, Feb 2011.

[5] C. Bregler, M. Covell, and M. Slaney, *Video Rewrite: Driving Visual Speech with Audio*, p.353-360, SIGGRAPH'97 Proceedings, ACM Press, 1997.

[6] E. Cosatto, H.P. Graf, and J. Schroeter, *Coarticulation method for audio-visual text-to-speech synthesis*, US Patent 8,078,466, Dec 2011.

[7] T. W. Doeppner, *Operating Systems in Depth*, John Wiley & Sons, Inc., 2011.

[8] Forbes Magazine, *Android Solidifies Smartphone Market Share*, *http://www.forbes.com/*, Jan., 2013.

[9] , C.A.R. Hoare, *An Axiomatic Basis for Computer Programming*, p. 576-583, Comm. ACM 12, Oct. 1969.

[10] D. H. Hubel, and T. N. Wiesel, *Receptive Fields Of Single Neurones In The Cat's Striate Cortex*, Journal of Physiology,) p. 574-59I, (148), 1959.

[11] D. H. Hubel and T. N. Wiesel, *Receptive Fields, Binocular Interaction And Functional Architecture In The Cat's Visual Cortex*, Journal of Physiology, p. 106154, (160), 1962.

[12] F. June, *Android Programming and Open Source Tools*, CreateSpace, 2014.

[13] S. Mlot, *Google Adds Speech Recognition to Chrome Beta*, *http://www.pcmag.com/article2/0,2817,2414277,00.asp* , PC Magazine, Jan. 2013.

[14] *https://www.libsdl.org/*

[15] A. Silberschatz et al., *Operating System Concepts*, Addison-Wesley, 1998.

[16] M. Singhal and N.G. Shivaratri, *Advanced Concepts in Operating Systems*, McGraw-Hill, 1994.

[17] Sketchup, *http://www.sketchup.com/intl/en/product/gsu.html*, 2013.

[18] A.S. Tanenbaum, *Modern Operating Systems*, Third Edition, Prentice Hall, 2008.

[19] University of Maryland, *Blendshape Face Animation*, *http://userpages.umbc.edu/bailey/Courses/Tutorials/ ModelNurbsHead/BlendShape.html*, 2009.

[20] L. Wood et al., *Document object model (dom) level 1 specification*, W3C Recommendation, 1, 1998.

[21] Ronald Yu, Tong Lai Yu, and Ihab Zbib, *Animating TTS Messages in Android using OpenSource Tools*, Proceeedings of The 2013 International Conference on Computer Graphics & Virtual Reality, P.10-15, WORLDCOMP'13, July 22-25, Las Vegas Nevada, USA, 2013.

[22] T.L. Yu, "Chess Gaming and Graphics using Open-Source Tools", *Proceedings of ICC2009*, p. 253-256, Fullerton, California, IEEE Computer Society Press, April 2-4, 2009.

Stability and Performance Evaluation of Wireless Tele-Control System for MIMO Plant

[1]Faramarz Alsharif, [2]Shiro Tamaki, [3]Katsumi Yamashita, [4]Tustomu Nagado, [5]Tomokazu Nagata, [6]Mohammad Reza Alsharif , [7]Bruno Senzio-Savino and [8]Heung Gyoon Ryu

[1,2, 4,5,6,7] Graduate School of Engineering and Science
University of the Ryukyus
Nishihara, Japan
[3]Osaka Prefecture University
Osaka, Japan
[8] Chungbuk National University
Cheongju, South Korea
Department of Electronic Engineering
[1]faramarz_asharif@yahoo.com
[2]shiro@ie.u-ryukyu.ac.jp
[3]yamashita@comm.ees.osakafu-u.ac.jp
[4]nagado@eee.u-ryukyu.ac.jp
[5]nagayan@ie.u-ryukyu.ac.jp
[6]asharif@ie.u-ryukyu.ac.jp
[7]b.senzio@gmail.com
[8]ecomm@chungbuk.ac.kr

***Abstract*— In this research, we aim to evaluate the performance and stability of a jointed MIMO (Multi Input and Multi Output) plant and wireless communication system. In order to realize a distant control wireless communication can be applied to control system. The advantage of Wireless Tele-Control system is to minimize the weight of control object and can also be applied on unmanned vehicle system. However, In Wireless communication system channels are occurred due to utilization of wireless networks that can cause multipath channel. A multipath channel consists of accumulated delayed and attenuated reference signal. In wireless Tele-Control system, output signal are fed back to the controller in order to stabilize the closed-loop or compensate the performances. Due to this matter we have two multipath channels in the closed-loop system. One is feedforward channel and the other is feedback channel. Because of existence of multipath channel, control input and feedback signal are affected by multipath channels. As a result undesired signal are observed due to existence of multipath channel and may cause the closed-loop instability or performance degradation. General Speaking, design of controller with multipath channel becomes stiff problem to satisfy the stability of the closed-loop. Thus, in order to simplify the design of controller, we have jointed equalizers in the control loop. Meanwhile the effects of channels are reduced by equalizer which is consisted of FIR (Finite Impulse Response) filter. The stability and performance of the closed-loop system can be evaluated by step response. The plant is set to be a drone that aims to control the attitude. In conclusion we discussed about the performance and stability of Wireless Tele-Control system in frequency domain. Eventually we could confirm the simplification of controller design for MIMO and Wireless Tele-Control System.**

Keywords: Wireless Tele-Control System, Multipath Channel, Equalization, MIMO system, Unmanned Aerial Vehicles

1 Introduction

The utilization of Wireless Tele-Control system is one of the significant issues in the servo systems. Especially, when system requires control in distant. The advantage of Wireless Tele-Control system is that maintenance and management of controller can be done easily since controller is located in observation center and plant may be located in distant. One more thing is that by Wireless Tele-Control system since controller is not loaded on plant, it could be considered as reduction of load in plant and makes system performance enhanced. Let us clarify the Wireless Tele-Control system. Basically, in Wireless Tele-Control system they are always two channels. One is the feedforward channel to send the optimal or compensated input to the control plant and the other one is the feedback channel since output signal should be sent to the controller side in order to calculate the error and to minimize it or stabilize the closed-loop. So, these channels are disadvantages of utilization of Wireless Tele-Control system. First of all due to the usage of communication system in the closed-loop system we would have some impairment such as phase noise, Doppler effects, frequency offset, delays and attenuations. The mentioned impairment can be solved by implanting the system that has high function capabilities. Therefore, phase noise, Doppler effects, frequency offset can be repaired by installing the advanced function capability. However, the received signal should be equalized to get the original information from sender. Therefore, in order to get the exact data from sender it is required to equalize the received signal. The received signal may be distracted by the multipath channel. Multipath channel effect occurs concerning the circumstances of the environment of control plant. In other words, multipath channel is inclusion of accumulated delayed and attenuated direct path signal. Even though sender has sent the original signal but interfered signal will be received in receiver side. Thus, equalization of signal is required in receiver side. For equalization, first we have to compose the replica of the unknown channel. The composition of the replica of the unknown channel can be done by FIR adaptive filter. However, the composition of the replica channel is not sufficient. In order to get reference signal it is required to realize the inverse transfer function of replica Channel. Therefore, the inverse channel is realized after the receiving the distracted signal. This has role of equalizing the received signal. These processes should be implemented in two different stages. One is the feedforward side of the receiver and the other one is the feedback part of the receiver since we have round trip multipath channel in the closed-loop system. After realizing the equalizer, implementation can be done in the closed-loop system. Furthermore, controller can be designed according to plant without considering multipath channel. After designing a controller equalizers and controller are jointed in cascade. Thus, controller and equalizers are jointed in the closed loop system. In Next chapter we will introduce unmanned aerial vehicles and controller design of it.

2 Design of Feedback Controller for Unmanned Aerial Vehicles

In this chapter we introduce briefly design of an unstable system in order to stabilize the closed loop system. The plant is set to be unmanned aerial vehicles. First of all let us introduce unmanned aerial vehicles briefly.

The applications of unmanned aerial vehicles have been extended in both military and civilian fields around the world in the recent years. Unmanned aerial vehicles are used in all military ranging from investigation, monitoring, intelligence gathering, battlefield, distant investigation and several other aims. As well Civilian applications include remote sensing, transport, exploration, and scientific research. Because of vast application unmanned aerial vehicles is expected to be in vogue. Therefore, in the view of reliability of stabilization and performance, even in drastically changing environment such as strong storm unmanned aerial vehicles should be operated to do its duty. Therefore, controller required to be installed in the closed loop system. However, as we have mentioned the mass of unmanned aerial vehicles is not recommended to be increased as a stability and performance point of view since it may cause instability and performance degradation. Thus Wireless Tele-Control system is proposed. Following shows how to stabilize a plant.

First of all let us consider a plant which is that $G = \frac{1}{s^2}$ for a >0 . A basic PID (proportion integral and derivative) feedback controller can be considered as

$$K = Kp\left(1 + \frac{1}{T_i s} + \frac{T_d s}{as+1}\right) \tag{1}$$

that set (K_p, T_i, T_d, a) are tuned to make the closed loop system internally stable.

Here, K_p is proportional gain, T_i is integral gain, T_d is derivative gain and a is derivative approximated gain. The following conditions should be satisfied to maintain internal stability and performance enhancement of the closed loop system.

(1)- If and only if the real part of solutions of characteristic equation are less than zero.

Characteristic equation becomes as follows:

$$s^4 + \frac{1}{a}s^3 + K_p\left(1 + \frac{T_d}{a}\right)s^2 + K_p\left(\frac{1}{T_i} + \frac{1}{a}\right)s + \frac{K_p}{aT_i} = 0 \quad .$$

If the above equation's pole is set of number such as

$$\Pi = (p_i, \forall i = 1,2,3,4 \mid p_i \in C) \tag{2}$$

Stability condition is Re (Π) < 0.

(2)- Frequency response of Sensitivity function *S(jω)* should contain the following specification.

ω_b : Band width frequency

For ω < ω_b | S(jω) | << 0 [dB]

For ω > ω_b | S(jω) | ≃ 0 [dB]

The above condition described that when for sensitivity in the low frequencies it has small gain most of interferences and disturbances with direct current component charactresitc that can affect as external disturbances are not influenced the closed loop system. For high frequency domain it is desirable to maintain 0 [dB] to reduce the tracking error.

(3)- Frequency response of Complementary sensitivity function *T(jω)* should contain following specification.

For ω < ω_b | T(jω) | ≃ 0 [dB]

For ω > ω_b | T(jω) | << 0 [dB]

The above condition described that when for Complementary sensitivity in the low frequencies it nearly equal to 0 [dB] that means for reference signal that contains direct current componenet, it becomes almost same in the output of plant that is desiable. For high frequencies, it is desirable drop to small gain since can reduce the affect of feedback noise and plant uncertainty.

(4)- Frequency response of open loop function $G(j\omega)K(j\omega)$ should contain following specification.

ω_c: Cross frequency

$$\text{For } \omega < \omega_c \quad |\, G(j\omega)K(j\omega) \quad | >> 0\ [dB]$$
$$\text{For } \omega > \omega_c \quad |\, G(j\omega)K(j\omega)\, | << 0\ [dB]$$

The above condition describes that when open loop in low frequency domain contain large gain, it can enhance the performances and when it is in high ferqnecies gain should be small values to reduce the effect of uncertainty in plant and external disturbances.

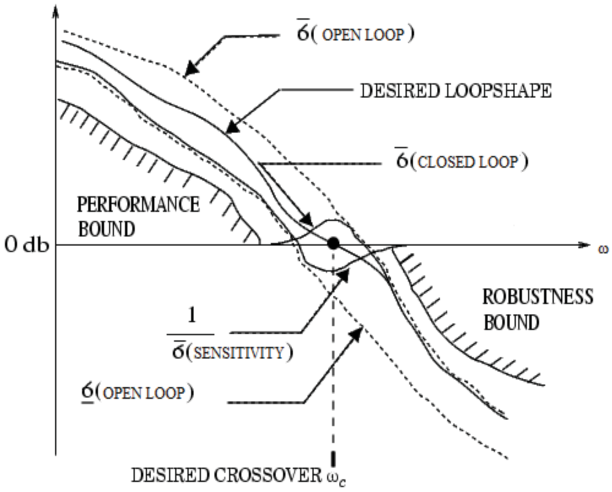


Fig.1 Singular value specifications on open loop, sensitivity, and closed loop

Here is an example of the described conditions.

$$plant\!: G(s)=\frac{1}{s^2}I_{3\times3}$$

$$Controller\!: K = Kp\left(1+\frac{1}{T_i s}+\frac{T_d s}{as+1}\right)I_{3\times3}$$

$$Open \quad Loop : G_O(s)=G(s)K(s)$$

$$Sensitivity : S(s)=\left(I+G_O(s)\right)^{-1}$$

$$Complementary \quad Sensitivity : T(s)=G_O(s)\left(1+G_O(s)\right)^{-1}$$

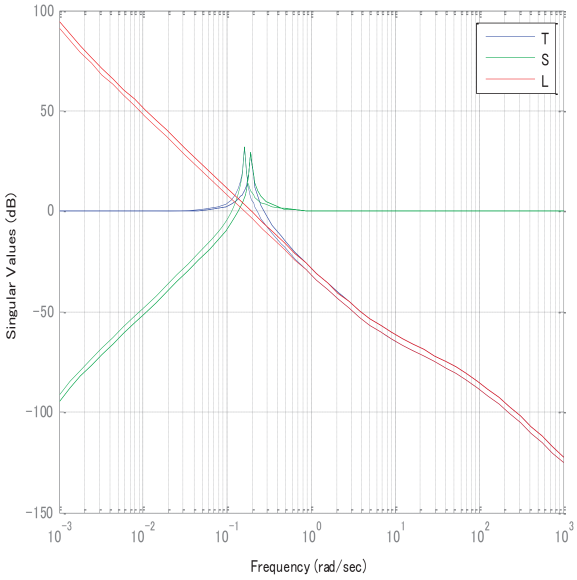


Fig.2 Specification of desired frequency response

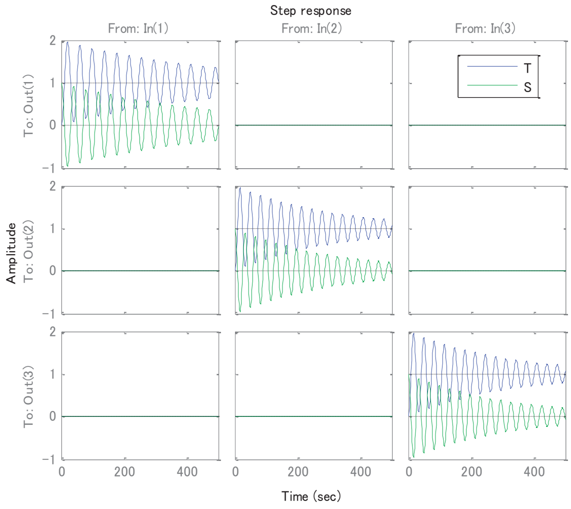


Fig.3 Specification of desired step response

As there are shown in the above figures, Fig.2 shows the desired frequency response of sensitivity, complementary sensitivity function and open loop, respectively. It is obvious that when frequency increase open loop and complementary sensitivity function overlapping each other. This matter can be confirmed mathematically.

As it shown on above equation, open loop has significant influence on the closed loop stability and performances. Therefore, controller design should be done by acquiring the plant characteristic and frequency response. However, in actual and practical cases, there would be parameters perturbation in plant and uncertainty always presents. Thus, a control scheme should overcome the uncertainty problem when design a controller. Next chapter joint system of feedforward and feedback control is introduced

3 Wireless Tele-Control System

So far we have discussed about the stability condition and performances of the closed loop system for ordinary case(without Channel). Here, let us define Wireless Tele-Control System as follows. Following figure shows the structure of Tele-Control system.

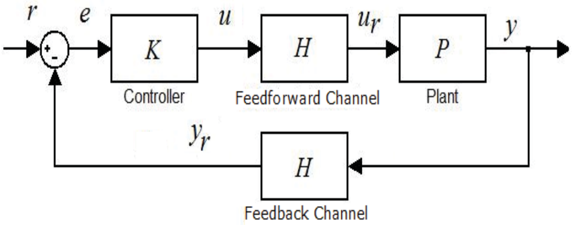


Fig. 4 Tele-Control System

Here H is Multipath channel and u_r, y_r are received input and received output signal, respectively. Through Fig.4, we can get the closed-loop system's transfer function according to following equations.

$$y = PHKe \qquad (3)$$

$$e = r - Hy_r \qquad (4)$$

Afterward we get the transfer function between r and y which is complementary sensitivity transfer function as follows.

$$y = \Delta_H PHKr \qquad (5)$$

Where, $\Delta_H = (1 + HPHK)^{-1}$ stands for sensitivity transfer function which is from r to e.

As we can see in sensitivity function of the closed-loop system, it has been involved with Channel's square.
The stability condition of closed-loop system is

$$\bar{\sigma}\left((1+H(j\omega)P(j\omega)H(j\omega)K(j\omega))^{-1}P(j\omega)H(j\omega)K(j\omega)\right)$$
$$<1 \qquad \forall\omega.$$

Which $\bar{\sigma}(.)$ indicate the maximum singular values.
However, due to existence of channels, it is stiff problem to satisfy the above condition. Therefore, our proposed method is to reduce the effect of the channel in the sensitivity function. The proposed method has shown in following Fig. 5.

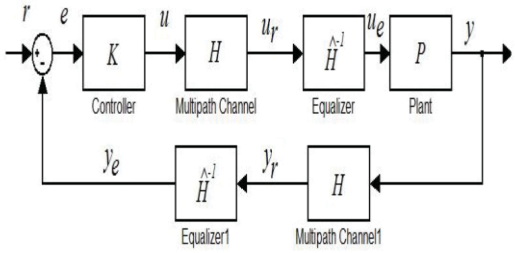


Fig. 5 Configuration of the proposed method

Here, y_e, u_e and $\hat{H}^{-1}$ stand for the equalized input signal, equalized output signal and Equalizer, respectively. $\hat{H}$ it self is the replica channel of H that is estimated with adaptive filter. However, before getting starting the proposed method let us see how we can design a controller for Tele-control system without considering channel equalizer. For fig.5, we can design a controller according to the plant only and neglect channels.
Assuming $H=I$ Then our complementary sensitivity function becomes

$$T(s) = (1 + P(s)K(s))^{-1}P(s)K(s)$$

and the necessity of stabilizing a MIMO system is that if and only if the complementary sensitivity function's maximum singular values is satisfying the equation (8).

For s=jω

$$\bar{\sigma}\left((1+P(j\omega)K(j\omega))^{-1}P(j\omega)K(j\omega)\right) < 1 \qquad (8)$$

3 Reduction of Multipath Channel Effects in the Closed-loop System

As we have discussed previously, existence of the multipath channel make the systemunstable and it is very hard to determine the PID parameter that satisfies the small gain theorem which has been mentioned in Equation (7). Therefore, somehow the multipath channel should be eliminated in order to get rid of the instability. Thus, equalizer is required in the receiver side of the plant for the feedforward multipath channel and another equalizer is required in the controller side for feedback multipath channel. By implementation of the equalizer we can reduce the effect of the multipath channel. However, before equalizing the received signal estimation of multipath channel is required. Estimation of multipath channel can be done by adaptive filter. After reconstructing the replica of multipath channel, inversion of the replica channel should be implemented in cascade to vanish the multipath channel. In following figure the process of the proposed system is indicated in detail.

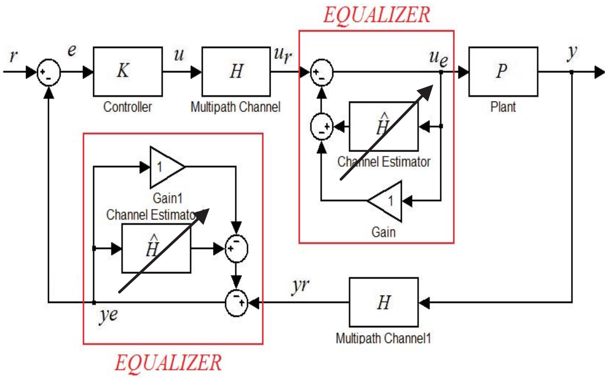


Fig. 6. Configuration of the proposed method in detail

$$y = KH\hat{H}^{-1}Pe \qquad (9)$$

$$e = r - H\hat{H}^{-1}y_e \qquad (10)$$

Afterward, we have as follows.

$$y = \left(1 + PK\left(H\hat{H}^{-1}\right)^2\right)^{-1} PH\hat{H}^{-1}Kr \qquad (11)$$

That $\left(1 + PK\left(H\hat{H}^{-1}\right)^2\right)^{-1}$ is the sensitivity function of the proposed method.

If and only if $\hat{H} = H$, then we obtain the conventional feedback control system. In other words $\hat{H}^{-1}H = I$ that $\hat{H}^{-1}$ is an unitary matrix. However, the result of multiplication of $\hat{H}^{-1}H$ is rarely becomes identity matrix since replica channel cannot realize the exact characteristic of multipath channel. Nevertheless, we can reduce the effect of multipath channel in the sensitivity function and in the open loop.

4 Multipath Channels

Basically, multipath channel is consequences of the reflected desired signal or in other words accumulation of several attenuated and delayed reference signal. Especially, this phenomenon would be occurred easily and frequently in metropolitan ambit which comprised of high density of building and so. Also, it would occur in mountainous area as well. Following shows the signal composition in time domain of reflected signal which comprise of multipath channel.

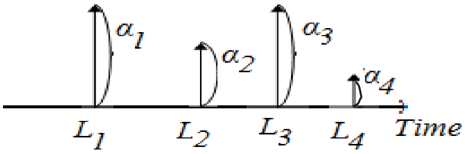


Fig. 7. Signal compositions in time domain

The mathematical model for a multipath channel H can be expressed as follows.

$$h(n) = \sum_i \alpha_i \delta(n - \tau_i) \qquad (12)$$

Where, α and τ stands forattenuation andtime delay factor of multipath channel, respectively. As it is clear in equation (11), we need to estimate the multipath channel in order to get rid of instability in the closed-loop system. Therefore, for estimation of multipath channel FIR (finite impulse response) adaptive filter is utilized. The tap number of FIR adaptive filter concerns the length of multipath channel. Hence, the length of filter should exceed the length of multipath channel. Otherwise reconstruction of replica multipath channel becomes hard. In next chapter several adaptive filter algorithms are introduced.

6 Adaptive Filters

An adaptive algorithm [6-7]is a set of recursive equations used to adjust the weight vector of replica multipath channel H automatically to minimize the effect of multipath channel in sensitivity function. Such that the weight vector converges iteratively to the optimum solution that corresponds to the bottom of the performance surface, i.e. the minimum of MSE (Mean Square Error). The Least- Mean- Square (LMS) algorithm is the most widley used among various adaptive algorithm because of its using the negative gradient of the instantaneous squared error. In general expression for H that intent to adapt itself to H. The derivation of updated weight vector of LMS algorithm can be shown as follows. Here the adaptation done in time domain so we consider the *h(n)* as the imverse Laplace trasfer of *H(s)*. As well for $\hat{h}(n)$ is the inverse Laplace transfer of $\hat{H}$ (s). For feedforward equalizer let us define the error signal in the adaptive filter e_f(n).

$$e_f(n) = h(n) * u(n) - \hat{h}(n) * u_e(n) \qquad (13)$$

According to stochastic gradient algorithm, we would have as follows. Here n and i are iteration and filter's tap number, respectively.

$$\hat{h}_i(n+1) = \hat{h}_i(n) - \mu \nabla e_f^2(n)$$

$$= \hat{h}_i(n+1) = \hat{h}_i(n) - \mu \frac{\partial e_f^2(n)}{\partial \hat{h}}$$

Calculation of gradient of square error is given by below.

$$\nabla e_f^2(n) = \frac{\partial}{\partial \hat{h}(n)} (h(n) - \hat{h}(n))(h(n) - \hat{h}(n))^T * (u(n) \times u^T(n))$$

$$= \frac{\partial}{\partial \hat{h}(n)} (h(n) - \hat{h}(n)) \times (h^T(n) - \hat{h}^T(n)) * (u(n) \times u_e^T(n))$$

$$= \frac{\partial}{\partial \hat{h}(n)} h(n) \times \hat{h}^T(n) - 2\frac{\partial}{\partial \hat{h}(n)} h(n) \times \hat{h}^T(n) + \frac{\partial}{\partial \hat{h}(n)} \hat{h}(n) \times \hat{h}^T(n)$$

$$= (0 - 2h(n) + 2\hat{h}(n)) * (u_e(n) \times u_e^T(n))$$

$$= -2(h(n) - \hat{h}(n)) * u(n) \times u_e^T(n)$$

$$= -2e_f(n)u_e^T(n)$$

Eventually, we obtain the following equation.

$$\hat{h}_i(n+1)=\hat{h}_i(n)-2\mu e_f(n)u(n-i) \qquad (14)$$

where μ is the step size or convergence factor that determines the stability and the convergence rate ofthe algorithm.

In the case of Normalized LMS, the LMS algorithm normalizes the step size with respect to the input signal power.

$$\hat{h}_i(n+1)=\hat{h}_i(n)-\frac{2\mu e_f(n)u(n-i)}{N\sigma_x^2} \qquad (15)$$

Where, $\sigma_x=\frac{1}{N}\sum_{i=0}^{N-1}u^2(n-i)$

N is tap number of adaptive filter.
Step size is now bounded in the range of 0 to 2. It makes the convergence rate independent of signal powerd by normalizing the input vector with the energy of the input signal in the adaptive filter.

7 Simulation and Results

In order to evaluate the perforamnce and stability of the proposed method, we have simulated for a system that it requried wirelss Tele-Control system. The plant is chosen to be a Drone. Here, we are going to stablize the attitude of Drone by a conventional PID contoller and reduce theeffect of multiplath channel by equlizer. Attidude control consist of roll, pitch and yaw. The plant is MIMO 6 dimentional plant matrix with 3 inputs and outputs. As we disscused, in the case of Wireless Tele-Control System we have multipath channel certainly. In this situation we can consider that in the closed-loop system we would have two idenditical multipath channel, one for feedforward and the other for the feedback. In the result, in order to see the charachtresitc and influences in control system we show the singular values of open loop and the closed-loop system without equalizer. Thereafter, we simulate the step response of the closed-loop system with equzlier to confirm the stability and feasibility of equalizer which is implemented in closed-loopsystem. Following shows the Conditions of Simulation.

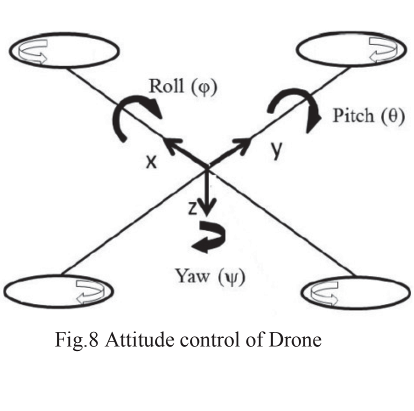


Fig.8 Attitude control of Drone

・Doyle Expression of the simplified atitude dynamics of Drone:

$$P(s)=\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

$$=\left[\begin{array}{cccccc|ccc}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 5.1282 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 7.4074 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 7.4074 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{array}\right]$$

where A , B, C and D are plant, input , output and direct matrices , respectivley .

・Controller of attitude:

$$K(s)=K_P\left(1+\frac{1}{T_i s}+\frac{T_d s}{as+1}\right)I_{3\times3}$$

$$=0.0005\left(10+\frac{0.01}{s}+\frac{2s}{0.01s+1}\right)I_{3\times3}$$

・Channel Specification (with 10 taps M=10) :

$$H(s)=\sum_{i=1}^{M}\alpha_i e^{-sL_i}$$

.

Feedforward channel and feedback channel are assumed to be identical.

Where, attenuated and time delay factor are indicated below.

$\alpha_i=rand(i)e^{\frac{-0.1i}{M}}$, $L_i=0.5i\times sort(|rand(i)|)$

That rand(.) is uniformly distributed random numbers.

Results

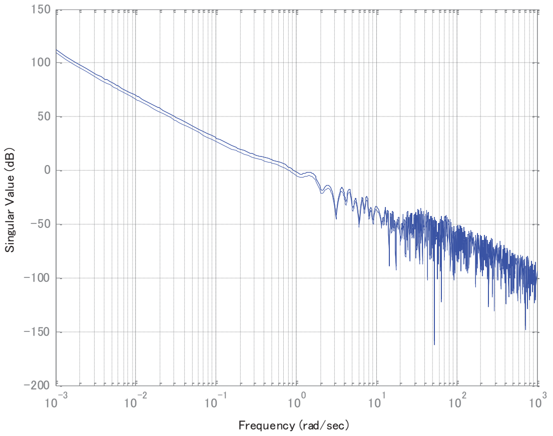


Fig.9 SVD of open loop without equalizer

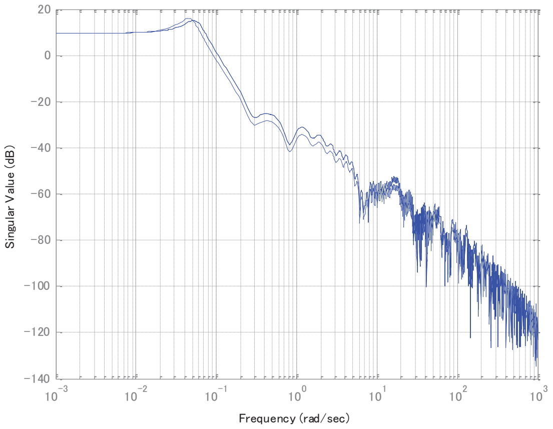


Fig. 10 SVD of closed loop without equalizer

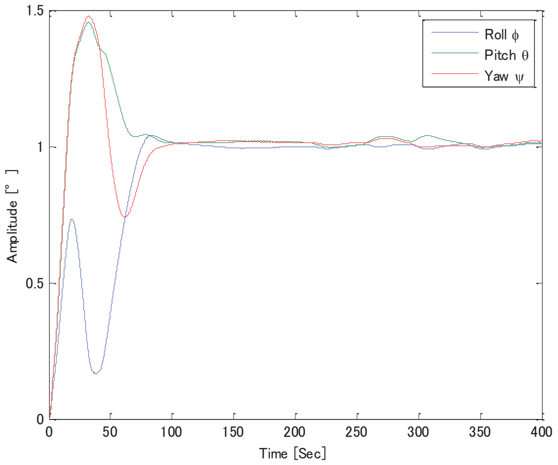


Fig. 11 Step response of closed-loop with equalizer

Fig. 9 shows the singular values of open loop that contains multipath channel without equalizer. As it is clear singular values in haigh frequencies, it flucuates frequently. Therefore, these kinds of charactrestic for open loop is not desired since it may casuse the intability of the closed-loop system. In Fig.10 we can see the singular values of the closed- loop system without equalizer. In this figure as well as Fig.9 in the high frequencies singular values fluctuates and in low frequencies it contain high gain that is not satisfying the stability condition. Thus, it would become unstable and sensitive to disturbances. As it is shown in Fig.11 eventhough roll and pitch angle have 50% overshoot, the step response of the closed-loop with equailzer is stable. Thus we could confirm the stablity of closed loop system even multipath channel exist in system. However, in this study we were not able to analyze the singular values of the closed loop and open loop with equilizer since equalizer is a time variant system . Therefore as a future work we are going to obtain the singular values with equalizer. In addtion, in this paper singular values of open loop and the closed-loop with equalizer could not be obtained due to equalizer. Since equalizer is a variant system singular values cant not be determined. Therefore, we also going obtain the singular values and even stability margin of system with equalizer in order to anaylize in frequency domain.

8 Conclusion

In this paper we implemented equalizer in Wireless Tele-Control system in order to get rid of instability in the closed-loop system which cuased by multipath channel. As a result we could confrim the stability of the closed-loop system with step response. However, the performance is not accepatble due to existence of overshoots. Moreover, in the actual case parameters of plant may be perturbed due to environmental and phisycal conditions. Therfore as a future work enhancment of performance and internal stability of MIMO system including parametres perturbation should be considered.

Acknowledgment

This work was supported by Japan Society for Promotion of Science Grant Number 25420924. We wish to deeply express our gratitude to their supports.

REFERENCES

[1] R. C. Dorf, R. H . Bishop " Modern d Control System ", Prentice Hall2002

[2] Witold Pedrycz, " Robust Control Design an Optimal control Approach " , Wiley 2007

[3] R. Oboe, K. Natori, K. Ohnishi, "A Novel Structure of Time Delay Control System with Communication Disturbance Observe" International Workshop on Advanced Motion Control, AMC '08. 10th

[4] F. Asharif, S. Tamaki, T. Nagado, T. Nagata and M. R. Alsharif, " Design of Adaptive Friction Control of Small-Scaled Wind Turbine System Considering the Distant Observation" ICCA, LNCS Springer pp213-221, Nov. 2012.

[5] Guillermo J., Silva Aniruddha Datta, S.R Bhattacharyya,"PID Controller for Time-Delay System" Birkhauser, 2004

[6] Kong-Aik Lee, Woon-SengGan and SenM.Kuo,"Subband Adaptive Filtering Theory and Implementation," 2009, John Wiley & Sons, Ltd

[7] F. Asharif, S. Tamaki, M. R. Alsharif, H. G. Ryu"Performance Improvement of Constant Modulus Algorithm Blind Equalizer for 16 QAM Modulation"InternationalJournal on Innovative Computing, Information and Control, Vol. 7, No. 4, pp.1377-1384, April 2013.

Optimization Modeling in a Smart Grid

Damian Lampl, Md Chowdhury, Pranav Dass, Kendall E. Nygard[1]
Dept. of Computer Science
North Dakota State University
Fargo, ND, USA
{Damian.Lampl, MD.Chowdhury,Pranav.Dass,Kendall.Nygard}@ndsu.edu
1. Corresponding Author

Vahid Khiabani
Dept. of Construction Management and Operations Management
MSUM Moorhead
Moorhead, Minnesota, USA
Vahid.Khiabani@Mnstate.edu

Abstract—**Communication and control in a fully realized Smart Electrical Grid involves heterogeneous wired and wireless networks working cooperatively, supporting data streams among many types of sensors. We address the self-healing problem, in which the goal is to intelligently automate corrective actions when a disruption to Grid operation occurs. Such actions include redirecting electricity flows along alternative pathways, and selectively tripping breakers. A primary objective of a self-healing method is to prevent cascading failures. Motivated by the need for corrective actions in self-healing to produce efficient and reliable grid operations, we formulated and developed an optimization model that generates sets of high performance electricity flows in an arbitrary Grid configuration. The model is a Capacitated Transshipment Problem (CTP) that we solve using a very fast and customized algorithm. The versatility of the model in supporting multiple performance metrics and the speed achieved in generating sets of optimal electricity flows makes the model useful in evaluating self-healing strategies.**

Index Terms — smart electrical grid, self-healing capacitated transshipment problem, linear programming, network flow optimization

I. Introduction

A Smart Grid is an electrical generation and distribution system that is fully networked, instrumented, and automated [2]. From a communication network perspective, there are three distinct levels. At the most distributed level, within a demand site such as a home, a wireless network is typically used to interconnect appliances and various other devices and systems. Intelligent control is called for to regulate consumption of energy for such things as heating water and living spaces. At a second level, smart meters receive information from the low level network, and are in turn themselves networked within neighborhoods. Other devices are also in the neighborhood network with the smart meters and form the distribution system. Wireless networking is typical within a neighborhood. Finally, a wide area network (WAN) interconnects utility owned and operated equipment and systems, such as distribution substations, power plants, and long-haul transmission lines. Multitudes of sensing devices, such as Phasor Measurement Units (PMUs) that report detailed waveform information, are deployed throughout the grid. Self-healing functionality relies heavily on streaming sensing data to drive models and analytics aimed at choosing effective actions for maintaining safe, efficient, and reliable grid performance.

An electrical grid experiences faults caused by numerous factors such as failures of generators or routers; or power lines damaged by weather events or vandals. Faults can propagate through the connected networks of an electrical grid and result in remote butterfly effects. The effects can be cascading failure and consumer power outages over wide areas. It is not possible to prevent such faults [3], but their effects can be minimized by isolating fault sources with sensor information and taking corrective actions. Corrective actions taken by power companies traditionally are mostly focused on scheduling and dispatching crews and equipment to make repairs and replace devices or connections in the grid infrastructure. However, human decision making and actions often cannot be fast enough to avoid significant downtimes for consumers, providing a basic motivation for intelligent automation in a Smart Grid.

One strategy for mitigating the effects of malfunctions in the grid is to dynamically reroute power to physically avoid trouble spots. However, rerouting power can itself be a source of problems, as power lines that are overloaded or nearly so can result in cascading failures over wide areas. Thus, control decisions and actions to reroute power must be done with full consideration of possible ramifications distributed in the grid infrastructure. The software tool that we have developed serves the purpose of rapidly determining optimal distribution patterns and dispatches of power along available channels, including the reporting of metrics that evaluate costs and quality of service.

Another important consideration in optimizing grid operations is the emerging deployment of microgrids. A microgrid is a local energy generation system, powered by small-scale generators, batteries, or alternative sources like solar panels. A microgrid is coupled with a primary grid, and can be disconnected as needed so that a local area can function as an island during an emergency, or to cut costs. Thus, microgrids provide a decentralized control function that can help maintain quality of service. Our self-healing model supports the use of microgrids.

The mathematical model that we have developed is a linear programming optimization model with a special structure that can be conceptualized as an abstract network with nodes and arcs. As described in the literature, the model is a Capacitated Transshipment Problem (CTP). One type of parameter for the model pertains to known data on grid topology such as locations of sites where power is generated or demanded and interconnection nodes. Another type of parameter pertains to the capabilities of grid devices to do useful work, such as capacities of transmission lines to carry power and of power plants to generate electricity. The output of the model is the values of variables that specify dispatching decisions, flows of power, and performance metrics. Under conditions of normal operation or of disruption, data from distributed sensors are streamed to populate the model and trigger computational devices within the Grid to solve the model. Our customized model solver is fast and modest in terms of computational resources, so it can be preinstalled on computational devices distributed in the Smart Grid. General linear programming solvers could be applied to the model accurately, but would have the disadvantage of requiring unacceptably long computation times.

This remainder of paper is organized as follows. Section III provides a brief overview of linear programming modeling. In section IV, the CTP formulation is presented and is applied to the Smart Grid. The algorithmic process for solving the model is detailed in section V. Section VI provides the results and analysis, followed by the conclusion in section VII.

II. Objective

Representing the Smart Grid network using a CTP model allows multiple different cost and network flow related problems to be easily solved. To make a Smart Grid self-healing, whenever a critical failure is detected, the CTP solver can be used to find an optimal and inherently feasible redirected path for redistributing energy throughout the grid, resulting in minimizing customer outages.

Apart from the self-healing aspect of the Smart Grid, the CTP solver offers other key benefits such as its ubiquitous availability to any machine or mobile device connected to the internet, regardless of the operating system. Since the CTP Solver is able to connect to a database as well as read XML files, it could be easily integrated with other Smart Grid systems such as failure notification solutions, providing automatic optimal electric flow rerouting based on the supplied network topology of available nodes and arcs. Since arc capacities are taken into consideration, the cascading failure dynamic could possibly be avoided by ensuring network flow is feasibly rerouted.

The CTP solver incorporates an object-oriented approach, thereby ensuring ease of use and maintainability for its users. This further allows the developers to quickly determine the application areas that need updates and implement them in a timely and efficient manner. The CTP solver automates its processes so that the user does not need to learn a new application-specific language or syntax to follow them. The CTP solver involves use of bidirectional arcs in its design, thus allowing the network flow in both directions between a node pair, resulting in effectively limiting the network file size and memory requirements of a dataset containing all bidirectional arcs.

In this work, we have developed the mathematical models based on the design goals of the CTP solver we have already discussed in order to determine the optimal network flow of a given Smart Grid network.

III. LINEAR PROGRAMMING MODELS

Linear Programming models are formulated to maximize or minimize an objective function that is devised to measure performance of a solution. Linear constraints in the form of equations or inequalities are supported. Linear programming is an exact model, in that once solved, the solution is guaranteed to be the very best (genuinely optimal) as measured by the objective function. In some applications heuristic models are applied as an alternative, but such models do not guarantee optimality. The three basic steps given below are followed when formulating a linear programming model.

1. Determination of the decision variables
2. Formulating the objective function
3. Formulating the constraints

The decision variables are the quantities that the model seeks to calculate, providing the solution to the problem. The objective function is the expression that the modeler wishes to optimize, and the constraints are limitation requirements. The general form of a linear programming model is given below in Figure 1 [4].

Parameters

C = [c_j] = Vector of costs or value measures per unit of decision variable value
A = [a_{ij}] = matrix of technological coefficients that measure the rate at which variable x_i consumes resource j.
b = [b_i] = Vector of coefficients that measure constraint limitations of resource

Variables

Z = Objective function that measures the value of a solution
x = [x_i] = Vector of decision variables

Formulation

Optimize $z = c_1x_1 + c_2x_2 + . . . + c_nx_n$

Subject To:

$$
\begin{array}{l}
a_{1,1}x_1 + a_{1,2}x_2 + . . . + a_{1,n}x_n \{\leq, =, \geq\} b_1 \\
a_{2,1}x_1 + a_{2,2}x_2 + . . . + a_{2,n}x_n \{\leq, =, \geq\} b_2 \\
\qquad\qquad\qquad\qquad . \\
\qquad\qquad\qquad\qquad . \\
\qquad\qquad\qquad\qquad . \\
a_{m,1}x_1 + a_{m,2}x_2 + . . . + a_{m,n}x_n \{\leq, =, \geq\} b_m \\
\qquad x_1, x_2, . . . x_n \geq 0
\end{array}
$$

Figure 1: Linear Programming Model General Form

When instantiated to model electricity distribution in the Smart Grid, we think of the decision variables as representing flows of power, and resource constraints as representing capacity limitations on devices and power lines.

IV. THE CAPACITATED TRANSSHIPMENT MODEL

The CTP is conceptualized as a network problem with supply and demand nodes, transshipment nodes, and connective arcs. The basic concept is to find an optimal set of flows that transfers units from supply nodes through the network to meet requirements at the demand nodes, conserving flow at transshipment points, and without violating capacity constraints.

The CTP is presented in algebraic form in Figure 2.

Parameters

c = [c_{ij}] = Measures of the costs or values per unit of flow through arcs indexed by tail and head nodes i and j
u = [u_{ij}] = Vector of flow capacities on arcs
l = [l_{ij}] = Vector of lower bounds for flow on arcs
b = [b_i] = Vector of supplies and demands at nodes indexed by i. Positive values for supplies, negative for demands

Variables

Z = Objective function that measures the value of a solution
x = [x_i] = Vector of optimal flows

Formulation

Minimize $z = \Sigma c_{ij}x_{ij}$

Subject To:

(1) $x_{ji} - x_{ij} + b_i = 0$ for all arcs i,j
(2) $x_{ij} \geq 0$ for all arcs i,j
(3) $x_{ij} \leq u_{ij}$ for all arcs i,j
(4) $x_{ij} \geq l_{ij}$ for all arcs i,j

Figure 2: CTP Standard Form

The objective function is to minimize the total of all arc flows multiplied by their costs. Constraint (1) ensures flow balance at every node by ensuring that total flow out of a node is the same as the total flow in, adjusted for supplies or demands at the node itself. These constraints also ensure that supply units are fully distributed from all supply nodes to all demand nodes, creating flow balance for the entire network. Constraint set (2) ensures that all arcs have a non-negative unit flow. Constraint (3) ensures that no arc capacities (upper bounds) are violated. Constraint set (4) ensures that no arc lower bounds are violated. In a self-healing application to the Smart Grid, a candidate grid configuration, even one that reflects serious disruptions or damage, can be optimized. This then supports a best possible means of running the grid under adverse conditions. The special CTP formulation allows for a customized solver with highly desirable characteristics to be developed as detailed in the following section.

A standard Smart Grid test problem is the IEEE 14-Bus System, illustrated in Figure 3. The corresponding flow network configuration that can be modeled as a CTP is illustrated in Figure 4.

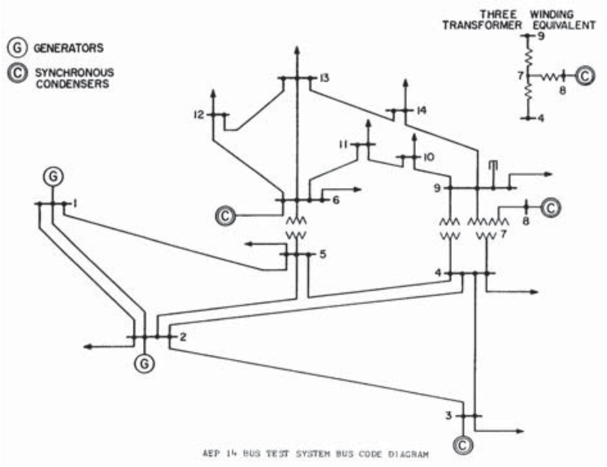

THREE WINDING TRANSFORMER EQUIVALENT
G GENERATORS
C SYNCHRONOUS CONDENSERS
1
2
3
4
5
6
7
8
9
10
11
12
13
14
AEP 14 BUS TEST SYSTEM BUS CODE DIAGRAM

Figure 3: IEEE 14-Bus Test System Diagram

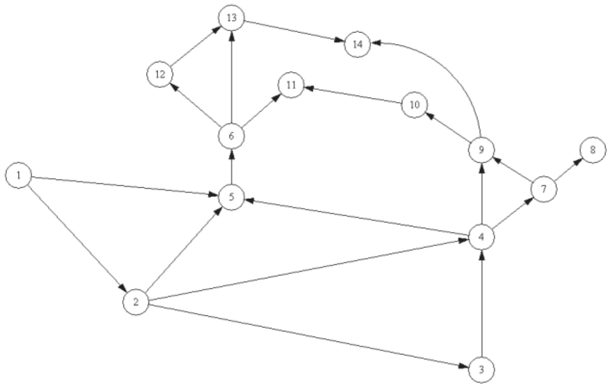


Figure 4: IEEE 14-Bus Test System Network Representation

V. MODEL SOLVER

There are currently a number of solvers that have been developed and are available for producing optimal solutions to general linear programming problems. However, we needed a solver that would scale extremely well and produce solutions in near real time. Our custom solver software was written as an ASP.NET C# application using a simplex algorithm modified to exploit the special structure of the model. The powerful characteristic of the CTP that we exploit is that any linear programming basis corresponds to a spanning tree of the network representation. This enables simplex basis changes to be carried out in all integer arithmetic on graphical tree structures, greatly expediting the computations when compared with working inverses of basis matrices. Following the general scheme for applying the simplex method, we carried out the following five steps:

1. Initialization
2. Reduced Cost Calculation
3. Cycle Creation
4. Basis Update
5. Repeat Steps 2-4 Until Optimality

1. Initialization

An XML file or local database is populated with sensor readings and pre-established topological information. The initialization step reads the data and creates a candidate solution basis tree, as illustrated shown in Figure 5. From an artificial root node, a directed arc is connected to each actual network node using penalty values for the arc costs, which will force them out of the basis early. The absolute values of supplies (or negative demands) at the actual nodes are used to set initial values of the arc flows from the artificial node. In the algorithm, these artificial arcs are forced from the basis tree one by one due to their large penalty costs, leaving only actual network arcs in the final, optimal solution.

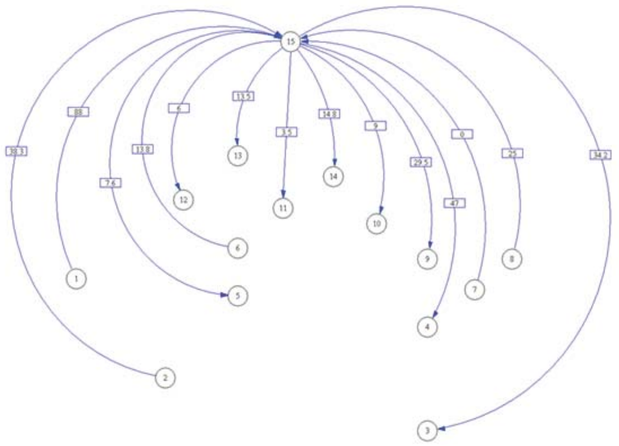


Figure 5: Example Initial Basis Tree

Node potentials are also calculated for the initial basis tree and used to determine the best candidate arc not already in the basis tree, to replace a basic arc. The node potentials are the dual variables in linear programming terms, represented algorithmically as the sum of the arc costs following the path from any given node back to the root node in the basis tree.

2. Reduced Cost Calculation

The reduced cost is the per unit rate at which the objective function would change if a given non-basic arc were inserted into the basis tree. If the evaluation metric is a cost that should be minimized, the best reduced cost belongs to the arc that will potentially lower the total network cost by the greatest per unit amount. For any given non-basic arc, the reduced cost is calculated by subtracting the node potential of the arc's tail node and its cost from the node potential of its head node. In effect, this evaluates an alternative pathway for power to flow.

If no candidate arc is found to reduce the total cost of the network, then the solution is optimal. Otherwise, the arc with the best reduced cost is chosen to enter the basis tree. At each step, both upper and lower bound on arc flows must be evaluated in order to maintain a feasible solution.

3. Cycle Creation

By definition, the basis tree is a connected graph with no cycles. This means there is a path between any two nodes, but not a path from any node to itself. When a non-basic arc is added to the basis tree, a cycle is created and an arc must be removed to preserve the basis tree's acyclic property. An example cycle is shown in Figure 6.

Using the depth of the entering arc's nodes in the basis tree, the cycle is created by following the back path from each entering arc's node to the root node of the basis tree. The node depth allows the two back paths to be traversed in pairs during the same iteration, starting at the deepest node in the cycle and working up the basis tree until the two back paths meet at the same parent node or the root node is reached, either of which completes the cycle.

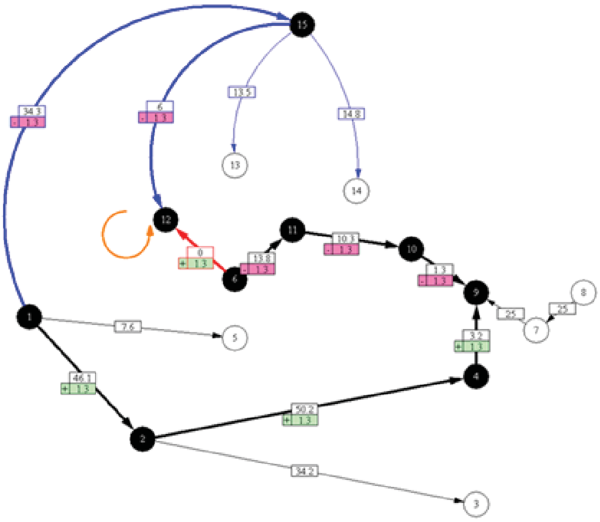


Figure 6: Example Cycle

As each arc is added to the cycle, its maximum feasible flow change is calculated based on the arc's direction in relation to the cycle created by the entering arc. This value is the largest flow that could be added or subtracted from a same- or opposite-cycle direction arc, respectively, without violating the arc's flow capacity or lower bound. Using this flow value, the algorithm adjusts the flow solution in the new basis tree to move the current solution incrementally toward the optimal solution. The solution adjustment respects upper and lower bounds at every step, ensuring that feasible solutions are found at every increment. These feasible solutions can be evaluated using auxiliary criteria that might be imposed when the Smart Grid experiences equipment failures. Robustness of the solution is one such auxiliary criterion.

4. *Basis Update*

Once a new arc enters the basis tree and one arc leaves, a new basis tree is determined. To fully specify the new tree, the node potentials and depth values are updated. Once the basis tree has been updated, it is ready to be used for the next iteration if the optimal solution has not yet been reached.

5. *Repeat Steps 2-4 Until Optimal*

At Step 2, the reduced cost information for all non-basic arcs is calculated. If there are no arcs that can improve the solution, optimality is guaranteed. For the example, the optimal basis tree is illustrated in Figure 7.

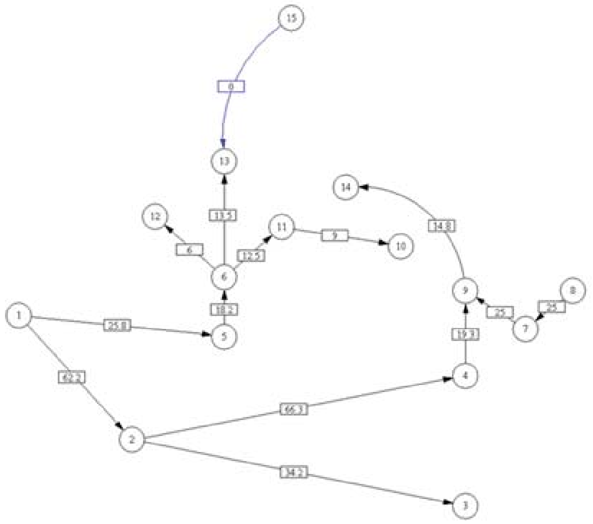


Figure 7: Example Optimal Solution

VI. Results and Analysis

The key question is whether the customized solver can compute solutions fast enough to function in a self-healing system. For all of our test problems an end to end process was followed, starting with a populated data set of parameters and network topology, following all steps to generate a full optimal solution, and reporting the results.

Several abstract networks with known solution obtained from the literature were tested to ensure that the solver accurately obtained the optimal solution [7].

To test the method on example power grid networks, test problems for the IEEE 14-Bus System, IEEE 30-Bus System, IEEE 57-Bus System, and IEEE 118-Bus System were downloaded from the University of Washington Electrical Engineering website. Various objective function evaluation metrics were evaluated for each problem, to generate realistic Smart Grid scenarios. The base case utilized simply used physical distances between nodes, to essentially determine overall shortest paths for electricity flow to follow. All computation tests were performed on an Intel Core 2 Quad 2.67GHz processor with 4GB DDR2 800 RAM, running on Windows Vista Ultimate x64. A local virtual directory was created for the solver, it using IIS and running the .NET 4.0 framework. Each suite of test problems was run multiple times with parameter changes, and average computation times recorded. The 14, 30, and 57 bus systems solved in well under .1 seconds and the 118 bus system solved in less than .2 seconds. Although we have not yet tested the solution algorithm on truly large problems, our computational experience thus far suggests that the algorithm will scale well. In any case, the procedure is clearly of potential value for dynamic power dispatching and allocation in smaller power grid components, such as microgrids. Another advantage is the ease of setting new parameters for the solver within a self-healing system. More specifically, when devices and lines modeled by nodes and arcs in a functioning Smart Grid system fail or malfunction, an updated network topology can easily be provided to the CTP solver. This produces the capability of quickly finding high performance ways to redistribute power throughout the network, meeting electricity demands with minimal interruption of service.

VII. Conclusion

In this research, a customized CTP solver was developed as a tool for formulating and analyzing the performance of a Smart Grid network. Multiple evaluation metrics are supported, allowing a diverse set of problems to be studied using the same solver. Solutions are generated with little required computation time, opening potential for use in self-healing Smart Grid situations which inherently demand near real-time results. These solutions are guaranteed to be optimal, ensuring the best possible flow of electricity throughout the network according to the provided parameters.

As sensors in the electrical grid become more sophisticated and high bandwidth communication networks are in place, the model provides the potential to receive the data streams and generate operational grid actions through the computational efficiencies of linear programming to minimize the effect of infrastructure failures.

References

[1] Kaplan, S. M. (2009). Smart Grid. Electrical Power Transmission: Background and Policy Issues. The Capital.Net, Government Series. pp. 1-42.

[2] Solanki, J., S. Khushalani, and N. N. Schulz, “A Multi-agent Solution to Distribution Systems Restoration,” IEEE Transactions on Power Systems, vol. 22, no. 3, pp. 1026–1034, 2007.

[3] Nygard, K. E., S. Ghosn, M. Chowdhury, D. Loegering, R. Mcculloch And P. Ranganathan, "Optimization Models For Energy Reallocation In A Smart Grid," In IEEE Infocom 2011 Workshop On M2mcn, 2011

[4] Ignizio, J. P., Linear Programming, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1994.

[5] Bazaraa, M., J. Jarvis, and H. Sherali, Linear Programming and Network Flows, 4th Edition, Hoboken, New Jersey: John Wiley & Sons, Inc.,2009.

[6] Brown, G. G. and G. H. Bradley, "Design and Implementation of Large Scale Primal Transshipment Algorithms," *Management Science,* vol. 24, no. 1, pp. 1-34, 1977.

Evaluation of a remote sensing architecture for precision irrigation using open hardware technologies

**A. Rafael Boufleuer[1], B. Alfredo Del Fabro Neto[1], C. Bruno Romero de Azevedo[1], D. Iara Augustin[1],
E. João Carlos D. Lima[2], F. Mirta T. Petry[3], G. Reimar Carlesso[3],
H. Laudenir J. Basso[3], and I. Douglas Haubert[3]**

[1]Informatics Graduation Program, Federal University of Santa Maria, Santa Maria/RS, Brazil
[2]Department of Languages and Computer Systems, Federal University of Santa Maria, Santa Maria/RS, Brazil
[3]Department of Rural Engineering, Federal University of Santa Maria, Santa Maria/RS, Brazil

Abstract— *Considering that irrigation consumes the major percentage of water in agriculture resources, there is a growing need in improving the irrigation water management in the world. This work proposes a remote sensing architecture for irrigation and its contributions are: the development of two prototypes of a moisture and precipitation meter based on the proposed architecture using open hardware technologies and a comparison between the resistive low-cost sensor produced (Federal University of Santa Maria, Brazil) and a high-accuracy and high-cost frequency domain reflectometry (FDR) sensor (CS616 - Campbell Scientific, United States). The results obtained were satisfactory, where it was verified that the architecture is viable, meeting the requirements to which it has proposed. In addition, the comparison performed showed a determination coefficient of up to 95 % between the resistive soil moisture sensor and the soil moisture sensor CS616 of Campbell.*

Keywords: Precision Irrigation, Architecture, Open Hardware.

1. Introduction

Agriculture is the largest consumer of water in the world (more than 70%), making the component of water at the same time the most abundant and the most limiting factor for the crop yields [1]. Considering that irrigation is responsible for the largest percentage of the water used in agriculture, there is a growing scientific interest in precision irrigation for its potential in improving the crop productivity and increasing the efficiency of resources usage, such as water and energy in irrigated agriculture [2]. This irrigation should be based on all the available environmental information in order to evaluate the amount of irrigation and the time it should take place, avoiding water waste by excessive irrigation and/or reduction in production by losses due to sub-irrigation [3].

Hargreaves et al. [4] emphasizes irrigation like one of the main instruments to maintain the farmer in the field and allow his economic and social development. To make this possible, it is necessary to provide options for irrigation management that may contribute to both conservation of natural resources and sustainable development [5]. One way to aid the development of irrigation is by using systems that consider information about weather, plants and soil in order to improve the water management in the irrigation process, supply the demands of the crops and increase the productivity in the field. In this sense, with the necessity of incorporating technology in the field, many studies that use open hardware technologies like the Arduino platform are being developed [6], [7], [8], which is a simple prototype development platform that enables the development of low-cost systems [9].

Several studies have also been conducted to evaluate new remote sensing-based soil moisture estimates in [10], [11], [12], which allows the gathering and interpretation of data from distance. Beyond that, several techniques to perform the soil moisture measurement have been developed ([13] for a review), including the measuring by the soil resistance with resistive sensors and by techniques that uses the frequency domain reflectometry (FDR), both techniques are compared in this paper. There are also many studies that are addressing the evaluation and comparison of the resistive and FDR sensors [14], [15], showing a high correlation between the soil resistivity and the soil moisture measurements [16]. Therefore, one can notice that the use of resistive sensors can be a viable alternative for monitoring soil moisture.

The main objective of this work is the development of two prototypes for a moisture and precipitation meter based on the remote sensing architecture proposed using open hardware technologies. For the validation of the proposed architecture, a study case was performed to verify the correct functioning of the architecture's components and to make a comparison between the moisture sensors. This paper is structured as follows: In Section 2 we present the main concepts related to remote sensing and the equipments for collection and transmission of data. In Section 3 is presented the proposed remote sensing architecture and the functioning of its components. The description of the study case and the alternatives to soil moisture adjustments are shown in section 4. In Section 5, we draw our final considerations and future work.

2. Core Concepts

In order to determine when irrigate and the amount of water to be applied for irrigation one can use the monitoring of the plant, the climate or the soil [3]. The management alternative used in this work is related to the quantification of the water content present in the soil by the usage of sensors that measure the soil moisture. In order to achieve this, the equipments used must be capable of gathering the data with precision and good response time due to the necessity of periodic and representative data of the soil moisture for a given region. This kind of monitoring was chosen because it has a good cost-benefit when we consider the spending on equipments and the precision in the soil moisture measurements.

2.1 Remote Sensing and Open Hardware

Several technologies are being used to increase the agricultural production and reduce its impact in the environment. The irrigation and the sensoring technologies have been used in agriculture as means to achieve such goals. With the remote sensing, it is possible to gather and interpret data remotely, allowing improvements in the control of the agricultural production [17]. Furthermore, independently of the irrigation techniques used, the irrigation systems are normally located far away from the cities, having a limited telecommunication infrastructure and costly services. These problems make it difficult for most of the farmers to perform the remote control and monitoring of their irrigation systems [18]. However, by using data transmission technologies such as GPRS (General Packet Radio Service) and the GSM network (Global System for Mobile Communication) it is possible to transmit data from the field with a relatively low-cost. This way, in the proposed prototypes we use a GSM/GPRS SIM 900 shield.

Open Hardware is a concept that is related to the sharing of the structure of physical objects to the community similarly to open source softwares, allowing for everyone to use and modify it. A Open Hardware solution used in our work is the Arduino platform, which is being used in sensing projects and enables the acquisition of data from various types of sensors in the environment [9].

2.2 Sensors

A sensor is a device that responds to a physical or chemical stimulus like heat, pressure or movement in a specific and measurable manner [19]. In the moisture and precipitation meters proposed we use sensors to measure temperature, soil moisture and a pluviometer to verify the precipitations. In this work, the soil temperature was measured using the Waterproof DS18B20. The soil temperature is an important information, because it influences directly in the seed germination, in the growth rate of plants and in the conditions for harvesting crops [20]. The pluviometer used to monitor the precipitations was the model WS1080.

For the measure of the soil moisture, we used resistive and capacitive sensors. The resistive sensors measure the levels of moisture based on the electrical resistance of the soil: as the water level in soil increases the resistance of the metal decrease, that is, the conductivity of the metal rises [21]. The low-cost resistive sensor used in this work was manufactured in the Federal University of Santa Maria. The capacitive sensors measure the soil moisture by the dielectric constant of the soil which is linked to the content of moisture in the soil: as the soil moisture changes, the dielectric constant also changes and is subsequently related to the volumetric water content of the soil [22]. The most widely used techniques for this type of measurement is the time domain reflectometry (TDR) and the frequency domain reflectometry (FDR). The capacitive sensor used in this work is the FDR sensor CS616 of Campbell [23], and it was used as reference in the comparison of the resistive sensors. The used sensors are presented in figure 1, the CS616 sensor is the on in the top.

Normally, the academy considers that the resistive sensors have a lower accuracy or that the interest in them has declined because of advances in other methods [13]. However, there is a significant difference in their prices, where the resistive sensors are more affordable. If a good calibration in these sensors is made, then it is possible to use them as a good alternative to make a feasible low-cost moisture and precipitation meter.

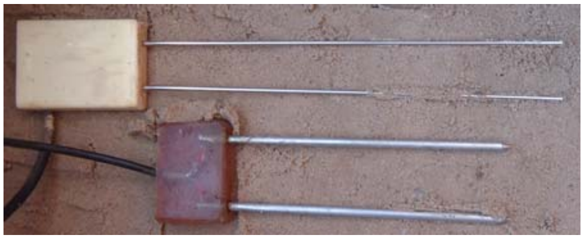


Fig. 1: Moisture sensors used.

2.3 Prototypes

The Figures 2 and 3 shows the prototypes for the moisture and precipitation meter, where Figure 2 shows the prototype with the Arduino and the GSM/GPRS board connected with a RTC (real time clock) DS1307 and Figure 3 shows the alternative prototype with an Arduino, a SD reader and a RTC. This prototype was designed for places where the GSM network is not available. In this case, it is necessary to manually collect the data stored in the SD card, so later it can be stored on the server.

3. Proposed architecture

The evaluated architecture is presented in Figure 4, and in sections 3.1, 3.2, 3.3 and 3.4 the layers composing it are briefly presented.

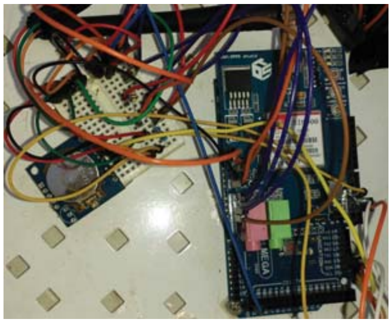


Fig. 2: Meter with a GSM/GPRS board.

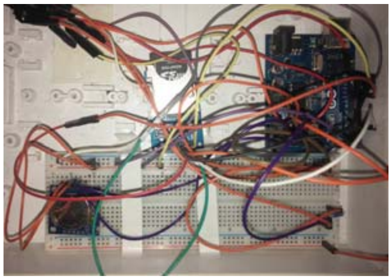


Fig. 3: Meter with a SD reader.

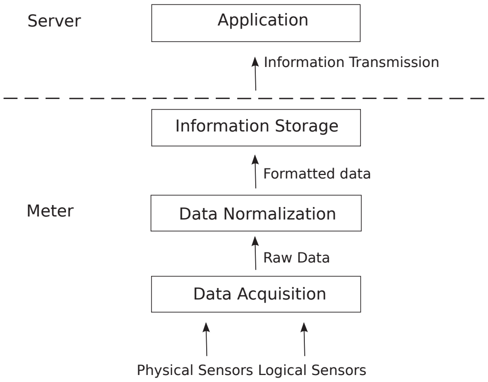


Fig. 4: Proposed architecture.

3.1 Data Acquisition

The main objective of the data acquisition layer is to gather data from physical and logical sensors. The data from logical sensors can be gathered by any external source of information (e.g., date and time information collected from the GSM network to update the RTC whenever the arduino is turned on). The raw gathered data is passed to the Data Normalization layer.

3.2 Data Normalization

The data normalization layer is responsible for formatting the raw data in a defined format so it can be used in the upper layers. This layer needs to know the format of the messages received from the sensors, which can be from different types and manufacturers. Thus, to begin using a new type or model of sensor from some manufacturer, is it necessary to develop a specific parser to obtain the sensor's data, because the data obtained from this new sensor is not recognizable by the Data Normalization layer beforehand. After the formatting is done, the data is passed to the Information Storage layer.

3.3 Information Storage

The Information Storage layer is responsible for receiving the formatted data from the Data Normalization layer and storing these information in the EEPROM or in the SD Card in order to maintain a history of such information. These information are important to determine when and where an event occurred and to assist in localized irrigation (e.g., rainfall occurrence). After the storage, the information is sent to the server.

3.4 Application

The application layer is located on the server and is responsible for receiving information from the Information Storage layer and performed the calculations required with it. Furthermore, in this layer, the information is visually represented (e.g., plots).

4. Study case

In this study case, we collected data from sensors connected to the prototype (moisture and temperature of soil and precipitation) and from the Campbell CS616 soil moisture sensors, that were used to compare the values with the resistive soil moisture sensors. Besides that, we used the environmental temperature and precipitation from the INMET (National Institute of Meteorology) station located in Santa Maria/RS, Brazil [24] and the precipitation from the Irriga System, which offers a set of monitoring and management services for irrigation [25]. The values for the precipitation and the environmental temperature from the INMET station as well as the precipitation values for the Irriga System were manually gathered for comparison purposes.

The sensors were installed within a sandbox with two holes at the bottom for water flow. The prototype was installed in a rural area of the Federal University of Santa Maria and gathered approximately 2000 values of sensor data in 3 weeks. The pluviometer gathered the rainfall in this period, the soil temperature sensor was installed horizontally at a depth of about 5 centimeters (cm) and the moisture sensors were installed horizontally and arranged in two layers, one at 10 cm from the surface and the other at 20 cm from the surface of the sandbox. In both layers, we placed one CS616 and one resistive sensor so we could correctly make the comparison of the moisture sensors. During the data gathering, the prototype was directly connected to the electric energy in order to disregard problems related with battery power consumption. The prototype with the GSM/GPRS board used in the study case is presented in figure 5.

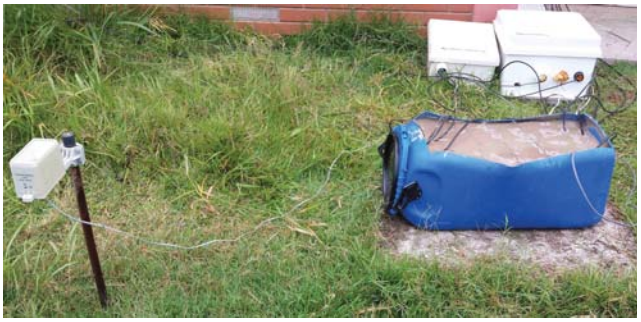


Fig. 5: Data gathering in the sandbox.

4.1 Soil moisture with raw data

This section presents the graphs generated with the raw data from resistive sensors in comparison with Campbell CS616 sensors. The CS616 sensor's data was collected with a Campbell CR1000 datalogger, which provides soil moisture data in the range 0-1 (the closer to 1, the moisture the soil is). The range of the obtained values for the resistive moisture sensors by the analog outputs of the Arduino vary in the range 0-1023. So, it is necessary to perform a normalization of these values for them to lie between the range of 0 and 1 in order to better compare the sensors values. The graphs presented similar variations, however it was necessary to perform an inversion of 180 degrees in the resistive sensor values, because values closer to 0 for the resistive sensors represent moisture soils, as opposite to the CS616 sensor's values. This inversion and normalization of the resistive sensor's values was performed according to equation 1:

$$UF = 1 - \frac{U}{L} \tag{1}$$

Where UF is the final soil moisture, U is the soil moisture gathered and L is the maximum value that can be obtained in the arduino. The resultant values of the equation vary between the range 0-1: the closer to 0, the least amount of water in the soil; the closer to 1, the greater the amount of water in the soil. The pluviometer values from the prototype were normalized between 0.0-0.1 in order to facilitate the understanding of the soil moisture sensors' behavior in the graphs. The amount of precipitation in millimeters is presented in figure 8.

The data collection was performed four times per hour. Then, we used the average of these values for the comparison in the graphs. One can notice in the figures 6 and 7 that the moisture sensors in the 10 cm layer vary more than the ones in the 20 cm layer. This is the case, because the rainfall water penetrates more easily in the top of the soil, so as the depth increases the water retention decreases.

Figure 8 shows the values for the precipitations per day gathered by the pluviometer of the prototype and figure 9 presents a comparison between the precipitations of the used pluviometers. In order to verify the values of our pluviometer we use the data from the Irriga System's pluviometer installed close to the location of where our prototype was. In addition, we also used the data of the INMET station located in Santa Maria, however not in the same location as the other two pluviometers. Figure 9 shows that the values gathered by our pluviometer were close to the Irriga System's pluviometer. In figure 10, one can notice that the mean data of the soil temperature vary more that the mean data of the environmental temperature.

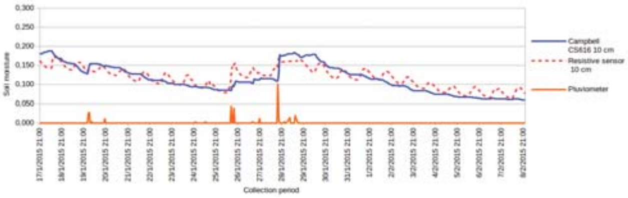


Fig. 6: Moisture sensors in the 10 cm layer.

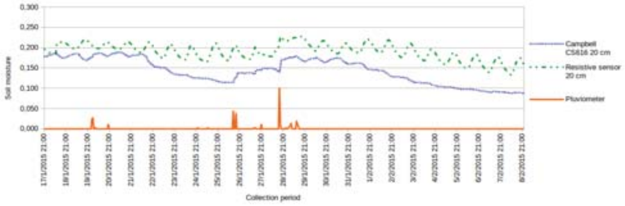


Fig. 7: Moisture sensors in the 20 cm layer.

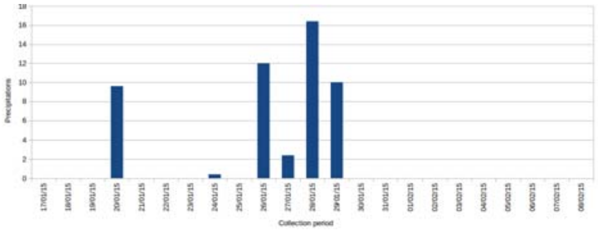


Fig. 8: Precipitation per day.

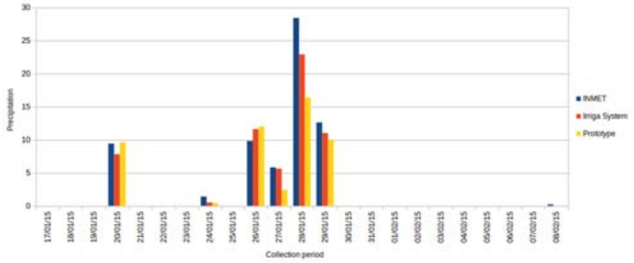


Fig. 9: Comparison between the pluviometers.

By analyzing the collected data, we notice that the correlation between the values for the moisture sensors could be

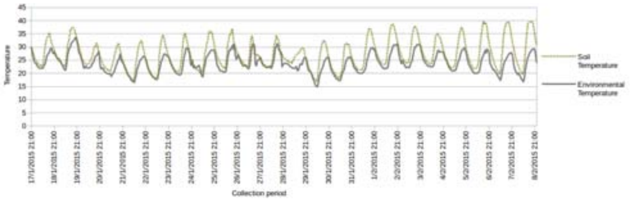


Fig. 10: Comparison between the gathered temperatures.

better. Therefore, we studied two alternatives, both are presented in section 4.2: (i) the adjustment of the soil moisture values using formulas that consider the soil temperature and (ii) use a data mining program with different algorithms, such as linear regression and neural networks, in order to generate equations that can be used to adjust the raw data. Therewith, we aimed to create adjustment equations for the resistive sensors using the CS616 as reference, which is widely used in the field. The statistics analysis were made to determine the correlations and the coefficients of determination of the resistive sensors using all the available information: soil moisture sensors, soil and environmental temperature and precipitation.

4.2 Alternatives for adjustment of soil moisture values

In this section, we present the two alternatives used to perform the adjustments in soil moisture values. The first one takes into account the soil temperature and the second one is bye the usage of a data mining software.

In the first alternative we verified that many factors can influence the electric conductivity of the soil, like soil texture and water content [26]. The temperature is also an important factor that affects the electric conductivity (according to the conductivity increase, the material resistance decreases, as explained in section 2.2), raising it to approximately 1,91% per C in temperature [27]. Therefore, to perform this adjustments in the soil moisture values, the measurements are made in the current soil temperature but are adjusted to a particular temperature. To do so, we used the relation model presented in equation 2, which is indicated in the study about correction models in [21], allowing an adjustment with satisfactory precision between a range of 3-50 C for measurements related to a default temperature of 25 C.

$$U25 = \frac{U_t}{1 + \delta(T - 25)} \tag{2}$$

Where U25 is the soil moisture in the base temperature of 25 C, δ is the temperature compensation informed in the model, U_t is the soil moisture gathered at the current temperature and T is the gathered soil temperature.

The second alternative was found by the use of the data mining software Weka (Waikato Environment for Knowledge Analysis), which has a set of machine learning algorithms for data mining tasks (WEKA, 2015). It was used to perform the measurement of the resistive soil moisture sensors using as reference the Campbell moisture sensors. The tests were performed using the k-fold cross-validation technique available in the software. We also perform the tests using a training phase, but the results were similar.

In the next section is presented the correlation and determination coefficients between the soil moisture sensors, which had the best results.

4.3 Determination coefficients between the soil moisture sensors

According to table 1, both of the values generated by the temperature adjustments (Temp Ad.) and the values obtained by the correction formulas indicated by the data mining software (Original data) had good results. The comparisons were made using three kinds of values: using only the sensor of the corresponding layer *without other information* - WOI, presented in the second and fifth rows of the table; using the sensor's data of the corresponding layer with other information (soil temperature, environmental temperature and precipitation) presented in the third and sixth rows of the table; and using the gathered data from the resistive soil moisture sensors of both layers with other information (soil temperature, environmental temperature and precipitation) presented in fourth and seventh rows of the table. In all tests the Campbell sensor of the corresponding layer (10 or 20 cm) was used as reference.

Table 1: Determination coefficients (R^2) between the soil moisture sensors.

Resistive	Linear R.	Linear R.	Neural N.	Neural N.
Sensor	Original data	Temp Ad.	Original data	Temp Ad.
at 10 cm (WOI)	0.78	0.90	0.72	0.91
at 10cm	0.93	0.93	0.94	0.95
at 10 cm (10 and 20)	0.93	0.93	0.94	0.94
at 20 cm (WOI)	0.51	0.75	0.42	0.74
at 20cm	0.78	0.78	0.78	0.75
at 20 cm (10 and 20)	0.88	0.88	0.94	0.94

Figures 11 and 12 present the determination coefficients related to the *original data - OD* and from the *temperature adjustment - TA* from the soil moisture sensors with the use of linear regression and neural network functions to better illustrate the values presented in table 1.

By using the linear regression and the comparison of the resistive sensor at 10 cm (WOI) with the sensor Campbell at 10 cm as well as the comparison of the resistive sensor at 20 cm (WOI) with the sensor Campbell at 20 cm, the formulas for this adjustment, taking into consideration the temperature, showed more satisfactory results if compared with the formulas for the original data used by the software Weka: the resistive sensor at 10 cm (WOI) had a R^2 of 0.90; the resistive sensor at 20 cm (WOI) had a R^2 of 0.75. Therefore, this option of adjusted values by the

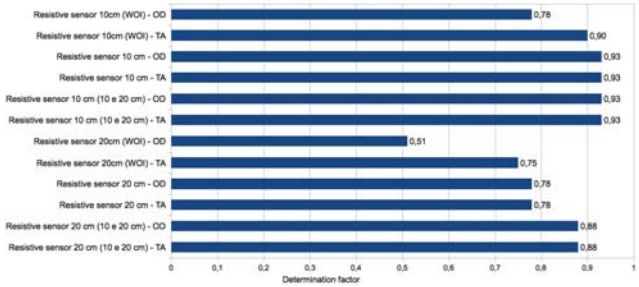


Fig. 11: Determination coefficients of the soil moisture sensors with linear regression.

temperature formula is interesting if only the moisture and soil temperature values are available, and there are no other information, such as environmental temperature or precipitation. However, using all available information (soil moisture, soil and environmental temperature and precipitation) for both of the 10 and 20 centimeters sensors, the best results were found by using the resistive sensors' values of the two layers simultaneously regardless of the technique used: resistive sensor at 10 cm (10 and 20 cm) had a R^2 of 0.93; and resistive sensor at 20 cm (10 and 20 cm) had R^2 of 0.88.

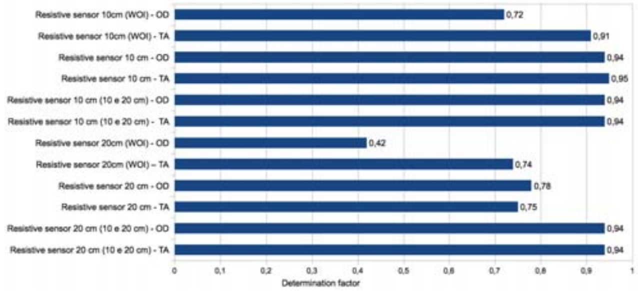


Fig. 12: Determination coefficients of the soil moisture sensors with Neural Network Multilayer Perceptron.

Using the neural network Multilayer Perceptron and performing the comparison between the moisture sensors without any other information, that is, only the comparison of the resistive sensor at 10 cm (WOI) with the Campbell sensor at 10 cm and the comparison of the resistive sensor at 20 cm (WOI) with the Campbell sensor at 20 cm, the formulas for adjustment, taking into account the temperature, also had the best results if compared to the formulas for the original data used in the software Weka: the resistive sensor at 10 cm (WOI) had a R^2 of 0.91; the resistive sensor at 20 cm (WOI) had a R^2 of 0.74. For the values of the neural network using all the available information for the sensor at 10 cm, the best results were: the resistive sensor at 10 cm had a R^2 of 0.95 (with temperature adjustment) and the resistive sensor at 10 cm (10 and 20 cm) had a R^2 of 0.94 (original data). For the sensor at 20 cm, again, the best result was using the data from both layers: the resistive sensor at 20 cm (10 and 20 cm) had a R^2 of 0.94.

Figure 13 presents a comparison between the original data for the resistive sensors adjusted by the formulas from Weka using all the available information and the CS616 sensors placed in the 10 cm layer in the sandbox, and figure 14 presents the comparison for the 20 cm layer. The formulas 3 and 4 were generated by the linear regression from the software Weka to illustrate the adjustments made in the soil moisture values. These formulas can be implemented directly in Data Normalization layer of the architecture (Arduino), as they require no big data processing, deferentially of a neural network.

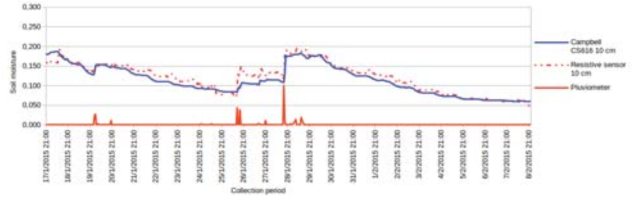


Fig. 13: Adjustment of the soil moisture for the sensor in the 10 cm layer using data from the 10 and 20 cm layers.

Formula for the adjustment of the soil moisture for the sensor in the 10 cm layer using data from the 10 and 20 cm layers (linear regression):

$$SM = (-0.0003 * RS20) + (-0.0012 * RS10) \\ +(-0.0009{*}ST)+(-0.003{*}ET)+(-0.0029{*}P)+1.55 \tag{3}$$

Where SM is the final soil moisture value, RS10 is the resistive sensor at 10cm, RS20 is the resistive sensor at 10cm, ST is the soil temperature, ET is the environmental temperature of INMET station, and P is the precipitation.

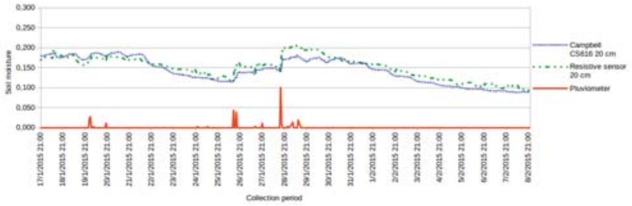


Fig. 14: Adjustment of the soil moisture for the sensor in the 20 cm layer using the data from the 10 and 20 cm layers.

Formula for the adjustment of the soil moisture for the sensor in the 20 cm layer using the data from the 10 and 20 cm layer (linear regression):

$$SM = (-0.0005 * RS2) + (-0.0007 * RS10) \\ +(-0.0013{*}ST)+(-0.0024{*}ET)+(-0.0024{*}P)+1.290 \tag{4}$$

5. Final Considerations and future works

This work proposes the development of a remote sensing architecture using open hardware technologies and all the available information in the environment for the soil moisture verification. In order to evaluate this architecture we built a study case for collecting and sending data through the GSM network, and performing a comparison between the low-cost resistive sensors of soil moisture produced in the Federal University of Santa Maria with the CS616 soil moisture sensors of Campbell.

We conclude that both alternatives used for the adjustments and comparisons, between the resistive and the CS616 sensors, had satisfactory results. When we use only the information about temperature and moisture of the soil, the formulas generated by the temperature adjustments had better results. However, when all available information (moisture and soil temperature, ambient temperature and precipitation) is used, both techniques show good results and higher determination factors due to the usage of more information to generate the formulas. One can notice that the Multilayer Perceptron neural network increases the determination factor in some cases, but in order to used it, the neural network should be implemented on the server, because it is more complex than using just the linear regression formulas. These formulas generated by the linear regression function can be implemented in the data normalization module of the proposed architecture, more specifically, they would be implemented in the Arduino itself.

In future works, we intend to perform more tests in different scenarios, that is, in other types of soil and with varied climates, in order to verify the behavior of the sensors under several different conditions. We also intend to perform tests using only the prototype connected to the battery and the solar panel, checking the power consumption and making improvements to reduce it.

References

[1] G. S. Rodrigues and L. J. M. Irias, "Considerações sobre os impactos ambientais da agricultura irrigada," Embrapa Meio Ambiente, 2004.

[2] A. Daccache, J. Knox, E. Weatherhead, A. Daneshkhah, and T. Hess, "Implementing precision irrigation in a humid climate–recent experiences and on-going challenges," *Agricultural Water Management*, 2014.

[3] R. Carlesso, M. Petry, M. Rosa, and B. Heldwein, *Usos e Beneficios da Coleta AutomÃqtica de Dados Meteorologicos na Agricultura.* Editora UFSM, 2007.

[4] G. H. Hargreaves *et al.*, "Food, water, and a possible world crisis." in *National irrigation symposium. Proceedings of the 4th Decennial Symposium, Phoenix, Arizona, USA, November 14-16, 2000.* American Society of Agricultural Engineers, 2000, pp. 187–194.

[5] T. A. Howell, "Enhancing water use efficiency in irrigated agriculture," *Agronomy journal*, vol. 93, no. 2, pp. 281–289, 2001.

[6] R. Salazar, J. C. Rangel, C. Pinzó, and A. Rodríguez, "Irrigation system through intelligent agents implemented with arduino technology," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journual*, vol. 1, no. 6, pp. 29–36, 2013.

[7] M. Thalheimer, "A low-cost electronic tensiometer system for continuous monitoring of soil water potential," *Journal of Agricultural Engineering*, vol. 44, no. 3, p. e16, 2013.

[8] M. M. Junior, R. O. Nunes, and V. G. Celinski, "Comparison of the responses of low cost eletrical soil sensors, and a arduino microcontroller platform," *Iberoamerican Journal of Applied Computing*, vol. 2, no. 1, 2013.

[9] Arduino. (2015) Arduino plataform. [Online]. Available: http://arduino.cc/

[10] W. Wagner, V. Naeimi, K. Scipal, R. de Jeu, and J. Martínez-Fernández, "Soil moisture from operational meteorological satellites," *Hydrogeology Journal*, vol. 15, no. 1, pp. 121–131, 2007.

[11] R. De Jeu, W. Wagner, T. Holmes, A. Dolman, N. Van De Giesen, and J. Friesen, "Global soil moisture patterns observed by space borne microwave radiometers and scatterometers," *Surveys in Geophysics*, vol. 29, no. 4-5, pp. 399–420, 2008.

[12] J. Haule and K. Michael, "Deployment of wireless sensor networks (wsn) in automated irrigation management and scheduling systems: a review," in *Science, Computing and Telecommunications (PACT), 2014 Pan African Conference on*. IEEE, 2014, pp. 86–91.

[13] D. Robinson, C. Campbell, J. Hopmans, B. Hornbuckle, S. B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth, "Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review," *Vadose Zone Journal*, vol. 7, no. 1, pp. 358–389, 2008.

[14] E. Scudiero, A. Berti, P. Teatini, and F. Morari, "Simultaneous monitoring of soil water content and salinity with a low-cost capacitance-resistance probe," *sensors*, vol. 12, no. 12, pp. 17 588–17 607, 2012.

[15] B. Böhme, M. Becker, and B. Diekkrüger, "Calibrating a fdr sensor for soil moisture monitoring in a wetland in central kenya," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 66, pp. 101–111, 2013.

[16] G. Calamita, L. Brocca, A. Perrone, S. Piscitelli, V. Lapenna, F. Melone, and T. Moramarco, "Electrical resistivity and tdr methods for soil moisture estimation in central italy test-sites," *Journal of Hydrology*, vol. 454, pp. 101–112, 2012.

[17] S. Liaghat and S. K. Balasundram, "A review: The role of remote sensing in precision agriculture." *American Journal of Agricultural & Biological Science*, vol. 5, no. 1, 2010.

[18] G. A. Mills, S. K. Armoo, A. K. Rockson, R. A. Sowah, and M. A. Acquah, "Gsm based irrigation control and monitoring system." *International Journal of Engineering Science & Technology*, vol. 5, no. 7, 2013.

[19] Z. Ahmed, "Design of autonomous low power sensor for soil moisture measurement." Master's thesis, LinkÃűping University, Electronics System, The Institute of Technology, 2013, 2013.

[20] Pessl. (2015) Metos compact. users' manual pessl instruments. [Online]. Available: http://efesaro.com/pdf/Manual_Metos_Compact_ingles.pdf

[21] R. Ma, A. McBratney, B. Whelan, B. Minasny, and M. Short, "Comparing temperature correction models for soil electrical conductivity measurement," *Precision Agriculture*, vol. 12, no. 1, pp. 55–66, 2011.

[22] J. S. Qu, J. Fan, and D. C. Huang, "The capacitive soil moisture sensor research," in *Applied Mechanics and Materials*, vol. 584. Trans Tech Publ, 2014, pp. 2142–2149.

[23] CS616. (2015) Instruction manual - cs616 and cs625 water content reflectometers. [Online]. Available: http://s.campbellsci.com/documents/us/manuals/cs616.pdf

[24] Inmet. (2015) Inmet - national institute of meteorology. [Online]. Available: http://www.inmet.gov.br/portal/

[25] SistemaIrriga. Sistema irriga®. [Online]. Available: https://www.sistemairriga.com.br/

[26] E. C. Brevik, T. E. Fenton, and R. Horton, "Effect of daily soil temperature fluctuations on soil electrical conductivity as measured with the geonics® em-38," *Precision Agriculture*, vol. 5, no. 2, pp. 145–152, 2004.

[27] D. Corwin and S. Lesch, "Apparent soil electrical conductivity measurements in agriculture," *Computers and electronics in agriculture*, vol. 46, no. 1, pp. 11–43, 2005.

[28] Weka. (2015) Data minig software. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/

A Novel Cosine-Phased Binary Offset Carrier Signal Tracking

S. Woo[1], K. Chae[1], H. Liu[2], and S. Yoon[1,†]
[1]College of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea
[2]School of Electrical Engineering and Computer Science, Oregon State University, OR, USA
[†]Corresponding author (Email: syoon@skku.edu)

Abstract— *This paper addresses a novel cosine-phased binary offset carrier (BOC) signal tracking, enabling an unambiguous signal tracking. Two novel locally-generated signals are first obtained by dividing the cosine-phased BOC sub-carrier into multiple parts, and then, cross-correlations between each of the locally-generated signals and the received signal are generated. Finally, the cross-correlations are efficiently combined, yielding a correlation function with no side-peaks, thus consequently, removing ambiguity in signal tracking. The tracking error performance comparison between the proposed and conventional correlation functions shows us that the proposed correlation function can offer a significant improvement in performance compared with the conventional correlation functions.*

Keywords: global navigation satellite systems (GNSSs), global positioning system (GPS), binary offset carrier (BOC), tracking

1. Introduction

Due to the sharp main-peak of the binary offset carrier (BOC) signal correlation function providing an improved location accuracy, next generation global navigation satellite systems (GNSSs) such as the modernized global positioning system (GPS) and Galileo have adopted the BOC as a modulation scheme, instead of the conventional phase shift keying (PSK) signal [1]. Despite the advantage in main-peak of the BOC, the BOC-modulated signal has multiple side-peaks in its autocorrelation, and the side-peaks could bring on an ambiguity in signal tracking, eventually could lead to a serious location error [2], [3].

Although several correlation functions [3]-[6] have been proposed to remove the side-peaks so far, most of the correlation functions can be used for sine-phased BOC signals only, i.e., they are inapplicable to cosine-phased BOC signals used in many GNSS bands including Galileo E1 and E6 bands. [4] is applicable to cosine-phased BOC signals to some degree; however, it cannot remove the side-peaks completely, thus leaving the ambiguity problem unsolved. So, in this paper, we propose a novel unambiguous correlation function for cosine-phased BOC signals. Removing the cosine-wave pattern in the BOC sub-carrier causing the side-peaks, first, we design two novel locally-generated signals and generate the corresponding cross-correlations. Combining the cross-correlations in a specialized way based on the absolute-value arithmetic, then, we create a novel correlation function with no side-peaks. The proposed correlation function has two significant advantages over the conventional correlation functions: First, the side-peaks causing an ambiguity in signal tracking are removed completely. Second, the sharpness of the proposed correlation function can be adjusted according to system design requirements, allowing us more flexibility in designing a system.

The rest of this paper is organized as follows: In Section 2, we describe the signal model of the cosine-phased BOC signal. In Section 3, we propose a novel correlation function for cosine-phased BOC signal tracking. In Section 4, we compare the tracking performances of the proposed and conventional correlations. Finally, we conclude this paper in Section 5.

2. Cosine-phased BOC signal model

The cosine-phased BOC signal is denoted by $\text{BOC}_{\text{cos}}(kn, n)$, where k is the ratio of the PRN code chip duration and the sub-carrier period, and n is the ratio of the PRN code chip rate and 1.023 MHz.

The $\text{BOC}_{\text{cos}}(kn, n)$ signal $C(t)$ can be expressed as

$$C(t) = \sqrt{P} \sum_{i=-\infty}^{\infty} h_i p_{T_c}(t - iT_c) s_{cs}^i(t), \tag{1}$$

where P is the signal power, $h_i \in \{-1, 1\}$ is the ith chip of a PRN code with a period T, $p_\alpha(t)$ denotes the unit rectangular pulse over $[0, \alpha)$, T_c denotes the PRN code chip duration, and $s_{cs}^i(t)$ denotes the cosine-phased sub-carrier.

Since the cosine-phased sub-carrier over one PRN code chip duration consists of $4k$ sub-carrier pulses, the sub-carrier $s_{cs}^i(t)$ can be expressed as

$$\begin{aligned} s_{cs}^i(t) &= \sum_{l=0}^{4k-1} m_l^i p_{T_s}(t - iT_c - lT_s), \\ &= \sum_{l=0}^{4k-1} s_l^i(t), \end{aligned} \tag{2}$$

where $m_l^i = (-1)^{2ki+\lceil l/2 \rceil}$ is the sign of the lth sub-carrier pulse in the ith PRN code chip, $T_s = T_c/(4k)$ is the sub-carrier pulse duration, $s_l^i(t)$ is the lth sub-carrier pulse in the ith PRN code chip, and $\lceil x \rceil$ denotes the smallest integer not less than x [7].

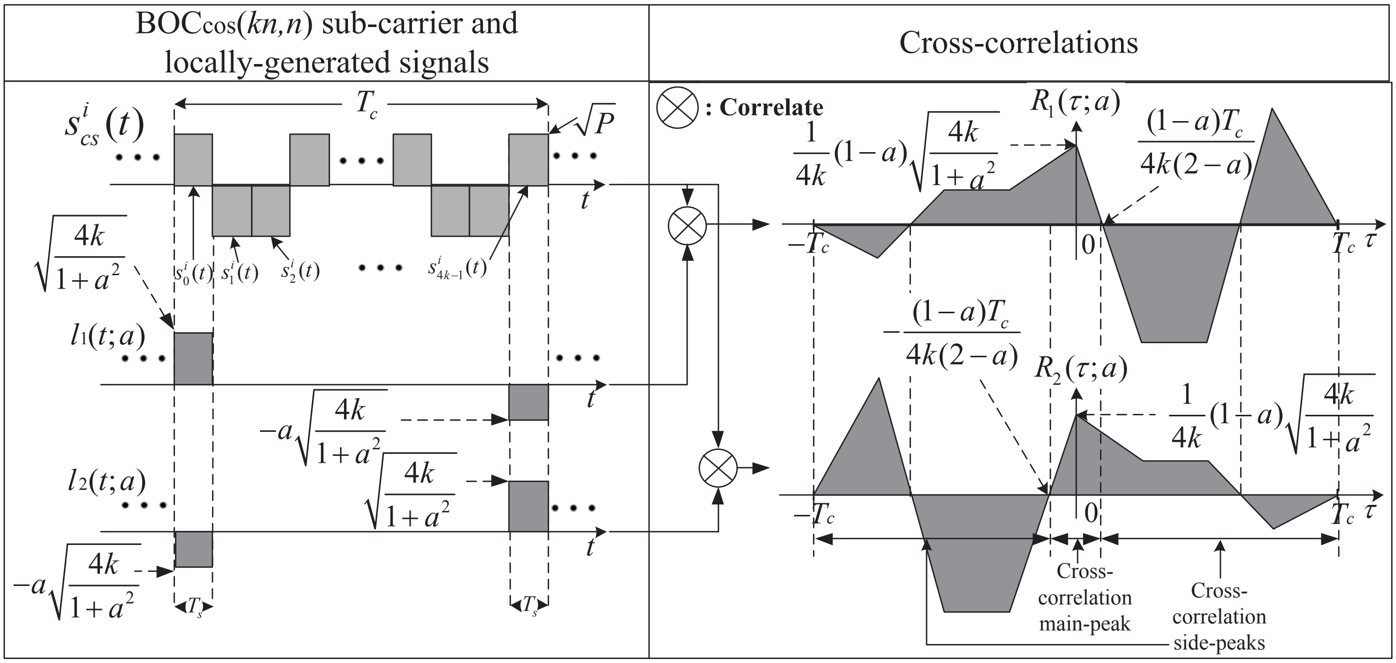

BOCcos(kn,n) sub-carrier and locally-generated signals
Cross-correlations
$s_{cs}^{i}(t)$
T_c
$\sqrt{P}$
t
$\sqrt{\frac{4k}{1+a^2}}$
$s_0^i(t)$
$s_1^i(t)$
$s_2^i(t)$
$s_{4k-1}^i(t)$
$l_1(t;a)$
$-a\sqrt{\frac{4k}{1+a^2}}$
$l_2(t;a)$
T_s
: Correlate
$R_1(\tau;a)$
$\frac{1}{4k}(1-a)\sqrt{\frac{4k}{1+a^2}}$
$\frac{(1-a)T_c}{4k(2-a)}$
$-T_c$
0
τ
$-\frac{(1-a)T_c}{4k(2-a)}$
$R_2(\tau;a)$
Cross-correlation main-peak
Cross-correlation side-peaks

Fig. 1: BOC$_{\cos}(kn,n)$ sub-carrier, locally-generated signals, and cross-correlations.

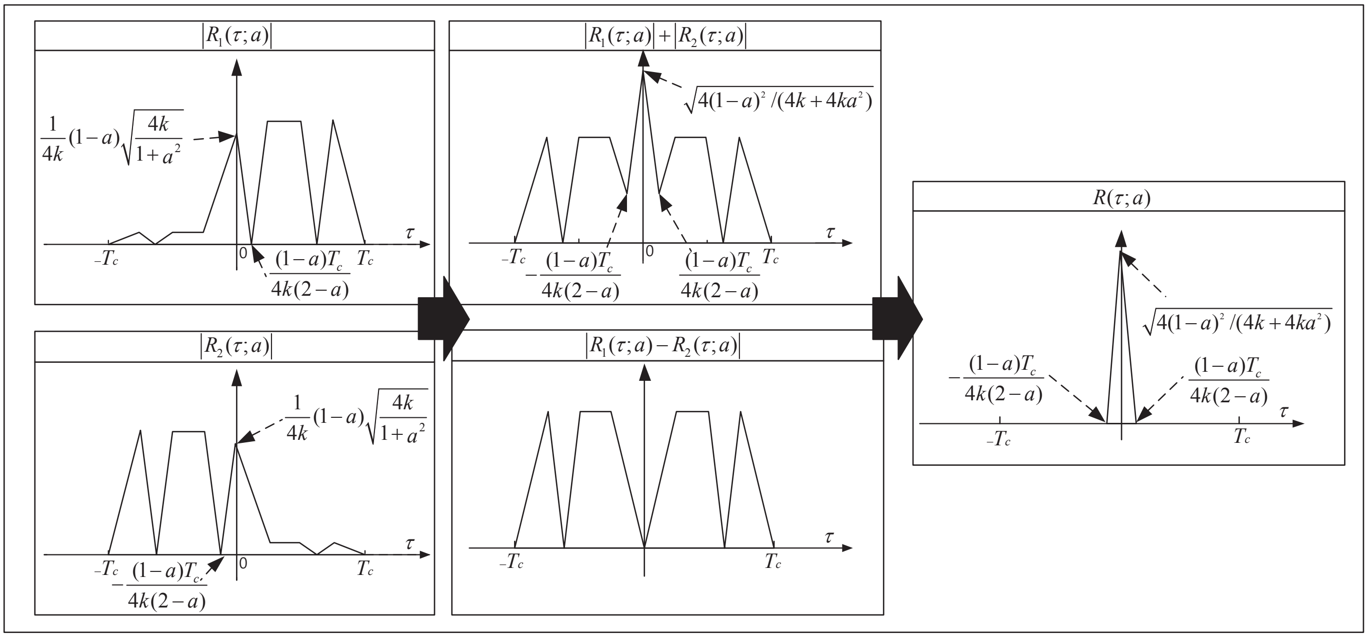

$|R_1(\tau;a)|$
$\frac{1}{4k}(1-a)\sqrt{\frac{4k}{1+a^2}}$
$-T_c$
T_c
τ
$\frac{(1-a)T_c}{4k(2-a)}$
$|R_2(\tau;a)|$
$-\frac{(1-a)T_c}{4k(2-a)}$
$|R_1(\tau;a)|+|R_2(\tau;a)|$
$\sqrt{4(1-a)^2/(4k+4ka^2)}$
$|R_1(\tau;a)-R_2(\tau;a)|$
$R(\tau;a)$

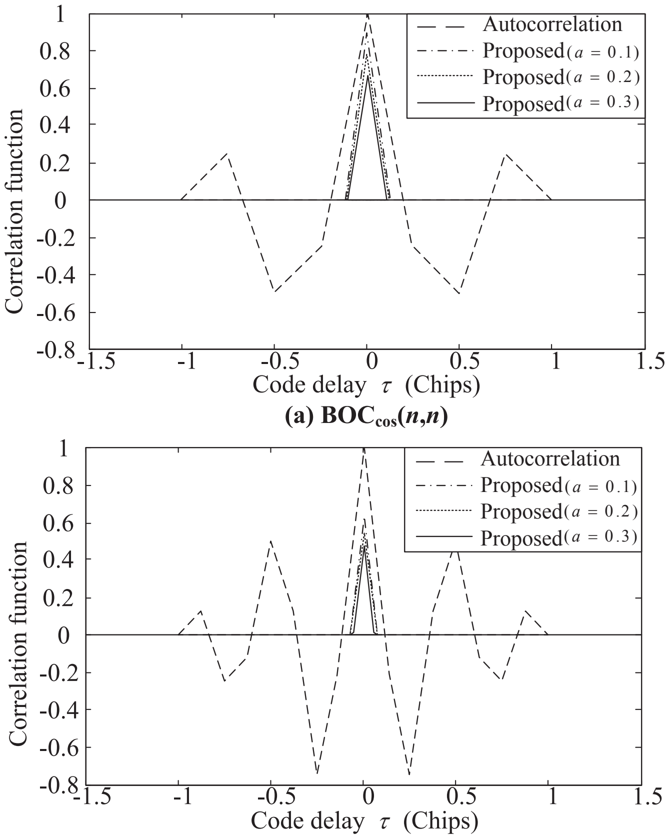
Fig. 2: Process of generating the proposed correlation function by using two cross-correlations.

3. Proposed correlation function

First, we design locally-generated signals to be used instead of the BOC sub-carriers, and obtain cross-correlations by correlating each of the locally-generated signals and received signal, respectively. To design the locally-generated signals, we use the following absolute-value arithmetic property

$$|A| + |B| - |A - B| = \begin{cases} > 0, & \text{for} AB > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

From (3), we can see that the side-peaks would be removed under the following three conditions: (i) The multiplication of the cross-correlation main-peaks is positive, (ii) the multiplication of the cross-correlation side-peaks is negative,



(a) BOC$_{\cos}$(*n*,*n*)

(b) BOC$_{\cos}$(2*n*,*n*)

Fig. 3: The proposed correlation and autocorrelation functions for BOC$_{\cos}(n, n)$ and BOC$_{\cos}(2n, n)$.

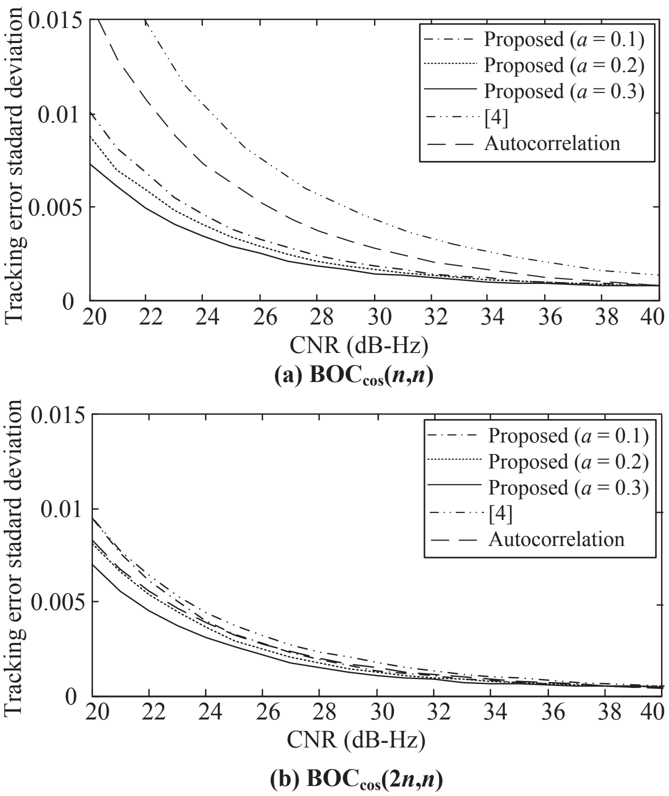
and (iii) the two cross-correlations must be symmetric to each other to generate a symmetric unambiguous correlation function, due to the side-peaks being caused by the cosine-wave pattern of the sub-carrier. Based on these observations, in this paper, we design two locally-generated signals as

$$\begin{cases} l_1(t;a) = \sum_{-\infty}^{\infty} \sqrt{\frac{4k}{1+a^2}}(s_0^i(t) - as_{4k-1}^i(t)), \\ l_2(t;a) = \sum_{-\infty}^{\infty} \sqrt{\frac{4k}{1+a^2}}(-as_0^i(t) + s_{4k-1}^i(t)), \end{cases} \quad (4)$$

where $l_1(t;a)$ and $l_2(t;a)$ are the locally-generated signals, $0 \leq a < 1$ is a parameter for adjusting the sharpness of the proposed correlation function. Then, we correlate the locally-generated signals and received signal, yielding cross-correlations

$$R_j(\tau;a) = \frac{1}{PT}\int_0^T C(t)l_j(t+\tau;a)dt, \;\; j = 1, 2. \quad (5)$$

In Fig. 1, we can see that the locally-generated signals are symmetric to each other and have no cosine-wave pattern, and also that the multiplication of the cross-correlations



(a) BOC$_{\cos}$(*n*,*n*)

(b) BOC$_{\cos}$(2*n*,*n*)

Fig. 4: TESD performances of the proposed and conventional correlation functions for BOC$_{\cos}(n, n)$ and BOC$_{\cos}(2n, n)$.

satisfy

$$\begin{cases} R_1(\tau;a)R_2(\tau;a) > 0, \\ \qquad \text{for} -\frac{(1-a)T_c}{4k(2-a)} < \tau < \frac{(1-a)T_c}{4k(2-a)}, \\ R_1(\tau;a)R_2(\tau;a) = 0, \text{ otherwise.} \end{cases} \quad (6)$$

Thus, from (3) and (6), we can generate an unambiguous correlation function

$$R(\tau;a) = |R_1(\tau;a)| + |R_2(\tau;a)| - |R_1(\tau;a) - R_2(\tau;a)|, \quad (7)$$

and the process of (7) is depicted in Fig. 2, where we can see that the side-peaks are removed, and the main-peak height and width of $R(\tau;a)$ are adjusted by the parameter a.

Fig. 3 depicts that the proposed correlation and autocorrelation functions for BOC$_{\cos}(n, n)$ and BOC$_{\cos}(2n, n)$. From the figure, we can see that the side-peaks on the autocorrelation are removed completely through the process (7). In addition, we can see that the height and width of the proposed correlation are adjustable by fixing the parameter a.

4. Numerical results

In this section, we compare tracking error standard deviation (TESD) performances of the proposed and conventional correlation functions. The TESD is defined as

$$\frac{\sigma}{G}\sqrt{2B_L T_I}, \quad (8)$$

where σ is the standard deviation of the discriminator output at $\tau = 0$, B_L is the loop filter bandwidth, T_I is the integration time, and G is the discriminator gain at $\tau = 0$ [8]. For simulations, we consider the following parameters: B_L = 1 Hz, $T = T_I$ = 4 ms, T_c^{-1} = 1.023 MHz, and the early-late spacing $\Delta = \frac{T_c}{16}$.

Fig. 4 shows the TESD performances of the proposed and conventional correlation functions for BOC$_{\cos}(n, n)$ and BOC$_{\cos}(2n, n)$ as a function of the carrier to noise ratio (CNR) defined as P/W_0 with W_0 the noise power spectral density. From the figure, it is observed that the proposed correlation function provides a better TESD performance than the conventional correlation functions including the auto-correlation function in the CNR range of $20 \sim 40$ dB-Hz of practical interest. In addition, it is seen that the proposed correlation function performs better as the value of a increases. This is because the correlation main-peak becomes sharper, as the value of a increases. However, it should be noted that the tracking range would be smaller for a larger value of a, and thus, the value of a should be determined according to system design requirements.

5. Conclusion

A novel correlation function has been proposed for unambiguous cosine-phased BOC signal tracking. We have first designed locally-generated signals to be used instead of the cosine-phased BOC sub-carrier and then have obtained the corresponding cross-correlation functions. Subsequently, by combining the cross-correlation functions in a specialized way, we have created an unambiguous correlation function. The numerical results has confirmed that the proposed correlation function has a much better tracking performance than the conventional correlation functions.

Acknowledgment

This research was supported by the National Research Foundation (NRF) of Korea under Grant 2012R1A2A2A01045887 with funding from the Ministry of Science, ICT&Future Planning (MSIP), Korea.

References

[1] M.-Flissi, K.-Rouabah, D.-Chikouche, A.-Mayouf, and S.-Atia, "Performance of new BOC-AW-modulated signals for GNSS system," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp.1-18, Jan. 2013.

[2] E. D. Kaplan and C. J. Hegarty, *Understanding GPS: Principles and Applications.* Artech House, 2005.

[3] O. Julien, C. Macabiau, M. E. Cannon, and G. Lachapelle, “ASPeCT: unambiguous sine-BOC(n, n) acquisition/tracking technique for navigation applications," *IEEE Trans. Aer., Electron. Syst.*, vol. 43, no. 1, pp. 150-162, Jan. 2007.

[4] A. Burian, E. S. Lohan, and M. K. Renfors, “Efficient delay tracking methods with sidelobes cancellation for BOC-modulated signals," *EURASIP J. Wireless Commun. Network*, vol. 2007, article ID. 72626, 2007.

[5] Z. Yao, X. Cui, M. Lu, Z. Feng, and J. Yang, “Pseudo-correlation-function-based unambiguous tracking technique for sine-BOC signals," *IEEE Trans. Aer., Electron. Syst.,* vol. 46, no. 4, pp. 1782-1796, Oct. 2010.

[6] K. Chae, H. Liu, and S. Yoon, “Unambiguous BOC signal tracking based on partial correlations," in *Proc. Vehic. Technol. Conf. (VTC)*, CD-ROM, Vancouver, Canada, Sep. 2014.

[7] J. W. Betz. “Binary offset carrier modulations for radionavigation," *J. Inst. Navig.*, vol 48, no. 4, pp. 227-246, Dec. 2001.

[8] A. J. Van Dierendonck, P. Fenton, and T. Ford, “Theory and performance of narrow correlator spacing in a GPS receiver," *J. Inst. Navig.,* vol. 39, no. 3, pp. 265-283, June 1992.

Transducer Module Recognition in a Network Environment Based on IEEE 1451

Tércio A. dos Santos Filho[1], **Antonio C. Oliveira-Jr**[1], **Dalton M. Tavares**[1]
Marcos A. Batista[1] **Alexandre C. R. da Silva**[2] **and Marcos N. Rabelo**[3]
[1]Federal University of Goiás - UFG/CAC
Institute of Biotechnology
Catalão - Goiás - Brazil
e-mail: tercioas@gmail.com, antonio@catalao.ufg.br,
dmatsuo@gmail.com, marcos.batista@pq.cnpq.br
[2]São Paulo State University - UNESP
Department of Electrical Engineering - DEE
Ilha Solteira - São Paulo - Brazil
e-mail: acrsilva@dee.feis.unesp.br
[3]Federal University of Goiás
Graduate Program in Modeling and Optimization
Catalão, GO, Brazil
e-mail: rabelo@dmat.ufpe.br

Abstract— *The development of smart transducers networks demand a detailed study of processes to be controlled in order to achieve the user needs. This research proposes the development of networks with their own architectures in order to avoid compatibility issues at software or hardware level, which are common in proprietary communication protocols. The IEEE 1451 standard defines a set common commands and describes the features to NCAP and TIM to connect the transducers in network of way plug-and-play. This work presents a NCAP embedded using hardware and system operating dynamic implemented in DE2 kit. The NCAP has two interfaces based on IEEE 1451.5 (ZigBee) and IEEE 1451.2 (RS-232). Another important feature in this paper is the TEDS that were described of two ways, in STIM stored in no-volatile memory and to WTIM was used a text file stored in NCAP denominated virtual TEDS. The results were obtained through web page using a microcomputer to access the data in NCAP, like: temperature sensor, logs and the TEDS.*

Keywords: IEEE 1451, NCAP, RS-232, TIM, ZigBee.

1. Introduction

The transducers are components that convert a type of energy in other one, for example: electrical energy in mechanical (actuator) or physical in electrical (sensor). To realise the system data acquisition and control, the transducers are connected together with others devices, like: microcontroller and FPGA, denominated them smart transducers. The smart transducers network are developed using smart transducers module connected through a interface with others nodes or routers for realising the monitoring and/or control of the system in which being applied, these networks are denomineted DMC (Distributed Measurement and Control). The DMC systems development requires a great deal of engineering for its design, with the need to use tools and proprietary software, many with high costs for implementation. The transducer network can be applied in several environments in which wish monitoring or control, being used in smart houses, in industry, health and many others places, however, each manufacturer has a set of commands and protocols owner [1] [2].

To solve the problems of standardisation to NIST partnership with IEEE defined a set of standards and protocols for connecting smart transducers denomineted IEEE 1451 standard. The IEEE 1451 defines a set of protocols for wired and wireless distributed measurement, controls applications and plug-and-play through the TEDS (Transducer Electronic Data Sheet), in which is divided in two module: NCAP and TIM. The NCAP is a network node compost of hardware and software that provides the gateway functions between TIM and the user network, and the TIM is a module that contains signal conditioning, analog-to-digital or/and digital-to-analog conversion, frequency and digital converter and a interface to communicate with NCAP [3][4].

One technique to achieve plug-and-play was the definition of a minimum set of transducers and other optional features for more advanced functions. The TIM when is connected in NCAP it transfers TEDS information to protocol manager, in which performs the acknowledgment of transducers in the network automatically. each table in TEDS there is a format standardized based on IEEE 1451.0, being: Length, Data Block and Checksum, where, each line is denominated of TLV (Type/Length/Value). In Section 4.1 is described in more details the TEDS format.

The plug and play feature of smart transducers to its users and developers raises some advantages such as reduced time for parameterization of the system, advanced diagnostics, reduced time for repair and replenishment, advanced management of the hardware and automation of calibration [5].

For the implementation of TEDS, four blocks are required, Meta-TEDS TEDS TransducerChannel, User's Transducer Name TEDS TEDS and PHY. The other tables, Calibration TEDS, Frequency Response TEDS, Transfer Function TEDS, Text-based TEDS TEDS Commands, Identification TEDS, Geographic location TEDS TEDS extension Units, End User Application Specific TEDS TEDS defined and Manufacturer are optional. In this work, we used the TEDS mandatory for system testing and implementation of IEEE 1451.0-2007.

The NCAP described in this paper was developed over dynamic hardware and operating system embedded uClinux using C language based on IEEE 1451. It is important detach that the IEEE 1451.1 standard suggests object-oriented implementation, however, there is no requirement that actual implementation use of object-oriented implementation techniques [5]. The NCAP was made to be completely embedded and dynamic, both low level (hardware) like high level (software). To development in low level was used the SoPC Builder that there is default blocks described in VHDL or Verilog, and, in high level was configured and installed the operating system embedded uClinux. In Figure 1 presents the schematic of system.

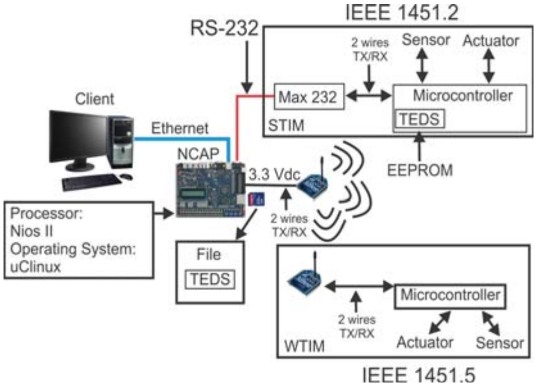


Fig. 1: Schematic of the implemented system.

The main propose of this paper is shows the TEDS of two ways, in TIM and in NCAP, and some features of sensor and actuator. The TEDS in TIM was stored using the memory flash of the Atmega8 microcontroller and in NCAP was stored at file. The TEDS stored in microcontroler describes the STIM developed with three transducers (temperature sensor and two step motors) and a RS232 interface to communication with NCAP. The TEDS stored at file in NCAP describes two transducers (temperature sensor and DC motor) and a wireless interface.

2. Related Works

The IEEE 1451 defines the TEDS description and communication between NCAP and TIM , involving a particular standard of the family and different technology approaches. Several implementations and applications of the IEEE 1451 smart transducer interface standards have been carried out using microprocessors, microcontroller, embedded commercial solutions and multicore technologies. Thus, many studies have been realized as: in [6] the TEDS format was described, and it was implemented a stand alone NCAP, using the RS-232 interface for realizing the communication with the TIM. The author also showed the system using a wireless communication based on IEEE 802.15.4. In [7] was described a NCAP developed in XML (eXtensible Markup Language) and RPC (Remote Procedure Call) with wireless interface IEEE 802.15.4 standard using a ZigBee module. The connection between NCAP and ZigBee module was made using the USB (Universal Serial Bus) standard.

In [8] described the IEEE 1451 and its main features, such as information model in which consists of a hierarchy of classes divided into three main categories: Block, Component and Service. The authors described the main blocks within each class (Block, Component and Service) and system tests, was made a software control water level in a tank. The authors also reported that, despite the advantages of the IEEE 1451.1 standard, there are still no commercial NCAPs available in the market.

In [9] was presented a prototype to the networks of mixed signal (Analogical/Digital) based on IEEE 1451.4 standard, in which was used a accelerometer sensor and was described the TEDS. The authors described the matter of a standardization to analogical transducers and the TEDS format. The authors showed a prototype developed in laboratory in which had a better understand about the standard and the integration between the IEEE 1451.4 and the others standards of IEEE 1451 family.

In [10], the author presented the development of an NCAP interfaces with two USB (Universal Serial Bus) implemented on a microcomputer. As the first step, the author provides a brief description of the IEEE 1451, each committee and when the standards were adopted. The second part presented the development of the system in which each TIM was implemented using the Freescale microcontroller (HC9S12DT256) and a memory (AT24C64A). The protocol used for communication between the NCAP and TIM was PTP (Precision Time Protocol) described by the IEEE 1588 standard in which the author describes the characteristics and size of messages.

In the work of [11] showed communication between the NCAP and WTIM, using the IEEE 1451.5, using the wireless interface based on IEEE 802.11 standard. The authors described the messages, commands, control patterns of the TIM, common commands, such as: TEDS reading, writing the TEDS, the TEDS request commands etc. And response commands to the NCAP, all represented by the class diagram. As a case study, the authors presented the recognition modules WTIMs and showed each step in recognition of the PHY TEDS in a graphical interface done in Java, validating recognition modules.

In other work [12] the author proposed the use of TEDS to store information concerning patient clinical history and diagnostic criteria in order to project learned and patient-adapting devices. The system uses the built-in information to optimize the data processing by adapting the diagnostic algorithm to the specific patient. The author presents the potential of TEDS in use of others applications.

3. Network Node Embedded

The NCAP is a network node with the function of getting data of external network, processing it and sending it to interface based on IEEE 1451.X. The NCAP is divided to hardware and software. The hardware is composed of processor, memory, I/O pins, driver to network and communication interface. The software is composted of an operating system, network protocols and transducers firmware. The operating system provides the logic interface between applications and hardware, and tools to realize the configurations necessary for the system. In Section 3.1 describes the hardware to development of network node and the Section 3.2 describes the operating system embedded to development of NCAP.

3.1 Hardware

The FPGA consists of logic cells arrangement, or configurable logic blocks, contained in a single integrated circuit. Each cell contains computational capacity to implement logic functions and perform routing for communication between blocks. The FPGA basically have logic blocks, blocks inbound and outbound keys and interconnection.

Among the families of commercially available FPGA by Altera, we have: Stratix II, Stratix, StratixGX, Cyclone II, APEX, APEX II, APEX 20K, Mercury, FLEX 10K, ACEX 1K, FLEX 6000 Devices and Excalibur [13].

To development of projects, the Altera developed a processor denominated Nios II that can to be described in VHDL or Verilog using the SoPC Builder. The SoPC Builder is a tool of the Quartus II environment from Altera, that there are blocks defined to implement in FPGA using the VHDL or Verilog language. The blocks provide the communication between the processor and the components, like: memory, I/O pins and interfaces, for example: UART and Ethernet [13].

The NCAP was made with two interfaces and with the Nios II/fast processor using 50MHz clock frequency internal and some components necessary to implement the embedded operating system, like: a timer block, Static Random Access Memory (SRAM) block, Syncronous Dynamic Random Acess Memory (SDRAM) block (for uClinux the minimum requirement is 8MB), Joint Test Action Group Universal Asynchronous Receiver/Transmitter - JTAG UART block, button block, SPI block (SD Card communication) and two Universal Asynchronous Receiver/Transmitter - UART blocks, used to communicate with WTIMs and STIMs.

3.2 Operating System Embedded - uClinux

The uClinux is a operating system embedded focused to work with devices without Memory Management Unit - MMU and offers support many processors, like: Coldfire, Axix, Etrax, ARM (Advanced RISC Machine), Atari 68k, Nios II and others distributed in the market. Today, uClinux is an operating system includes Linux kernel releases 2.0, 2.4 and 2.6 and user applications, libraries and toolchains [14]. The reconfiguration and recompilation of uClinux to each device is made using the software called uClinux-dist.

The uClinux-dist has configurations standards to build image with a set minimum resources predefined which can be modified or include others components [14]. In this paper the following applications were defined to create the image of the operating embedded system for NCAP: chmod, chown, date, df, echo, install, ls, mkdir, mv, pwd, clear, reset, vi, find, lsmod, modprobe, fdisk, arp, ftpget, ftpput, httpd, ifconfig, IP, ping, route, wget, free, kill, ps, uptime, msh and many others components available in the uClinux-dist.

The NCAP was implemented over operating system embedded uClinux. To communication were defined two interfaces in which the both owns the same features defined in the hardware and the operating system. The operating system was defined using the uClinux-dist software, that there are many options of tools to setup the system, as like: “Device drivers", “Character devices" and “Serial drivers". Select the options: “Altera JTAG UART support", “Altera JTAG UART console support" (if use a USB Blaster cable on a nios2-terminal), “Altera UART support", define “Maximum number of Altera UART ports", “baud rate" (in this project was defined 4800 bps) default baud rate for Altera UART ports and “bypass output" when no connection. Defined the interfaces and the applications was possible connect two TIMs (WTIM and STIM) and develop the NCAP using C language over operating system uClinux. For uClinux in the DE2 kit each serial port is defined by a special file in the /dev/ttySX, where, the dev represents the device directory and the ttySX (X port number) represents the access port. The ports defined in this project were ttyS0 (STIM) and ttyS1 (WTIM).

4. IEEE 1451.0 - TEDS

The IEEE 1451.0 - 2007 is a project that presents a common commands feature and TEDS family standard of smart transducers. This feature is independent of the physical medium of communication. This includes the basic functions required to control and manage smart transducers, protocols command and independence of the media format. The IEEE 1451.0 specifies the format of TEDS that are “tables” of data containing information of transducers (sensors/actuators) stored in nonvolatile memory inside the TIM. However, there are applications where storage in non-volatile memory is not practical for application, for example: when there is not memory in TIM, then the IEEE 1451.0 allows storage at other locations remotely calling them Virtual TEDS. The Virtual TEDS are electronic files that provide the same functionality implemented in memory of the TIM, however, they are not in TIM [5].

One technique to achieve feature plug-and-play was to define a minimum set of information from the transducers and other optional features for more advanced functions. The TIMs, to be connected to the NCAP, transfers data from the TEDS to manager protocol, which makes the recognition transducers network automatically, the system working of way plug-and-play.

The plug and play feature of smart transducers to its users and developers raises some advantages such as reduced time for parameterization of the system, advanced diagnostics, reduced time for repair and replacement, advanced management and automation hardware calibration [5].

For the implementation of TEDS, four tables are required, Meta-TEDS, TransducerChannel TEDS, User’s Transducer Name TEDS TEDS and PHY TEDS. The others tables, Calibration TEDS TEDS Frequency Response, Transfer Function TEDS, Text-based TEDS TEDS Commands, Identification TEDS, Geographic location TEDS TEDS extension Units, End User, Application Specific TEDS and Manufacturer TEDS are optionals. In this work was used the TEDS mandatory for system testing and implementation based on IEEE 1451.0-2007.

To test of feature plug-and-play, in this project was developed a module defined by IEEE 1451.2 denominated STIM using interface RS-232 and a module defined by IEEE 1451.5 denominated WTIM using interface ZigBee (point to point). The STIM there are three transducers, being: two step motors and one temperature sensor. The WTIM there are two transducers, being: temperature sensor and DC motor. To make the logic in both modules were used Atmega microcontroller.

4.1 TEDS Format

The TEDS format is common to any TEDS, where, the first field is the length and it is represented by 4 octets unsigned integer. The next block represents the TEDSt’s content, it can be represented by data binary or based on text. The last field in any TEDS is the checksum. The checksum is used to check the integrity of TEDSt’s data [5]. The Table 1 presents the structure TEDS general.

The TEDS fields are described like:

Table 1: Format generic to TEDS.

Field	Description	Type	Octet
—	TEDS length	UInt	4
1 to N	Data block	Variable	Variable
—	Checksum	UInt16	2

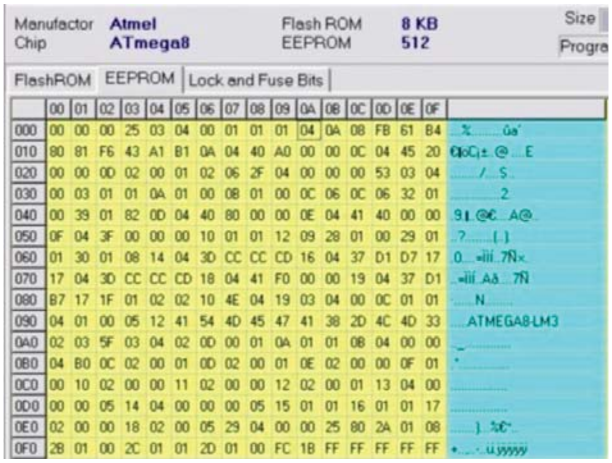
- **TEDS lenght -** it is the number of octets in the data block more two octets in the checksum;
- **Data block-** field that contains specific information according to each table TEDS. The fields that make up this structure are different for each type of TEDS and its structure is based on TLV, except the IEEE 1451.2-1997 and IEEE 1451.3-2003 [5]. The construction of each row within the table structure TEDS uses TLV defined as:
  - ***Type*** - field defined by 1 octet where this field represents the identification of TLV line;
  - ***Length*** - specific the number of octets in the field Value;
  - ***Value*** - content the informations of the TEDS field;
- **Checksum -** it is the complement of the sum of all octets preceded field including the initial size of the TEDS.

5. Transducer Interface Module - TIM

The TIM was developed using the microcontroller ATmega8 from Atmel, integrated circuit MAX 232 (wired) and a X-Bee Pro module (wireless) for interface between NCAP and TIM as demonstrated in Figure 1. To tests were used sensors and actuators, and the TEDS were described in EEPROM memory in the microcontroller like described in Figure 2. When the user connect the TIM in the NCAP, the TEDS are transfers to NCAP and made the recognition of the transducers connected to the TIM. Another way is when the TIM receives a command request, the TIM process the command and sends the reply message to the NCAP based on IEEE 1451.0 standard. The command can be data request from the sensor, actuator’s control or kind of TEDS.

6. TEDS Implementation in the Embedded Node

The NCAP software was developed in C language and using HTML CGI communication protocol. The IEEE 1451.1 standard suggest the implementation object orientation, however, does not demonstrate applications of structured programming. In this context, the work was developed using a new representation structure of NCAP, it being for each class suggested by the standard was represented by a file and each received a file name based on the IEEE 1451.1 standard, such as: IEEE 1451_<name file>.

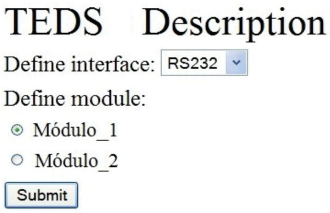


```
Manufactor  Atmel          Flash ROM  8 KB     Size
Chip        ATmega8        EEPROM     512      Progra

FlashROM | EEPROM | Lock and Fuse Bits |

     00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F
000  00 00 00 25 03 04 00 01 01 01 04 0A 08 FB 61 B4   ...%........ûa´
010  80 81 F6 43 A1 B1 0A 04 40 A0 00 00 0C 04 45 20   €ıöC¡±..@ .....E
020  00 00 0D 02 00 01 02 06 2F 04 00 00 00 53 03 04   ......../....S..
030  00 03 01 01 0A 01 00 08 01 00 0C 06 0C 06 32 01   ..............2.
040  00 39 01 82 0D 04 40 80 00 00 0E 04 41 40 00 00   .9.‚..@€....A@..
050  0F 04 3F 00 00 00 10 01 01 12 09 28 01 00 29 01   ..?........(..).
060  01 30 01 08 14 04 3D CC CC CD 16 04 37 D1 D7 17   .0....=ÌÌÍ..7Ñ×.
070  17 04 3D CC CC CD 18 04 41 F0 00 00 19 04 37 D1   ..=ÌÌÍ..Að....7Ñ
080  B7 17 1F 01 02 02 10 4E 04 19 03 04 00 0C 01 01   ·......N........
090  04 01 00 05 12 41 54 4D 45 47 41 38 2D 4C 4D 33   .....ATMEGA8-LM3
0A0  02 03 5F 03 04 02 0D 00 01 0A 01 01 08 04 00 00   .._.............
0B0  04 B0 0C 02 00 01 0D 02 00 01 0E 02 00 00 0F 01   .°..............
0C0  00 10 02 00 00 11 02 00 00 12 02 00 01 13 04 00   ................
0D0  00 00 05 14 04 00 00 00 05 15 01 01 16 01 01 17   ................
0E0  02 00 00 18 02 00 05 29 04 00 00 25 80 2A 01 08   .......)...%€*..
0F0  2B 01 00 2C 01 01 2D 01 00 FC 1B FF FF FF FF FF   +..,..-..ü.ÿÿÿÿÿ
```

Fig. 2: TEDS description in EEPROM memory of ATmega8.

In “TEDS Descriptions” option presents which TEDS that are stored on network node NCAP. In this paper was described the four tables required for each TIM, where each TEDS table was stored in differents files and specified based on communication interface to which it belongs. As an example of description of the file name, have, RS232_meta_TEDS, which sets the RS-232 interface and Meta-TEDS TEDS table describes what is being described.

The NCAP recognizes the TIM when both connects and transfers the UUID of the TIM to NCAP, if the identifier is not present in NCAP is done reading the TEDS and stored in a separate file. However, if the identifier UUID is present in NCAP, then the reading is performed in the TEDS itself, that is done for recognition. In Figure 3 presents a user interface for selecting modules reading the TEDS to show to user.



TEDS Description

Define interface: RS232

Define module:

● Módulo_1

○ Módulo_2

Fig. 3: Interface selection reading the TEDS.

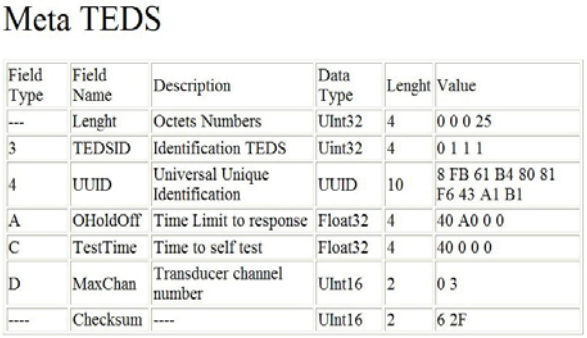
In Figure 4 shows an example of reading of the TEDS in hexadecimal.

In Figure 5 presents an example of features relating to the STIM1-RS232 module.

In Figure 6 presents the description TEDS for step motor.

In Figure 7 presents the description TEDS for interface.

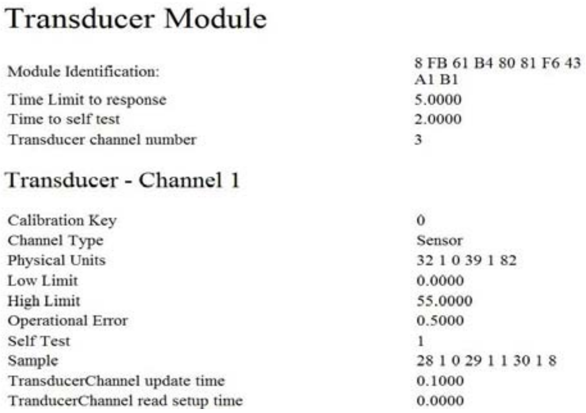
Figure 8 shows the system implemented in the laboratory with the NCAP connected to microcomputer through the Ethernet network and the two TIMs (STIM and WTIM) connected to the NCAP using the wireless and wired network.



Meta TEDS

Field Type	Field Name	Description	Data Type	Lenght	Value
---	Lenght	Octets Numbers	UInt32	4	0 0 0 25
3	TEDSID	Identification TEDS	Uint32	4	0 1 1 1
4	UUID	Universal Unique Identification	UUID	10	8 FB 61 B4 80 81 F6 43 A1 B1
A	OHoldOff	Time Limit to response	Float32	4	40 A0 0 0
C	TestTime	Time to self test	Float32	4	40 0 0 0
D	MaxChan	Transducer channel number	UInt16	2	0 3
----	Checksum	----	UInt16	2	6 2F

Fig. 4: Example of reading the TEDS stored on file in NCAP.



Transducer Module

Module Identification:	8 FB 61 B4 80 81 F6 43 A1 B1
Time Limit to response	5.0000
Time to self test	2.0000
Transducer channel number	3

Transducer - Channel 1

Calibration Key	0
Channel Type	Sensor
Physical Units	32 1 0 39 1 82
Low Limit	0.0000
High Limit	55.0000
Operational Error	0.5000
Self Test	1
Sample	28 1 0 29 1 1 30 1 8
TransducerChannel update time	0.1000
TranducerChannel read setup time	0.0000

Fig. 5: Description of TEDS for the STIM1_RS232 module.

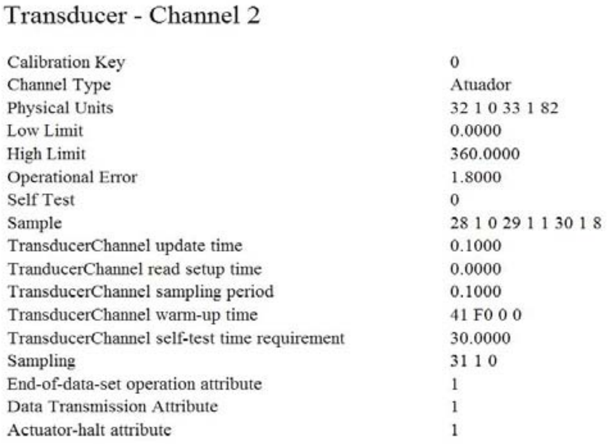
7. Conclusion

In this paper has presented the implementation of an embedded NCAP to access the TIMs through wireless and wired interfaces. The NCAP was fully implemented in the DE2 kit, using the Nios II soft-core and the uClinux embedded operating system. For this project two interfaces, wired and wireless have been considered to test the system and to implement the IEEE 1451 concepts, by means of a WTIM and a STIM module. With the implemented uClinux operating system both interfaces can be configured.

Another important aspect was the TEDS described using the file in the NCAP and it stored in TIM using the microcontroller in the flash memory, in which demonstrated the same functionality to both implementation and it becoming the TIMs plug-and-play. This approach allows introducing relevant aspects in the creation of digital systems such as code reuse and technological independence.

Acknowledgment

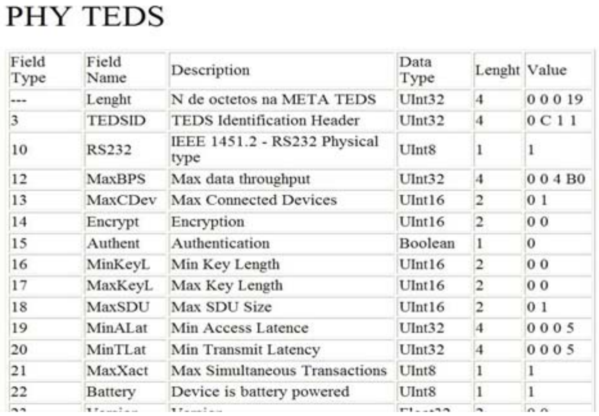
This works was carried out with the financial support of CNPq (National Council for Scientific and Technological Development).



Transducer - Channel 2

Calibration Key	0
Channel Type	Atuador
Physical Units	32 1 0 33 1 82
Low Limit	0.0000
High Limit	360.0000
Operational Error	1.8000
Self Test	0
Sample	28 1 0 29 1 1 30 1 8
TransducerChannel update time	0.1000
TranducerChannel read setup time	0.0000
TransducerChannel sampling period	0.1000
TransducerChannel warm-up time	41 F0 0 0
TransducerChannel self-test time requirement	30.0000
Sampling	31 1 0
End-of-data-set operation attribute	1
Data Transmission Attribute	1
Actuator-halt attribute	1

Fig. 6: Description of TEDS for step motor.



PHY TEDS

Field Type	Field Name	Description	Data Type	Lenght	Value
---	Lenght	N de octetos na META TEDS	UInt32	4	0 0 0 19
3	TEDSID	TEDS Identification Header	UInt32	4	0 C 1 1
10	RS232	IEEE 1451.2 - RS232 Physical type	UInt8	1	1
12	MaxBPS	Max data throughput	UInt32	4	0 0 4 B0
13	MaxCDev	Max Connected Devices	UInt16	2	0 1
14	Encrypt	Encryption	UInt16	2	0 0
15	Authent	Authentication	Boolean	1	0
16	MinKeyL	Min Key Length	UInt16	2	0 0
17	MaxKeyL	Max Key Length	UInt16	2	0 0
18	MaxSDU	Max SDU Size	UInt16	2	0 1
19	MinALat	Min Access Latence	UInt32	4	0 0 0 5
20	MinTLat	Min Transmit Latency	UInt32	4	0 0 0 5
21	MaxXact	Max Simultaneous Transactions	UInt8	1	1
22	Battery	Device is battery powered	UInt8	1	1

Fig. 7: Description of TEDS for the STIM1_RS232 module.

References

[1] ABATE, F. and Paciello, V. and Pietrosanto, A. and Guia, S.S. and Santo, A.E.; *AISEM Annual Conference, 2015 XVIII*, Period measurement with an ARM microcontroller, 2015, pages 1-4,

[2] Anuj K., Hiesik K., and Gerhard P. H.; *Environmental Monitoring Systems: A Review*, IEEE SENSORS JOURNAL, Journal, pp. 1329-1339, 2013.

[3] Yurish S.Y. *IEEE 1451 Standard and Frequency Output Sensors: How to Obtain a Broad-Based Indus-try Adoption?* Sensors and Transducers, Vol.59, Issue 9, September 2005, pp.412-418.

[4] Yurish S.Y. *Extension of IEEE 1451 Standard to Quasi-Digital Sensors, in Proc. of the IEEE Sensors Applications Symposium 2007(SAS-2007).* San Diego, California, USA, February 6-8, 2007.

[5] IEEE. *IEEE Instrumentation and Measurement Society, Sponsored by the Techinical Committee on Sensor Technology*, IEEE standard for a smart transducer interface for sensors and actuators Ű common functions, communication protocols, and transducers electronic Data Sheet (TEDS) formats,p. 330, 2007.

[6] Eugene Y. Song; Kang B. Lee . *Sensor Network based on IEEE 1451.0 and IEEE p1451.2-RS-232, IEEE Measurement Technology*, Conference, Victoria, Vancouver Island, Canada, pp. 1728-1733, May 2008.

[7] Darold Wobschall; *Network Sensor Monitoring Using the Universal IEEE 1451 Standard*, IEEE Instrumantation e Measurement Magazine, pp. 18-22, 03 April 2008.

[8] Eugene Y. Song, kang B. Lee; *An Implementation Transducer Web Service for IEEE 1451 Based Sensor System*, 2011 IEEE Sensor and Applications Symposium, San Diego, USA, 2011.

[9] Eugene Y. Song, WESTBROOK, D., LEE, K. B.; *A Prototype IEEE 1451.4 Smart Transducer Interface for Sensors and Actuators*, Electronic Measurement and Instruments (ICEMI), 2011 10th International Conference on. p. 6. 2012.

[10] Viegas, V.; Pereira, M.; GIRAO, P. *A brief tutorial on the IEEE 1451.1 standard.*, IEEE Instrumentation and Measurement Magazine, v. 13, p. 38-45, 2008.

[11] Ramos, H. G. *IEEE standard 1451 and a proposed time synchronization approach.*, IEEE Instrumentation and Measurement Magazine, Lisbon, p. 29-37, 2008.

[12] Morello, R., *Sensors Journal, IEEE*, Use of TEDS to Improve Performances of Smart Biomedical Sensors and Instrumentation, 2015, volume 15, pages 2497-2504.

[13] Deivis Borgonovo, Marcelo L. Heldwein, Samir A. Mussa; *Application of the NIOS II Processor FPGA on the Digital Control of a Single-Phase PFC Rectifier* Control and Modeling for Power Electronics, COMPEL 2008, 11th Workshop on, pp. 7, 26 September 2008.

[14] Jiangchun Xu, Jiande Wu, Yuhui Li; *A Networks Data Collection Embedded System Based on ARM-uCLinux*, 2011 2009 WASE International Conference on Information Engineering , Taiyuan, China, pp. 452-454, 7 Aug. 2009.

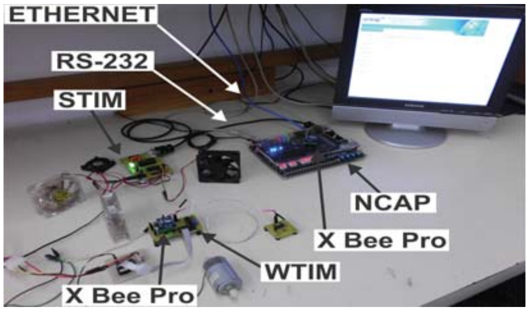


Fig. 8: Picture of the system implemented in laboratory.

PERFORMANCE PERSPECTIVE OF REAL TIME MONITORING OF AVAILABILITY OF RADIO ACCESS IN A NETWORK

Okonigene Robert[1], Ikhine Matthew[2], Samuel John[3], Agbator Austin[4]

[1,2,4]Department of Electrical and Electronic Engineering, Ambrose Alli University Ekpoma Edo State Nigeria.

[3]Department of Electrical and Information Engineering, Covenant University Otta, Ogun State, Nigeria.

Abstract - *The work presented in this paper laid more emphasis on the availability of radio signals to a consumer in a communication network. Attempts were made to have real time monitoring of several communication networks in Nigeria to ascertain the availability of essential services to a subscriber. The objective of these monitoring was to have a common characteristics behavior of all the networks, in terms of subscriber access to the services paid for. Availability as a percentage value of the amount of time the network delivered services as against the amount of time it was expected to deliver services is most critical to a subscriber. Thus, the findings reveal the facts that power outages, location of base stations and lack of security personnel were mostly responsible for lack of network availability. Also the findings reveal that the factors militating against network availability or access to network service across Nigeria are the same. This is clear violation of the policy that binds the service providers and the subscriber. Therefore, in this report we focused on these factors citing specific base station for emphasis.*

Keywords: Communication network, access to network, base station, subscribers

1 Introduction

Most companies rely on data-carrying networks such that any form of interrupt can cause considerable economic losses. As networks grow bigger and more complex, the factors influencing the network availability increased. Thus, many of these companies have invested in securing their networks with redundancy and quality of service as well as demanding high network availability from their network operators. For the network operator measuring and quantifying the network availability has become an important issue, not only to attract customers, but also as an indicator of the variation of quality in the network helping to organize maintenance and expansion of the network [1-3].

Service availability is also very important in e-business economy, customer satisfaction and corporate reputation [4- 8]. A lot of effort and improvement have been made to ensure high availability in each technology industry [9-11]. In this work we focus on a process that specifically define and measure access to network availability in Nigeria. We studied the compliance of four major GSM service providers to their policies as it relates to the subscribers.

2 Methodology and Discursion

Comprehensive on-site Monitoring of Base station performance using real-world scenarios both at initial installation and then during ongoing maintenance, plays a vital role in identifying and preventing performance problems. Poorly performing base stations have significant negative impact on the quality of service (QoS) experienced by users, particularly the higher data rate services available on 3G networks. Whether poor network performance is caused by incorrect installation, hardware/software incompatibility, gradual degradation or complete failure of a particular module, the end result is that the subscriber experience will be less than satisfactory and network operator revenues adversely affected. Making sure a cell site works in line with designed specifications ensures that problems are isolated before the network goes "bad". This is also the most cost effective monitoring solution as it is harder to troubleshoot when a cell site is active and some performance issues may not actually become visible until network capacity limits are tested. In carrying out this research work we interviewed a number of field support staff (FSO) from different network providers across Nigeria as it relate to epileptic power supply. Inefficient power supply is the major causes of poor performance of base stations in Nigeria. In general, we monitored the performance and consequently measured the Network availability from base stations across Nigeria. The studies were premeditated on the following questions identified:

- How is Network Availability defined?
- How can Network Availability be measured?

- Why should Network Availability be measured?
- What standards of Network Availability exist?
- Are there any recommended values for Network Availability parameters?
- How can Network Availability measurements be applied to radio access networks in Nigeria?

2.1 Major identified factors militating against network availability

The most direct consequence of power outages is in the form of network availability. It is a common practice to observe battery back-up and generators in most of the Base stations. However, with no primary power source for more than ten to fifteen hours, it becomes impossible to ensure complete successful site operations. In Nigeria the public utility power supply is very unstable.

Prolonged power cuts result in serious amount of site outages. Key performance indicators such as Call Setup Success rate and SDCCH Blocking rate were observed to show considerable degradation. These indicators are a direct measure of network availability and accessibility experienced by the end user. In some cases, networks suffered up to 100% degradation due to the prolonged primary power source outage.

However, contrary to the expectation that base stations are built with state of the art technology we observe that most of the base stations are not built to standard. This is another factor responsible as to why outages are on the high side in Nigeria. A major failure that occurred in Asaba Region, Delta State, Nigeria, which affected fifty seven (57) Base Transmission Station (BTS) sites, was investigated. This type of major failure was also observed in several BTS sites across Nigeria. The outcome of the investigation reveal that the root cause of the failure was power outage at one of the BTS site (hub link site) bringing down the remaining fifty six BTS sites. This affected hundreds of subscribers who could not access the network services for fifty six minutes. The time of restoration of power supply depends on time of the day, availability of security personnel and location of the BTS sites. Some of the base stations are located in remote areas and when power outage occurs at Night (from 19:00 hours to 06:00 hours), in most cases no security personnel are available to escort the site engineer to the BTS site. Under this circumstance the network downtime is longer.

The following factors also significantly impact on transmission outages of network availability and these are; fibre failure, Bad weather (due to Heavy rainfall and/or wind) resulting in flapping of sites, Microwave HOPs and Antenna misalignment.

It provides a centralized network management platform for supporting telecommunication operators in their long-term network evolution and shielding the differences between various network technologies. The M2000 focuses on continuous efforts that telecommunication operators have made for network OM and inherits the existing OM experience. The sample measurements were generated from the BTSs which also provided the File Transfer Protocol (FTP) services to the Network Management System (NMS). The NMS is hosted on a central server which is connected to other network elements such as BSC, BTS, MSC, HLR etc. The server with its NMS software was configured to retrieve BTS measurements of remote MSCs and its BSCs. The NMS was used to monitor the network under review (MTN Network) and data were collected hourly for site (BTS) down count at Asaba region. The number of site going down on an hourly basis has a direct effect on the availability of GSM Network in Nigeria. From our observations it is safe to provide information about one particular network provider, which is MTN network, without any consequence on the generality of the research. Figure 1 shows the M2000 client interface (**Physical Topology** window) and Figure 2 shows the main menu screen of the software. Figure 3 is a Screen shot from M2000 mobile monitoring system showing BTS down under their various BSCs in Asaba region. Figure 4 is a screen shot showing the ouput of the software ACC monitoring tool that is used for creating trouble ticket.

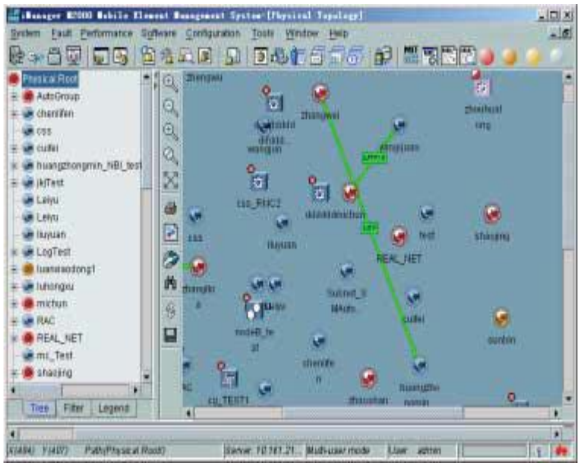
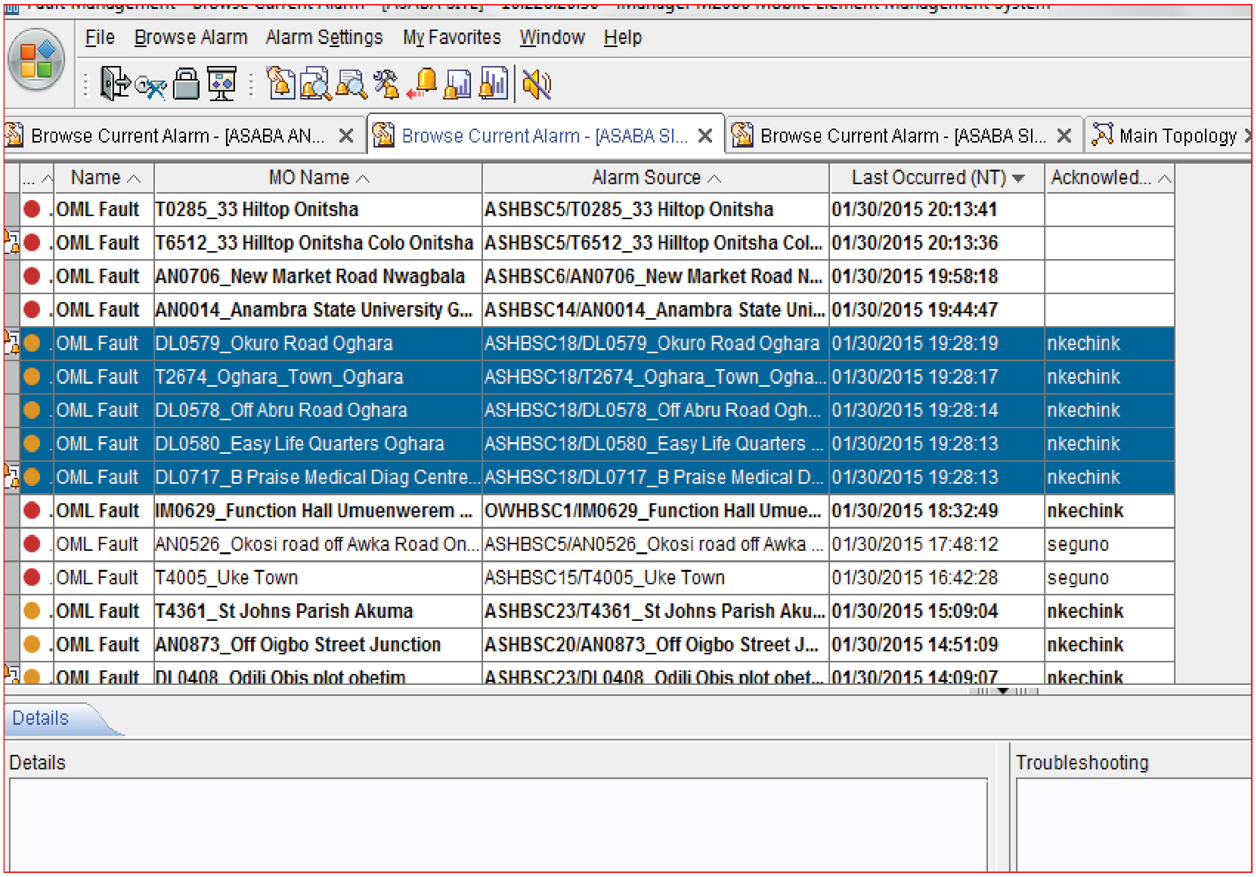


Figure 1: M2000 client interface (Physical Topology window)



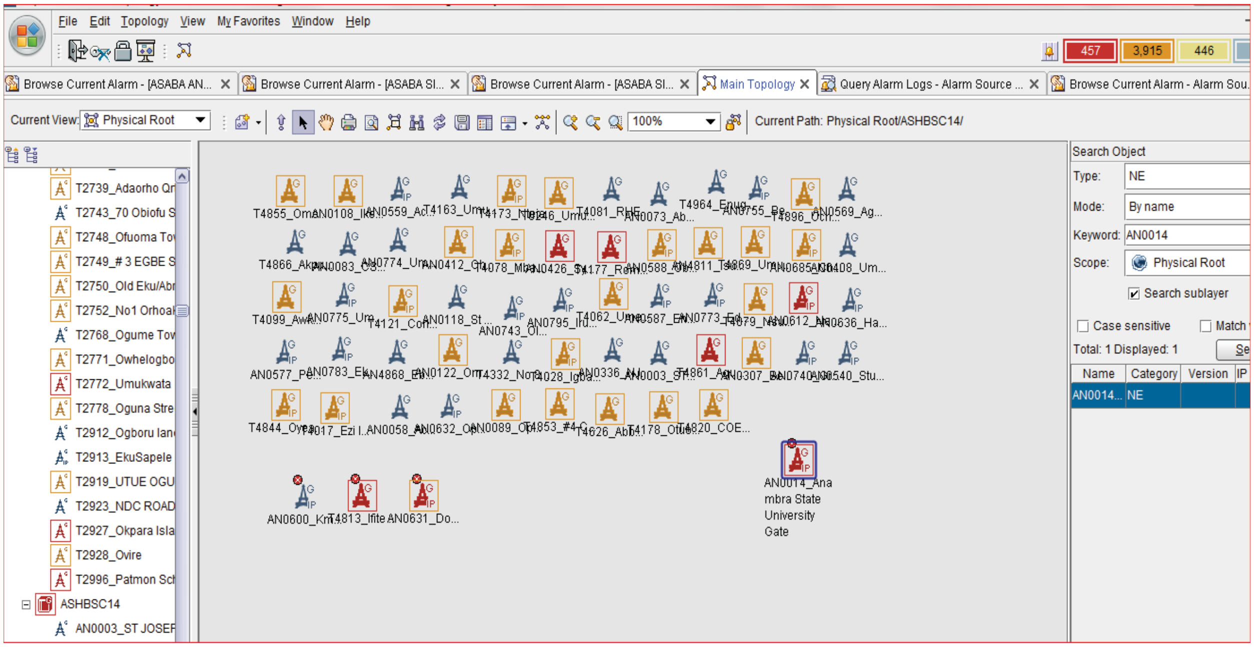
File
Browse Alarm
Alarm Settings
My Favorites
Window
Help
Browse Current Alarm - [ASABA AN...
Browse Current Alarm - [ASABA SI...
Browse Current Alarm - [ASABA SI...
Main Topology
Details

...	Name	MO Name	Alarm Source	Last Occurred (NT)	Acknowled...
●	OML Fault	T0285_33 Hiltop Onitsha	ASHBSC5/T0285_33 Hiltop Onitsha	01/30/2015 20:13:41	
●	OML Fault	T6512_33 Hilltop Onitsha Colo Onitsha	ASHBSC5/T6512_33 Hilltop Onitsha Col...	01/30/2015 20:13:36	
●	OML Fault	AN0706_New Market Road Nwagbala	ASHBSC6/AN0706_New Market Road N...	01/30/2015 19:58:18	
●	OML Fault	AN0014_Anambra State University G...	ASHBSC14/AN0014_Anambra State Uni...	01/30/2015 19:44:47	
●	OML Fault	DL0579_Okuro Road Oghara	ASHBSC18/DL0579_Okuro Road Oghara	01/30/2015 19:28:19	nkechink
●	OML Fault	T2674_Oghara_Town_Oghara	ASHBSC18/T2674_Oghara_Town_Ogha...	01/30/2015 19:28:17	nkechink
●	OML Fault	DL0578_Off Abru Road Oghara	ASHBSC18/DL0578_Off Abru Road Ogh...	01/30/2015 19:28:14	nkechink
●	OML Fault	DL0580_Easy Life Quarters Oghara	ASHBSC18/DL0580_Easy Life Quarters ...	01/30/2015 19:28:13	nkechink
●	OML Fault	DL0717_B Praise Medical Diag Centre...	ASHBSC18/DL0717_B Praise Medical D...	01/30/2015 19:28:13	nkechink
●	OML Fault	IM0629_Function Hall Umuenwerem ...	OWHBSC1/IM0629_Function Hall Umue...	01/30/2015 18:32:49	nkechink
●	OML Fault	AN0526_Okosi road off Awka Road On...	ASHBSC5/AN0526_Okosi road off Awka ...	01/30/2015 17:48:12	seguno
●	OML Fault	T4005_Uke Town	ASHBSC15/T4005_Uke Town	01/30/2015 16:42:28	seguno
●	OML Fault	T4361_St Johns Parish Akuma	ASHBSC23/T4361_St Johns Parish Aku...	01/30/2015 15:09:04	nkechink
●	OML Fault	AN0873_Off Oigbo Street Junction	ASHBSC20/AN0873_Off Oigbo Street J...	01/30/2015 14:51:09	nkechink
●	OML Fault	DL0408_Odili Obis plot obetim	ASHBSC23/DL0408_Odili Obis plot obet...	01/30/2015 14:09:07	nkechink

Details

Troubleshooting

Figure 2: Screenshot from M2000 monitoring system showing site down and their downtime.



Current View: Physical Root

Current Path: Physical Root/ASHBSC14/

Search Object

Type: NE

Mode: By name

Keyword: AN0014

Scope: Physical Root

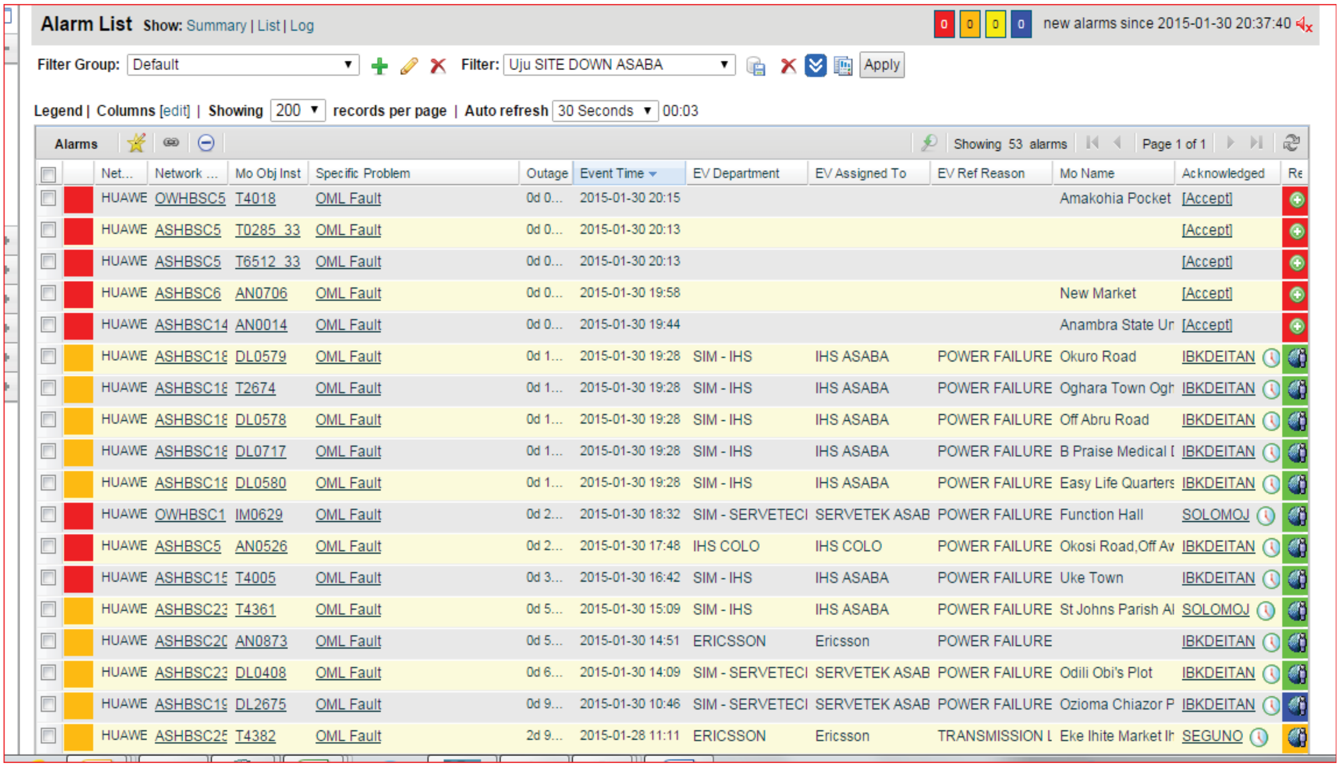
☑ Search sublayer

☐ Case sensitive ☐ Match

Total: 1 Displayed: 1

Name	Category	Version	IP
AN0014...	NE		

Figure 3: Screen shot from M2000 mobile monitoring system showing BTS down under their various BSCs



Alarm List Show: Summary | List | Log — 0 0 0 0 new alarms since 2015-01-30 20:37:40

Filter Group: Default ▾ Filter: Uju SITE DOWN ASABA ▾ Apply

Legend | Columns [edit] | Showing 200 ▾ records per page | Auto refresh 30 Seconds ▾ 00:03

Alarms — Showing 53 alarms — Page 1 of 1

□		Net...	Network ...	Mo Obj Inst	Specific Problem	Outage	Event Time ▾	EV Department	EV Assigned To	EV Ref Reason	Mo Name	Acknowledged	Re
□	🟥	HUAWE	OWHBSC5	T4018	OML Fault	0d 0...	2015-01-30 20:15				Amakohia Pocket	[Accept]	🟥
□	🟥	HUAWE	ASHBSC5	T0285_33	OML Fault	0d 0...	2015-01-30 20:13					[Accept]	🟥
□	🟥	HUAWE	ASHBSC5	T6512_33	OML Fault	0d 0...	2015-01-30 20:13					[Accept]	🟥
□	🟥	HUAWE	ASHBSC6	AN0706	OML Fault	0d 0...	2015-01-30 19:58				New Market	[Accept]	🟥
□	🟥	HUAWE	ASHBSC14	AN0014	OML Fault	0d 0...	2015-01-30 19:44				Anambra State Ur	[Accept]	🟥
□	🟧	HUAWE	ASHBSC18	DL0579	OML Fault	0d 1...	2015-01-30 19:28	SIM - IHS	IHS ASABA	POWER FAILURE	Okuro Road	IBKDEITAN	🟩
□	🟧	HUAWE	ASHBSC18	T2674	OML Fault	0d 1...	2015-01-30 19:28	SIM - IHS	IHS ASABA	POWER FAILURE	Oghara Town Ogh	IBKDEITAN	🟩
□	🟧	HUAWE	ASHBSC18	DL0578	OML Fault	0d 1...	2015-01-30 19:28	SIM - IHS	IHS ASABA	POWER FAILURE	Off Abru Road	IBKDEITAN	🟩
□	🟧	HUAWE	ASHBSC18	DL0717	OML Fault	0d 1...	2015-01-30 19:28	SIM - IHS	IHS ASABA	POWER FAILURE	B Praise Medical I	IBKDEITAN	🟩
□	🟧	HUAWE	ASHBSC18	DL0580	OML Fault	0d 1...	2015-01-30 19:28	SIM - IHS	IHS ASABA	POWER FAILURE	Easy Life Quarters	IBKDEITAN	🟩
□	🟥	HUAWE	OWHBSC1	IM0629	OML Fault	0d 2...	2015-01-30 18:32	SIM - SERVETECI	SERVETEK ASAB	POWER FAILURE	Function Hall	SOLOMOJ	🟩
□	🟥	HUAWE	ASHBSC5	AN0526	OML Fault	0d 2...	2015-01-30 17:48	IHS COLO	IHS COLO	POWER FAILURE	Okosi Road,Off Av	IBKDEITAN	🟩
□	🟥	HUAWE	ASHBSC15	T4005	OML Fault	0d 3...	2015-01-30 16:42	SIM - IHS	IHS ASABA	POWER FAILURE	Uke Town	IBKDEITAN	🟩
□	🟧	HUAWE	ASHBSC23	T4361	OML Fault	0d 5...	2015-01-30 15:09	SIM - IHS	IHS ASABA	POWER FAILURE	St Johns Parish Al	SOLOMOJ	🟩
□	🟧	HUAWE	ASHBSC20	AN0873	OML Fault	0d 5...	2015-01-30 14:51	ERICSSON	Ericsson	POWER FAILURE		IBKDEITAN	🟩
□	🟧	HUAWE	ASHBSC23	DL0408	OML Fault	0d 6...	2015-01-30 14:09	SIM - SERVETECI	SERVETEK ASAB	POWER FAILURE	Odili Obi's Plot	IBKDEITAN	🟩
□	🟧	HUAWE	ASHBSC19	DL2675	OML Fault	0d 9...	2015-01-30 10:46	SIM - SERVETECI	SERVETEK ASAB	POWER FAILURE	Ozioma Chiazor P	IBKDEITAN	🟦
□	🟧	HUAWE	ASHBSC25	T4382	OML Fault	2d 9...	2015-01-28 11:11	ERICSSON	Ericsson	TRANSMISSION L	Eke Ihite Market Ih	SEGUNO	🟧

Figure 4: Screen shot from ACC monitoring system for creating trouble ticket (TT) for the various alarms on the BTS

3 Conclusions

This study reveals that the most common cause of lack of availability of network to a subscriber is power failure. Even though all the BTS are provided with generators sometimes it takes as long as 10 hours to restore network service. Sometimes subscribers are unable to have access to network service and as a consequence are unable to communicate with one another. Under these circumstances the service providers cannot guarantee efficient platform for e-learning. Sometimes customers in banks are delayed for a long time due to lack of availability of network. When power failure occurs in these BTS, some businesses, such as online transactions, are put on hold. The service providers spend millions of Naira monthly to run their generators in all the BTS sites. As a result their annual profit margin is reduced.

4 References

[1] Mathias Thulin September (2004), Measuring Availability in Telecommunications Networks

[2] Cisco systems, Availability measurement 2004

[3] Dr.-Ing. Nikola Milanovic, berlin (2010), Models, Methods and Tools for Availability Assessment of IT-Services and Business Processes

[4] *David M. Fishman* (2000), Application Availability: An Approach to Measurement

[5] Ericsson Radio Systems AB, Basic Concepts of WCDMA Radio Access Network 2001

[6] Evan Marcus (2003), Blueprints for High Availability Second Edition

[7] Huawei Technologies, M2000 operators guide Issue 02 (2006-10-31)

[8] Jihong Zeng December (2008), A Case Study on Applying ITIL Availability Management Best practices

[9] Janet Kreiling, High Availability Networking, Packet Magazine p.54, Volume 15, No. 3, 2003

[10] Joseph Isabona September (2013), Real Time Monitoring of Service Quality of a Deployed sHow Cisco IT-LAN-SJ Achieved High Availability, Cisco Whitepaper,

SESSION

WIRELESS SENSOR NETWORKS

Chair(s)

TBA

Wireless Sensor Network Wormhole Detection using an Artificial Neural Network

Mohammad Nurul Afsar Shaon and Ken Ferens
Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba, Canada
{Shaonmna@myumanitoba.ca, Ken.Ferens@umanitoba.ca}

Abstract— This paper presents an innovative wormhole detection scheme an using artificial neural network for wireless sensor networks (WSNs). Most detection schemes described in the literature are designed for uniformly distributed sensors in a network, using statistical and topological information and special hardware. However, these schemes may perform poorly in non-uniformly distributed networks. Accordingly, the aim of the proposed research is to detect wormhole attacks for both uniform and non-uniform network environments. Furthermore, the proposed research does not require any special hardware to discover wormhole attacks and causes no significant communication overhead as well. The efficacy of the proposed detection model is measured in terms of detection accuracy, false positive rate, and false negative rate. The results show that the proposed algorithm achieves higher detection and lower false positive rates in comparison with existing statistical wormhole detectors.

Keywords—Artificial neural network; wormhole attack; non-uniform distribution; wireless sensor networks

1. INTRODUCTION

A wireless sensor network is simply a pool of self-directed devices organized into a mutually connected network**.** Sensors are usually autonomous and spatially distributed within a certain area to monitor targeted physical and environmental conditions, such as temperature, sound, and pressure. In WSNs, free frequency band and open architecture are used for supporting mission critical application in a hostile environment; thus, they are highly prone to various security attacks, such as the wormhole attack.

The wormhole attack is recognized as one of the most detrimental security threats for WSNs [1]. In WSNs, known communication channel is used so that the wormhole attack can be deployed silently without raising any security concerns. Wormhole attackers (node) are connected via virtual tunnel which can be established in many ways (e.g. out of bound hidden channel, packet encapsulation and high powered transmission) [2]. During this attack, a malevolent node, which is controlled by an adversary, records packets from one location in the network, and replays them in another location through a virtual tunnel to another malevolent node. As shown in Fig. 1, the two wormhole nodes E_1 and E_2, connected by a dedicated link, can capture the packets from one location and replay them to another location.

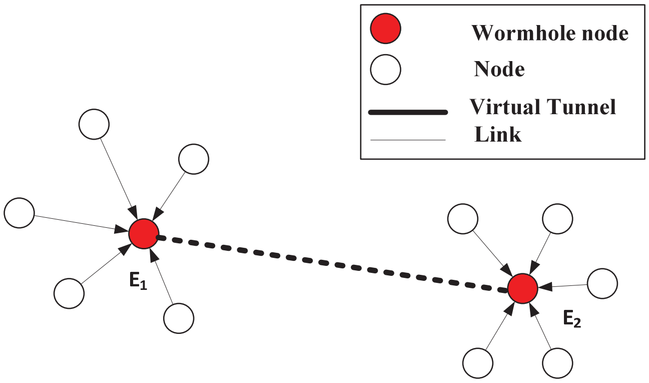


Fig. 1 Wormhole attack.

Subsequently, this wormhole attack becomes so severe that it might destroy the network or hamper the usual operation of the network by selective dropping of packets; manipulation of traffic; or modifying data packets without revealing their identities.

Therefore, detection of wormhole nodes is an essential task for ensuring the security of wireless sensor networks. Most of the existing countermeasures use distance between nodes, direction, and location abnormality among claimed neighbour nodes as detection features to fight against wormhole attack. To gain a certain level of accuracy, many existing schemes have used complex and highly advanced devices such as directional antenna [3], GPS [4], or ultra sound for distance measurement [5]. In fact, those special devices are very costly for practical deployment. Some statistical wormhole detection schemes based on hop count

[6], node connectivity [7], or neighbourhood count [8][9] do not need any special hardware. Those schemes are usually used with hardware supported scheme as a secondary approach. Furthermore, centralized statistical wormhole detector [8] caused significant network and communication overhead in contrast to distributed approach statistical approach [9]. However, most of the wormhole detection schemes are made to apprehend wormhole nodes where sensor nodes are distributed uniformly, but their performance in case of non-uniformly distributed networks is in question.

In recent years, artificial intelligence technology is combined with network anomaly detection scheme to improve its detection accuracy. It is one of the exemplary intelligent models that is extensively used in a detection system. However, an artificial neural network is a very simplified model of the information processing in the human brain. This network consists of interconnected processing units, works in a parallel fashion to find non-linear solution to a particular problem. Its adaptive and self-learning criteria helps to increase the competence of an anomaly detection model [10].

In this paper, we propose a novel detection scheme based on an artificial neural network using neighborhood count. The proposed detection model is able to detect wormhole attacks in non-uniform sensor distributions and does not need any special hardware. Here, we have introduced a mobile node, called as detector node (D_N) that visits a random location within the network area and collects neighborhood counts. When D_N moves into a wormhole attack zone, the collected number of neighbors by D_N are increased abruptly (uniform network scenario) or slightly (non-uniform network scenario) compared to non-affected zone. This abnormality is captured by D_N as evidence of the presence of wormhole attack and gathered in a dataset. D_N collects the number of neighbors both in the presence and absence of wormhole nodes. Dataset is used for training and testing of neural network. After training phase, test dataset is fed into a neural network and based on the output of the network, we decide the existence of wormhole attack in the network.

We studied detection accuracy, false positive rate and false negative rate through simulation in detail. Our simulation results have confirmed that an artificial neural network based wormhole detector can detect wormhole attack with high precision and negligible false positive and false negative rates compared to statistical based wormhole detector.

The remainder of this paper is organized as follows: Section 2 presents the literature survey of detecting wormhole attack and its counter measure for WSNs. We discuss the artificial neural network in Section 3. The proposed artificial neural network based Detector is detailed in Section 4. The evaluation results are discussed in Section 5. Section 6 includes concluding remarks and future scope of work.

2. RELATED WORK

Numerous counter measures have been proposed to confront the wormhole attacks in WSNs. In [4], The authors have proposed packet leashes to detect wormhole nodes. Two types of packet leashes are used, such as temporal packet leash and geographical packet leash. In temporal leash (TL), a sender adds either sending time or expiration time of packet so that the receiver can verify if the packet has made a journey too far based on maximum transmission speed and time. In the geographical leash (GL), sender includes its own location (using GPS) and sending time. Using GL, the maximum distance between the sender and receiver can be estimated by receiver. This scheme can perform better if strict time synchronization and additional device like GPS are provided.

In [3], a new idea has been introduced to detect wormhole attack. A directional antenna is attached to each sensor node to detect wormhole node. According to the authors, if a sensor sends a packet in a given direction, its receiver will receive it in the opposite direction. Therefore, authenticity of neighbor can be verified by their sending and receiving directions. This scheme appears to require additional hardware (i.e., directional antenna).

The method in [7] detects wormhole node by looking at the connectivity graph for forbidden substructures. Two non-neighbor nodes might have at most f_k common independent k-hop neighbors; attack is spotted if the opposite happens. Compared to dense network, forbidden substructures are very hard to find in spare network.

Another category of wormhole detectors has been proposed based on the investigation of the statistical parameters of network, such as number of neighbors and hop count etc. In [8], the statistics of total hop count and neighbor information are monitored by the base station. If the total number of hop counts decrease dramatically or whether the number of neighbors of all nodes increases over a threshold, presence of a wormhole node is declared. However, this scheme causes significant communication and co-ordination overhead.

In [9], another statistical approach is proposed, known as SWAN approach, in which each sensor collects the recent number of neighbor**s**. Wormhole attack is identified if the current number of neighbors exhibits unusual increase compared to the previous neighborhood counts taken outside of the wormhole zones. This is a distributed approach so that it doesn't cause any overhead unlike centralized approach. However, both schemes perform better in uniformly distributed network, but their performance is in question where sensors are distributed non-uniformly.

3. Artificial Neural Network

Artificial neural network (ANN) is a computational scheme that is modeled after the human brain. ANN consists of interconnected complex information processing unit, called neuron; they work together to find non-linear solution of certain problems [11]. Neural networks learn or adopt with input examples that flows through it. However, they consist of three general layers: input layer, hidden layer and output layer. Perhaps, the Multilayer layer perception (MLP) is the most widely used scheme of neural network. Fig. 2 shows a standard multilayer feed forward network.

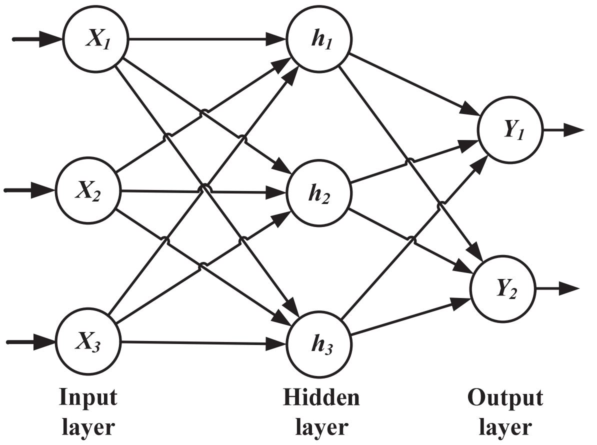


Fig. 2 Multi-layer neural network.

Operation of neural networks can be described in two phases: Training and Testing. There are several methods, but simplest and the most popular training method is generalized delta rule, which also known as backpropagation [12].

Input features from the input layer are shared with adjacent hidden layer through unidirectional branches [13]. Those input values are multiplied by some weights and then summed. Similarly, all output of hidden layer propagates to the output layer (This is called forward propagation). The value of the output layer is compared with desired output. This error between actual output and desired output are measured and propagated backward to adjust the branch weights. On other words, we minimize the cost or energy of the error function, $J(\theta)$, by using back pass of back propagation algorithm defined as:

$$J(\theta) = \frac{1}{2m}\sum_{a=1}^{m}(h_a(\theta) - y_a)^2 \tag{1}$$

The y_a defines the desired output of the a^{th} input training example, $h_a(\theta)$ represents actual output of neural network and m represents total number of training examples. During the back pass of back propagation, each branch weights is updated using (2):

$$\theta_{ij} = \theta_{ij} - \alpha \frac{dJ(\theta)}{d\theta_{ij}} \tag{2}$$

The θ_{ij} is the weight between i^{th} and i^{th} neuron and α is the learning rate. The weight adjustment procedure is performed recursively up to maximum epoch.

4. Proposed Algorithm

The proposed algorithm is a network based approach in which the number of neighbors is used as detection feature to confront wormhole attack.

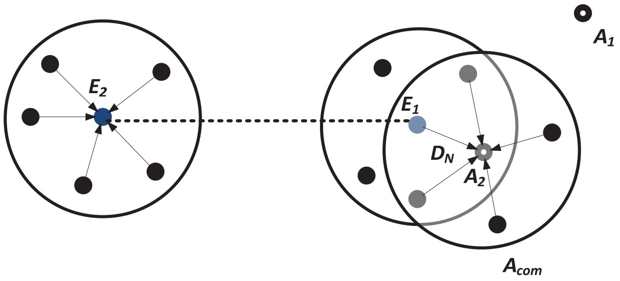


Fig. 3 Impact of wormhole attack.

However, a mobile sensor node, known as detector node (D_N) is deployed in an area where sensor nodes could be uniformly or non-uniformly distributed. D_N moves around this sensor field and collects the neighborhood count. When it reaches into the communication range of the wormhole node, counted number of neighbors would increase sharply or a little based on sensor distribution. This change is captured and gathered in a dataset, D_{set} along with other collected number of neighbors. For instance, as shown in Fig. 3, the detector node D_N moves from the one location A_1 to another location A_2. Then D_N will receive the new neighbor beacon message from sensor nodes within its communication range, A_{com}. At the same time, sensors, around the E_2, also send beacons via E_1 as they are connected through a virtual tunnel.

As we know, the performance of a neural network highly depends on how the neural network is trained and training dataset containing potential features. D_N involves in gathering D_{set} with adequate data samples. However, it is assumed that a wormhole attack does not exist in network in first half of the detection process. In this first half, detector node gathers *KxN* data samples which are called negative training examples. Similarly, same amount of neighborhood counts are collected in the presence of wormhole nodes, known as positive training examples. After that, those two types of data samples are mixed up so that training can be performed appropriately. Then *MxN* data samples are drawn and stored in a training dataset, $\underline{D_{train}}$. At the same time, *PxN* data samples from main dataset are stored in D_{test} for testing the trained neural network.

Proposed Algorithm:

1. Collect *KxN* negative Data samples
2. Collect *KxN* positive Data samples
3. Mix up the positive and negative data samples
4. Select *MxN* data samples from D_{set} and store in D_{train}
5. Select *PxN* data samples from D_{set} and store in D_{test}
6. Train the neural network with appropriate parameters
7. Test the neural network
8. If *output* ≥ 0.8 then wormhole attack exist
9. If *output* < 0.8 then wormhole attack does not exist
10. Update D_{train} by D_{test} for further training
11. Reset D_{test} and update with new data samples gathered by D_N

Furthermore, data samples of D_{train} are fed into input layer of neural network. Training procedure is performed repeatedly until it reaches the maximum epoch. Testing procedure involves the checking of the neural network whether it is able to classify the wormhole attack or not. If only the output of the trained network is greater than 0.8, then the presence of wormhole attack is declared.

After Testing, data samples of D_{test} updates D_{train} for further training by removing its old elements. This will minimize the error level that was achieved in the training phase. D_{test} entries are cleared up and updated with new data examples collected by the D_N in real time.

5. Simulation and Results

In this section, we have demonstrated the simulation and the results after applying our proposed model of detecting wormhole attacks. First phase of the experiment has been conducted to see if the proposed scheme is able to classify the wormhole attack in the network or not. In the second phase, we have evaluated the percentage of detection accuracy, false positives and false negatives of the proposed detection scheme. We have also investigated the performance of the proposed algorithm by deploying different sensor distributions.

In the simulation setup, 500 sensor nodes are distributed in the square field of 1000 meters by 1000 meters. Each sensor node including D_N has 50 meters radio range. A pair of wormholes is placed on a location of 300 meter by 300 meter and 700 meters by 700 meter. Random waypoint model is used as mobility model for the simulation[9].

A multi-layer, feed forward network with back propagation algorithm has been used for the experiments. Both Input and hidden layer contain of 100 neural nodes. On the other hand, the output layer has only one (01) neural node. Conversely, we have used a sub data set, $\boldsymbol{D_{train}}$ comprising of 9000 randomly selected data points from D_{set} for training, in which each data point consist of 100 neighborhood counts collected by D_N. During the training period, minimum error tolerance level was set to 1e-05. The Table 1 shows the parameters which are used during the training phase.

Table 1 Parameters used for training.

Parameter	Value
No of feature	1
No of Data points (training)	9000x100
No of Data points (testing)	1000x100
Architecture	[100,100,1]
Performance	1.00E-05
Learning rate,α	1.00E-02
Epoch	9000
CPU time	15 mins

In the testing phase, the testing dataset is fed into the input layer. Then we look into the output of the neural network to see whether it identifies the existence of wormhole attack in the given network. Fig. 4 shows that the proposed detection scheme can classify the “wormhole attack” and “no wormhole attack” successfully.

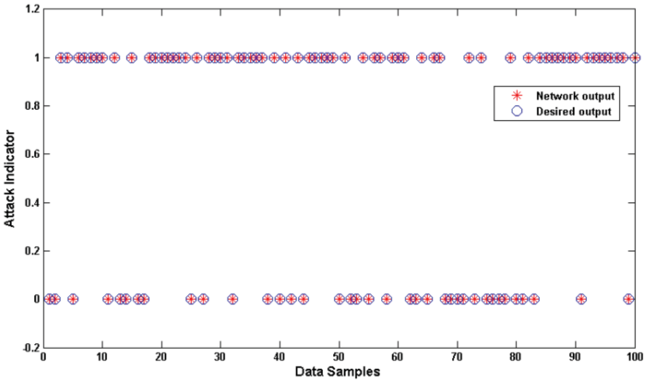


Fig. 4 Classification of wormhole attack (uniform distribution).

In the second phase, we wanted to evaluate the performance of the proposed scheme by using different sensor distributions in a given area. The total Number of sensors and radio range of each sensor including D_N remain same as the first phase. In this experiment, sensors are distributed according to several distributions such as uniform, Gaussian, Poisson, exponential, beta and gamma distribution. For each sensor distribution, we have calculated the percentage of the detection accuracy, false positives and false negatives.

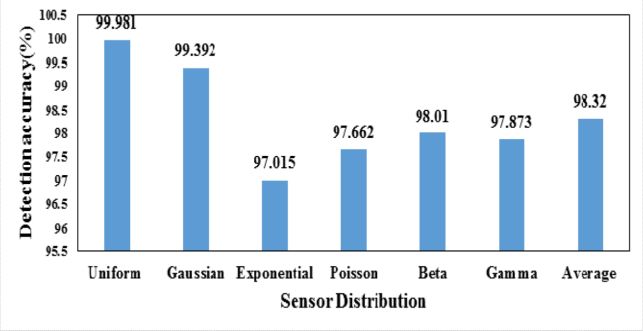


Fig. 5 Percentage of detection accuracy.

Fig. 5 shows the percentage of detection accuracy of ANN based detection scheme with different sensor distributions. In this graph, the highest detection accuracy is recorded as 99.981% when sensors are distributed according to Uniform distribution, whereas the lowest detection accuracy is measured 97.015% for exponential sensor distribution. Furthermore, detection accuracy for Gaussian distribution is almost same as the uniform distribution. Accordingly, 97.662%, 98.01%, and 97.873% detection rates are measured for Poisson, beta and gamma sensor distribution. Thus, the average detection accuracy calculated for the detection system is 98.32%.

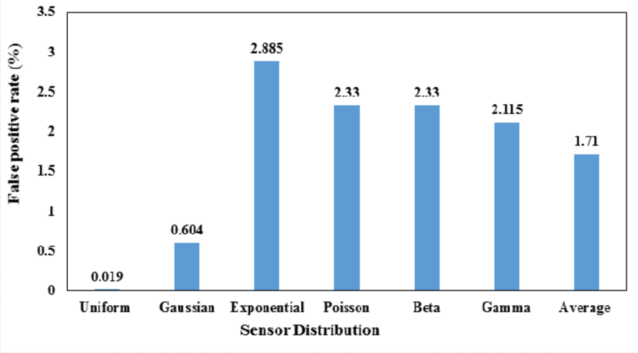


Fig. 6 Percentage of false positive.

Fig. 6 compares false positive rates of this scheme with the Variation of deployed sensor distributions. The lowest false positive rate is achieved for the uniform sensor distribution, and the highest false positive rate is recorded for the exponential sensor distribution, which are 0.019% and 2.885%, respectively. For the Gaussian sensor distribution, false positive rate is relatively low as uniform sensor distribution. The false positive rate for the beta sensor distribution is 2.115%. At the same time, false positive rates are approximately same for Poisson and Gamma sensor distribution, but not as high as Exponential sensor distribution.

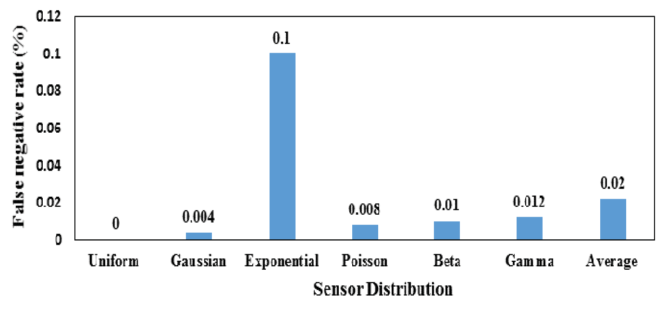


Fig. 7 Percentage of false negative rate.

Fig. 7 illustrates the changes in false negative rates over different sensor distributions. In this graph, lowest false negative 0% is obtained for uniform sensor distribution among all sensor distributions. However, false negative rate is relatively high for the exponential sensor distribution compare to other sensor distributions. Here, Gaussian and Poisson sensor distributions show almost equal false negative rates. Similarly, almost similar false negative rate is achieved when the sensors are distributed according to beta and gamma distribution.

An analysis of Fig. 5, Fig. 6 and Fig. 7 shows that the proposed detection scheme can perform better for the Uniform sensor distribution in contrast to the other non-uniform sensor distributions; though its performance is quiet improved compared to other existing wormhole detectors. In uniform sensor distribution, number of neighbors is increased abruptly when detector node is in the communication range of wormhole node. In contrast, the number of neighbors increases slightly or very small in magnitude when sensors are distributed non-uniformly. Therefore, detector node collects the number of neighbors as the evidence of wormhole attack more precisely in uniform sensor distribution compared to non-uniform sensor distributions.

In Fig. 8, we compare the performance of proposed detection scheme with other existing statistical wormhole detector. Proposed scheme is outperformed in all performance categories except false negative rate. Proposed detection scheme has higher detection accuracy with lower false positive rate, which are accordingly 98.25% and 1.71%.Though proposed scheme exhibits 0.02% false negative rates unlike other two statistical wormhole detectors, this false negative rate is apparently negligible considering its high detection accuracy and lower false positive rate.

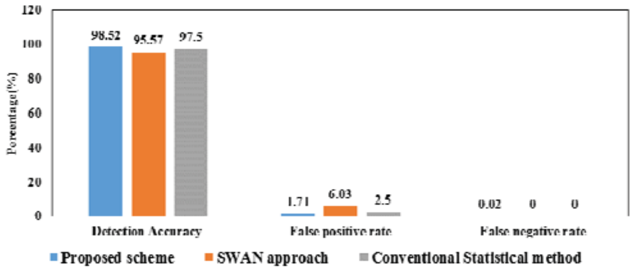


Fig. 8 Comparison of different statistical wormhole detectors.

6. Conclusions

This paper presents a novel detection model based on neighborhood count using ANN for wireless sensor networks. The goal of this proposed detection scheme is to detect wormhole attacks in any sensor distribution with high detection accuracy and low false positive rate, especially in non-uniform network environment. The proposed scheme shows promising performances through simulation. The detection accuracy is increased and false positive is decreased significantly compared to other statistical wormhole detectors. Future work is needed to enhance the detection scheme to locate wormhole nodes and to eradicate them from any sensor network.

References

[1] Y. Hu, A. Perrig, and D. B. Johnson, “Wormhole Attacks in Wireless Networks,” vol. 24, no. 2, pp. 370–380, 2006.

[2] M. E.-S. Marianne Azer, Sherif El-Kassas, “A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in wireless Ad Hoc Networks,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 1, no. 1, pp. 41–52, 2009.

[3] L. Hu and D. Evans, “Using Directional Antennas to Prevent Wormhole Attacks,” *Netw. Distrib. Syst. Symp. NDSS*, no. February, pp. 1–11, 2004.

[4] Y.-C. Hu, a. Perrig, and D. B. Johnson, “Packet leashes: a defense against wormhole attacks in wireless networks,” *IEEE INFOCOM 2003. Twenty-second Annu. Jt. Conf. IEEE Comput. Commun. Soc. (IEEE Cat. No.03CH37428)*, vol. 3, no. C, pp. 1976–1986, 2003.

[5] N. Sastry, U. Shankar, and D. Wagner, “Secure verification of location claims,” *Proc. 2003 ACM Work. Wirel. Secur. WiSe 03*, vol. 0, no. Section 2, pp. 1–10, 2003.

[6] N. Song, L. Qian, S. Ning, Q. Lijun, and L. Xiangfang, “Wormhole attacks detection in wireless ad hoc networks: a statistical analysis approach,” *Parallel Distrib. Process. Symp. 2005. Proceedings. 19th IEEE Int.*, p. 8 pp., 2005.

[7] Y. Xu, G. Chen, J. Ford, and F. Makedon, “Detecting wormhole attacks in wireless sensor networks using connectivity information,” *Crit. Infrastruct. Prot.*, vol. 2006, 2007.

[8] L. Buttyán, L. Dóra, and I. Vajda, “Statistical Wormhole Detection in Sensor Networks,” *Secur. Priv. Adhoc Sens. Networks*, pp. 128–141, 2005.

[9] S. Song and H. Wu, “Statistical Wormhole Detection for Mobile Sensor Networks,” pp. 322–327, 2012.

[10] J. Tian, M. Gao, and F. Zhang, “Network Intrusion Detection Method Based on Radial Basic Function Neural Network,” *2009 Int. Conf. E-bus. Inf. Syst. Secur.*, pp. 1–4, 2009.

[11] D. Devaraj, J. P. Roselyn, and R. U. Rani, “Artificial neural network model for voltage security based contingency ranking,” *Appl. Soft Comput. J.*, vol. 7, pp. 722–727, 2007.

[12] N. Etwork, “Anomaly Detection Using Artificial Neural Network,” vol. 2, no. 1, pp. 29–36, 2012.

[13] P. G. Kumar and D. Devaraj, “Network Intrusion Detection using Hybrid Neural Networks,” *2007 Int. Conf. Signal Process. Commun. Netw.*, pp. 563–569, 2007.

S-LEACH: A LEACH extension for Shared Sensor Networks

Gabriel Caldas, Claudio M. de Farias, Luci Pirmez, Flávia C. Delicato
PPGI
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brazil - 21941-901
{gcaldas08, cmicelifarias, luci.pirmez, fdelicato}@gmail.com
Corresponding author: Gabriel Caldas

***Abstract*— A new trend on Wireless Sensor Network field is the emergence of SSN (Shared Sensor Networks). SSN consists on multiple applications sharing the same sensoring and communication infrastructure. A major challenge for SSN is to extend the network's lifetime, since multiple applications potentially impose increased demand from the sensing and communication infrastructure. A promising solution to this challenge is the creation of algorithms that are capable to share the formation of clusters created by existing cluster-based routing algorithms in order to fully exploit the fact that a same sensing unit meeting the requirements of different applications is able to collect data only once and the collected result is shared by all of them. In this context, we propose an extension of LEACH algorithm [3], called S-LEACH (Shared LEACH). S-LEACH is able to handle the sensing requirements of several applications in SSN without executing redundant data collection. Our proposal is validated by experiments.**

Keywords*—*LEACH, S-LEACH, Shared networks

I. INTRODUCTION

Wireless sensor networks (WSNs) are composed of: (i) Sensors that are low-cost, energy and resource constrained devices equipped with sensing interfaces and communication capabilities via wireless links; and (ii) one or more Sink node(s), a device that works as a gateway between the WSN and external networks. WSNs contain a large number of sensor nodes working collaboratively to monitor the target area and perform actions that may actively affect the physical environment [12]. In general, the sensor nodes are typically deployed in difficult-to-access remote locations. Most deployments of WSNs require unattended operation, thus, sensor nodes have to rely on batteries for communication and information gathering. A new trend on WSN field is the emergence of SSN (Shared Sensor Networks) [1] [9]. In SSN multiple applications share the same sensoring and communication infrastructure.

A major challenge for SSNs is to extend the network's lifetime, since multiple applications impose a a greater demand from the sensing and communication infrastructure. One of the techniques to extend the lifetime of a WSN consists in creating an hierarchy of clusters in order to route the collected data [2]. Those cluster-based routing algorithms are responsible for organizing the network in groups, called clusters, whose members are: cluster leader, called Cluster Head (CH), and sensor nodes, called cluster members (CM). The main role of a CH is to route the collected data from the sensors of its cluster towards the sink node through multihop communication. Since data communication is an energy-demanding operation and the overall distance among cluster members and its respective cluster-head is generally smaller than the distance among these cluster members and Sink node, cluster members save transmission energy and thus extend the network operational lifetime [3]. Some clustering techniques uses distance criteria [8] [9] such as received signal strength indicator (RSSI) and shortest communication distance among others to form clusters.

There are several techniques to extend the lifetime of a WSN. Most of them focus on the fact that communication between the sensor nodes and the base station is expensive, and so intends to reduce the number of transmissions. Such techniques search to organize the WSN in order to reduce the number of hops between the collected data and the Sink Node or to aggregate data to reduce transmissions. One of the most famous and widely adopted cluster-based routing algorithm for WSN is LEACH (Low-Energy Adaptive Clustering Hierarchy.

In the SSN scenario a cluster-based routing algorithm will have to deal with several applications simultaneously sharing the same infrastructure. Then, in a SSN, it is possible that a same sensing unit meets the requirements of different applications. This is potentially efficient since a same sensing unit could collect data only once and then share the results with several applications so as to further improve the use of limited node resources. So, a challenge for SSNs is related to the improvements and adaptation of existing cluster-based routing algorithms in order to fully exploit the fact that a same sensing unit that meet the requirements of different applications could collect data only once and the collected result be shared by all of them.

This paper proposes an application-aware extension of LEACH for SSN, called S-LEACH (Shared LEACH). S-LEACH is an application aware cluster-based routing algorithm for SSN because is designed to deal with several applications simultaneously sharing the same infrastructure of wireless sensor network. Therefore, in S-LEACH the clusters formation is created in order to route the data for multiple applications by transmitting these data once. Besides, by considering a context of shared applications in a common

This work is partly supported by the Brazilian Funding Agencies CNPq and FAPERJ under grants numbers FAPERJ - E-26/110.468/2012; CNPq - 307378/2014-4; CNPq 304941/2012-3; CNPq - 473851/2012-1; INMETRO - PRONAMETRO ; CNPq 477223/2012-5 (Universal ASGARD Project).

sensors infrastructure, the CH nodes of S-LEACH use data fusion algorithms designed for SSN [12] instead of traditional fusion techniques. The major problem with traditional fusion techniques when applied to SSNs is that they consider the universe (sensing units and data rates) of a single application. If the environment has a set of applications simultaneously running (as it is in a SSN scenario) and they use the same sensing unit (such as temperature), each application will apply fusion for each set of data instead of fusing these data for all applications. So, in order to deal with SSN scenarios, this work uses fusion techniques designed for SSN [12]. These fusion techniques designed for SSN reduce the number of collected data and the number of transmitted messages on SSNs, once that each sensor will send a single message independently of the number of applications, further extending the network lifetime [12].

The main benefits of using S-LEACH are: (i) S-LEACH extends the lifetime of SSN when compared to LEACH, since it reduces data traffic, by instead of transmitting the collected data of a same sensing unit that may meet the requirements of different applications several times, as LEACH would do, S-LEACH transmits this data only once for these applications; (ii), S-LEACH does not have a negative impact over the sensor resources (memory size). Our proposal is validated through simulations and tests on real nodes.

The rest of this paper is organized as follows. Section II reviews related works. Section III presents our designed clustering technique for SSN: S-LEACH. In Section IV, we describe the experiments to evaluate the proposal. And finally, at Section V, we conclude our paper and outline future works.

II. Related work

Several works have proposed energy-efficient cluster-based routing algorithms for WSNs [3] [4] [5] [6] [7]. Our main related work is [3], presenting LEACH, an adaptive self-organizing cluster-based routing algorithm for WSN. The main idea presented in [3] is the creation of clusters in a distributed random way over the network. During each round a set of Cluster-Heads (CHs) is elected at random in a distribute way. Based on node proximity each non-CH node joins the nearest newly elected CH. After the CH election and the formation of clusters, LEACH starts to perform steady-state phase, when data related to the application are transferred to the base-station. LEACH [3] is one of the best known cluster-based routing algorithm for WSNs and is extended by the works T-LEACH [4], TL-LEACH [5] and Pegasis [6], which basically changes the criteria used to form clusters (the received signal strength indicator (RSSI) and shortest communication distance are example of criteria) in order to extend the network lifetime. Thus LEACH [3] as its extensions [4], [5] and [6] were designed to meet the requirements of a single-application. The difference among our proposal and LEACH [3] and its extensions [4], [5] and [6] is that our proposal extends LEACH to handle the requirements of multiple applications in a SSN context and use data fusion algorithms tailored for SSN scenario, while [3] [4] [5] [6] are cluster-based routing algorithms for WSNs, not tailored for a shared sensor network infrastructure (SSNs).

The work [8] proposes a reconfigurable semantic middleware, capable of handling data from multiple and heterogeneous applications. The proposed middleware [8] provides a precise and formal semantic value for each data; also allowing integration of different applications and the share of data that belongs to the same context. The proposed semantic middleware uses ontology to process information from each sensor node as well information related to the current network state. The proposed middleware promotes data enrichment and adds an automatic association between the meaning of the data, temporal and spatial aspects. The work suggests that the energy economy occurs through the decrease of communication, and pre-processing and collecting data only when are considered significant. Also, [8] adds semantic value to each data and associates it with the context of each application, keeping unrelated nodes in sleep mode, thus saving energy. The main difference from [8] to our proposal is that S-LEACH transmits this data only once for the several applications, saving energy. Also, we use data fusion algorithms suited for SSN.

To the best of our knowledge, there are some proposed works that deal with the clustering challenge for SSNs, but they present various restrictions that these clusterings imposed on applications. The main difference between S-LEACH and other clustering approaches found in literature is that our proposal relies on the creation of clusters for a shared infrastructure, while other clustering approaches are concerned with a single application. We also use data fusion algorithms designed for SSNs scenario in order to reduce data traffic, extending further network lifetime, since instead of transmitting the same data several times (each one of the applications), as LEACH would do, S-LEACH transmits this data only once for the several applications.

III. S-LEACH

Our proposal, S-LEACH (Shared LEACH) extends LEACH to serve multiple applications efficiently. We have modeled the applications as a tuple, where an application App_i = <*Ident*, *Sts*, *Srs*>, where *Ident* represents the identification of the application, $Sts = <St_1, St_2, ..., St_n>$ represents the set of sensing units (for example, sensing temperature or humidity) required by the application and $Srs = <Sr_1, Sr_2, ..., Sr_n>$ represents the set of sensing rates that each sensing unit should perform to meet the requirements of this application.

This section is organized as follows: Section A presents an overview of the proposed algorithm and Sections B, C, D and E describe the phases of our algorithm.

A. S-LEACH Overview

S-LEACH is performed by all the SSN nodes. Our algorithm is performed in a periodic basis (similarly to LEACH). Each period is a round. In S-LEACH each round encompasses a sequence of procedures performed in three mains phases: the *Set-up Phase*, the *Application Awareness Transient State Phase* and the *Steady-State Phase*.

The Set-Up Phase (Section B) is responsible for setting the algorithm initial parameters, selecting the Role for each node, Cluster Head (CH) or Cluster Member (CM), and creating the clusters. The procedures of Role Selection and

Clusters Formation performed during this phase are the same as LEACH [3].

The Application Awareness Transient State Phase (Section C) is responsible for the network nodes to be aware of several currently running applications in network. Additionally this phase of S-LEACH shares the clusters formation among several applications in order to the data collection of a same sensing unit be shared among the applications that demands the data generated by this sensing unit. Also, by performing this phase, S-LEACH can save energy of the sensor node that does not meet at least one of the requirements of the currently running applications or there is no currently running application on the network, because it keeps the nodes that do not attend an application on sleep state.

In ***the Steady-State Phase*** (Section D), each CM that meets at least one of the requirements of the currently running applications on the network should collect and send the collected data for the respective CH. Each CH executes data fusion algorithms especially designed to deal with multiple applications simultaneously in the SSNs context [12].

B. Set-up phase

S-LEACH is a periodic algorithm that works in rounds. The first phase of each round is the Set-Up phase. All nodes should perform this phase at the same time, so the nodes must be synchronized. The clock synchronization problem is reported and addressed by several authors such as [3] and [8]. Considering the nodes synchronized, the *Set-Up phase* starts. This phase is divided in three procedures: (i) *Configurations*, (ii) *Role Selection* and (iii) *Clusters Creation*. The Role Selection and Clusters Creation procedures are the same as LEACH [3].

In the ***Configurations procedure*** all configurations are deployed to the nodes and the data structures are populated. Each CM node contains the following data structures: *SC, S, NV, L, CNI, A*, *FTU* and *APPs.*

The data structure *SC* is used by the sensor nodes in order to store its capacities. The *capacities* of the *sensor* node are the set of *Sts* (sensing units) and their respective *Srs* (sensing rates) that this node is able to perform and support the applications that demand these capacities. For each sensing unit, the quantity of data collected by the sensing unit (*St*) and its values is stored in the data structure *L*. The data structure *CNI* is is defined for each CM node in order to store the unique *network addresses* (NodeID) of the CH of its cluster. The data structure *APPs* is defined for all network nodes in order to contain all currently deployed applications in SSN. The sensor node uses the data structure *S* in order to store the identification of the applications that are supported by the sensor node. The data structure *A* stores the set of *St* and *Sr* for each application that is running on network and this node, when playing the CM node role, is able to attend its requirements. It is important to notice that differently from *A* data structure, *APPs* contains all currently running applications, not just the supported ones

Each CH node contains the following data structure: *NV* and *FTU*. The data structure *NV* stores the CM nodes that a certain CH has in its own cluster. This data structure has a field in order to store the size of the cluster, which is equal to its number of CM, and the rest of the structure is used to store the identifications (NodeID) of each CM node. The CH node stores in the *FTU* data structure the *identification* (id) of the *fusion technique designed for SSN* that should be performed given a set of currently running applications on network. As an example of a *fusion technique for SSN*, there is the *Enhanced Moving Average Filter* [13].

The ***Configurations procedure*** starts with the boot() procedure. This procedure represents the hardware initializations of each node, required for making the node operational and ready to start performing our algorithm. During the boot procedure the NodeID parameter is set. So, each node starts its operation knowing its own identification. The NodeID parameter stores a unique identification for each node in the network. After, the *init()* procedure is performed. Such procedure consists on setting the initial values of *S* (*supported applications*), *SC* (*capacities of the sensor node*) and *FTU* (*fusion technique given a set of running applications*). The other data structures are left empty and it will be filled during operation procedures of three phases of S-LEACH's, which are performed for all the nodes in the network. These nodes were already physically deployed over the structure

In the ***Role Selection procedure***, which is the same as LEACH, S-LEACH must select the roles of the nodes of the SSN as CH or CM. The role selection procedure in each node is made in the same way as in LEACH [3]. Each node chooses a random number α from the interval [0, 1] and calculates a threshold function $T(n)$, similarly to LEACH.

$$T(n) = \begin{cases} \frac{P}{1-P*(r \: mod \left(\frac{1}{p}\right))}, if \; n \in G \\ 0 \, , otherwise \end{cases} \qquad (1)$$

Where P is the desired percentage of cluster heads present in the network, r is the current round and G is the set of CM nodes in the last $\frac{1}{\mathrm{p}}$ rounds. At this point, since the ammount of nodes that are eligible to become CH decrease, the probability to become CH of the remaining nodes at G set must be increased. After $\frac{1}{\mathrm{p}} - \mathrm{P}$ rounds $\mathrm{T(n)} = 1$ for any CM node. Once a node has become CH, it does not perform $\mathrm{T(n)}$ again; in other words it is not eligible anymore. After $\frac{1}{\mathrm{p}}$ rounds all nodes are eligible again.

In ***Clusters Creation*** procedure, which is the same as LEACH, each node verifies its role. If the role is CH (line 1, Fig 1), this node broadcasts CH_ROLE_BROADCAST message for its neighboring CMs to inform its role (line 2, Fig 1). After that, this node waits for the reception of the CM_JOIN messages disseminated by its neighboring CMs (line 3, Fig 1). After the reception of all the CM_JOIN messages disseminated by its neighboring CMs, this CH creates a TDMA schedule for each CM node that joined its cluster (line 4, Fig 1). Finally, this CH sends TDMA_SCHEDULE message for each one of its CMs informing TDMA schedule (line 5, Fig 1).

If the role is CM (line 7, Fig 1), this node waits for the reception of the CH_ROLE_BROADCAST messages disseminated by the CHs (line 8, Fig 1). Upon receiving these messages, this node chooses the CHs with the strongest RSSI (Received Signal Strength Indicator) (line 9, Fig 1). Then, this node sends a CM_JOIN message to this nearest CH in order to join its cluster (line 10, Fig 1). Next, this node waits for the reception of the TDMA_SCHEDULE message informing its TDMA schedule disseminated by its CH (line 11, Fig 1).

By the end of this procedure the network is self-organized in clusters, like the original LEACH [3]. Next the Application Awareness Transient State procedure starts in order to share this clusters formation among the deployed applications.

```
Input: T(n), α (a random number α from the interval [0, 1]).
Output: Node role and NV.

1.  If T(n) > α my role is CH:
2.    Send CH_ROLE_BROADCAST message for the neighboring CMs.
3.    Wait for CM_JOIN messages from each CM that has chosen this CH node. CH updates the NV.
4.    Create a TDMA schedule for each member node that joined my cluster.
5.    Send to each CM a TDMA_SCHEDULE message.
6.  End If.
7.  If T(n) < α my role is Member Node:
8.    Wait all CH_ROLE_BROADCAST messages.
9.    Choose the CH with the best RSSI among the received CH_ROLE_BROADCAST messages.
10.   Send a CM_JOIN message for the chosen CH.
11.   Wait for CH to inform my TDMA schedule, through TDMA_SCHEDULE_MESSAGE.
12. End If.
```

Fig 1. Pseudocode of Set-Up PhaseNext the Application awareness transient state phase starts

C. The Application Awareness Transient State Phase

The pseudocode of the *Application Awareness Transient State* phase is shown in Fig 2.

```
Input: Applications to be deployed on the SSN.
Output: The set A, of the applications to be attended, SC the sensor capacities of each node.

1.  If my role is Cluster Head:
2.    Waits for the FUSION_DEPLOYMENT message from sink node informing the fusion technique to be used by the CH.
3.    Update the FTU structure, with the fusion technique to be used by a given set of currently running applications.
4.  End If.
5.  If my role is Member Node:
6.    Waits for the APP_DEPOYMENT message from sink node informing the ids of the applications to be deployed, and their respective Sts and Srs.
7.    Update the set APPS, to contain the identifications of all currently running applications on SSN, and their respective Sts and Srs.
8.    For each application identification i in set APPS:
9.      For Each sensing unit j in the set $Sts_{i,j.}$
10.       If $Sts_{i,j}$ from $App_i$ has the same $Sts_{\ ,j}$ from SC, for giving application i:
11.         For the giving application i , in set A update $Srs_{i,j}$ with highest Sr between the values of $App_i$ and SC.
12.       End If.
13.     End For each.
14.   End For each.
15.   If $A = \varnothing$:
16.     Sensor sleeps until next round.
17.   Else.
18.     Return set A, SC.
19.   End If.
20. End If.
```

Fig 2. Pseudocode of Application Awareness Transient State

First, each sensor node verifies its role (CH or CM) in network (lines 1 and 5, Fig. 2). If the sensor node is CH, it waits for the FUSION_DEPLOYMENT message informing the fusion technique to be used by it (line 2, Fig. 2). Next the CH updates the *FTU* structure with the fusion technique informed (line 3, Fig. 2).

If the role is CM, it waits for APP_DEPLOYMENT message (line 6, Fig. 2) from the *sink* node informing the currently running applications on SSN. The APP_DEPLOYMENT message contains the following fields: the number of applications to be created; for each application to be deployed, the message contains the Application 's ID (*id*), its sensing units (*Sts*), such as temperature, and its respective sensing rates (*Srs*) along with the quantity of sensing units that this application demands. Then, the node fills the set *APPS* to contain the identifications of all currently running applications (line 7, Fig. 2) in the SSN. Next, for each application in the *APPS* set (line 8, Fig. 2), the node has to verify if there is another application using the same sensing unit (line 10, Fig. 2). If the sensing unit is already being used by another application currently running in the SSN (in other words, the sensing unit already is in *Sc*), the node uses the most demanding rate for this sensing type (line 11, Fig. 2), and thus, serve all applications respecting the application that has the higher sensing rate. After performing these steps, the CM nodes that do not have any applications on set A (line 15, Fig. 2) remains in sleeping state in order to save energy (line 16, Fig. 2). Finishing the Application Awareness Transient State, the procedure has as result the sets *A* and the *SC*, containing the quantity of running applications, the identifications of the supported currently running applications on network, and for each application it stores the set of sensing tasks and sensing rates that the application needs and the *capacities* of the sensors (line 18, Fig. 2).

D. The Steady-State phase

Input: The set *A that contains all identifications of the currently running* applications to be attended.

Output: Collected data for each concurrent application.

```
1.  While the CM is in Steady-State:
2.    For each application identification i in APPS $\in$ A, do:
3.      For each sensing units $Sts_{i,j}$ of giving application i, do:
4.        Collect data for the sensing unit $Sts_{i,j}$ in the respective sensing rate $Sr_{i,j}$ for giving application i.
5.        Store the collected data of sensing unit j $Sts_j$ on set L, for giving application i.
6.        End If.
7.      End For each.
8.    End For each.
9.    CM send the SENSED_DATA_SAMPLES message for its CH containing all collected Data (content of L) during the node's TDMA time slot.
10. End While.
```

Fig 3. Pseudocode of the Steady-State on CM view

The procedure shown in Fig 3 presents the pseudocode of this phase in CM point of view. First, during the Steady-State time (line 1, Fig. 3), the CMs verifies each identification of application (line 2, Fig. 3) in set *APPS* and each sensing unit of this application (line 3, Fig. 3). To, on following, collect the data (line 4, fig. 3). Next CM stores the collected data in the corresponding sensing unit at the set *L* (line 6, Fig. 3). And finally, the CMs send its collected data to the CH node during its TDMA time slot (line 9, Fig. 3).

It is important to notice that those collected data must be sent in a single message. It happens, because the data communication procedure demands much more energy than the processing procedure [11]. It is important to note that this is one of the most important step in order to extend the network lifetime. By sending all collected data in a single message the node prevents that the number of messages on network increases as the current number of applications increases.

The procedure shown in Fig 4 presents the pseudocode of this phase in CH point of view. At this point, the CH performs the data fusion techniques. In this procedure S-LEACH uses fusion algorithms designed for the SSN scenario [12] [13]. These techniques should be used because a fusion technique when applied in SSN context should be performed considering the different characteristics of each application (such as sensing units and sensing rates).

First, the CH receives SENSED_DATA_SAMPLES messages from its CM nodes and stores these collected data in the set *L* (line 1, Fig. 4). Based on the *FTU* structure, that contains the fusion technique to be used, CH performs the fusion technique on the data in the *L* set (line 2, Fig. 4). This is represented by the function FUSION (*L, FTU*). After that, the CH node send APPS_FUSED_DATA message (hop by hop using a routing protocol such as CTP) to the sink Node containing the data fusion results (line 3, Fig. 4).

Input: Collected data of CM, the set ***A***, of applications identifications to be attended; ***FTU.***

Output: The fused data for each application.

```
//Step 1: Receive the collected data
1. Store in the set L the data received through the message SENSED_DATA_SAMPLES.
//Step 2: Data fusion
2. Executes the function FUSION (L, FTU) in order to select fusion technique and to fusion the data
3. Send APPS_FUSED_DATA message to the Sink Node containing the data fusion results.
```

Fig 4. Pseudocode of the Steady-State on CH view

IV. EXPERIMENTS

This section describes the experiments conducted with S-LEACH in SSN scenarios for evaluating the impact of the algorithm in the WSN, compared to LEACH approach, in terms of the network lifetime and the required amount of memory.

A. Experimental Settings

The experiments were conducted in the SUN SPOT platform [14], a sensor platform particularly suitable for rapid prototyping of WSNs applications. The SUN SPOT SDK environment includes Solarium, that is a tool to manage SPOTS that contains a SPOT emulator that is useful for experimenting SPOT software and/or to create scenarios with a large number of nodes whenever the real hardware is not available. The proposed algorithm was deployed on the SUN SPOT platform rev8 hardware [15] (1 Mb RAM memory sized, 8 Mb of Flash memory and AT91SAM9G20 with a master clock speed of 133.3248MHz).

In our experiments, we have used up to 10 applications (1, 2, 3, 5, and 10 applications). For each application, we assigned two randomly sensing units. Our implementation considered 1 – 5 different sensing units (accelerometers, temperature, light, humidity and presence) [10]. For each assigned sensing unit, we randomly assigned sensing rates varying from 1 to 5 seconds. The sensing units used in our applications represent the SUN SPOT embedded sensors.

All experiments were performed in a 100m x 100m field. The network nodes are in the Cartesian plane defined in the area {(0,0), (100,0), (0,100), (100,100)}. The sink is located far from any sensor node, at coordinates (200,100). All network nodes starts with 0.5 joules as initial energy within its batteries. We have randomly distributed 51 nodes in the network (50 nodes and 1 sink node). LEACH and S-LEACH are implemented using the message sizes in bits (Table I).

To simulate the message energy consumption in our experiments, the radio model used in the simulation is the same model as discussed in [3] which is the **first order radio**

model. Sending and Receiving messages are costly operations. Therefore, the usage of these operations should be minimal. Also it is assumed that the radio channel is symmetric so that the energy required to transmit a message from node *i* to node *j* is the same as energy required to transmit a message from node *j* to node *i*.

TABLE I. MESSAGES SIZES IN BITS

	S-LEACH	LEACH
CH_ROLE_BROADCAST	16	16
CM_JOIN	16	16
TDMA_SCHEDULE	80	80
SENSED_DATA_SAMPLES	176	80
APPS_FUSED_DATA	336	144
APP_DEPLOYMENT	26	26
FUSION_DEPLOYMENT	16	16

B. Metrics

The metrics used for assessing the impact of S-LEACH in the WSN are: (i) the lifetime of the network and (ii) the memory consumption. In this paper, we adopted the same definition of network lifetime used in [11], which is the time elapsed until the first node in the WSN is completely depleted of its energy. The memory consumption is defined as the amount of memory used by the implementation of S-LEACH installed in the sensors nodes (RAM and ROM).

C. Evaluating the impact of S-LEACH in the WSN

The main goal of the first set of experiments is to assess how long a SSN lasts using S-LEACH and LEACH algorithms by varying the number of applications (1, 2, 3, 5 and 10) simultaneously running in the network. Table II shows the network lifetime using S-LEACH and LEACH and the gains of S-LEACH, in terms of the lifetime, compared to LEACH for scenarios with 1,2,3,5 and 10 simultaneously running applications. The results of this experiment (TABLE II) shows that with the increment of the number of applications simultaneously running in the SSN, in both algorithms the network lifetime values are reduced. It is possible to observe in TABLE II that the network lifetime values achieved by S-LEACH were 28%, 101% and 340% higher than the network lifetime values achieved by LEACH for scenarios 3, 5, 10 applications, respectively.

TABLE II. USEFUL LIFETIME GAINED BY S-LEACH

	LEACH	SLEACH	GAIN
1 Application	22.45 h	9.5 h	-58%
2 Applications	10.45 h	9.57 h	-8%
3 Applications	7.63 h	9.75 h	+28%
5 Applications	4.85 h	9.75 h	+101%
10 Applications	2.25 h	9.90 h	+340%

As more applications are simultaneously running in SSN, there is naturally an increase in the possibility of finding common sensing units among them. S-LEACH algorithms well utilizes this idea to reduce energy consumption of nodes. Beside that, our implementation of S-LEACH also uses the *Enhanced Moving Average Filter* [13] data fusion algorithm designed for SSNs scenario in order to reduce the number of transmission made by the CH to the BS, further extending network lifetime, since instead of transmitting the same data several times (one for each application), as LEACH would do, S-LEACH transmits data only once for the several applications.

On the other hand, we can observe in TABLE II that LEACH presents better result (58%) in terms of lifetime than S-LEACH for scenarios with a single application. This happens because as LEACH was designed to attend a single application and S-LEACH was designed to meet the requirements of multiple applications, S-LEACH computation procedures are more complex than LEACH's. For the scenario of two applications, LEACH also presents better result (8%) than S-LEACH. This happens due to the communication overhead imposed by S-LEACH. In short, S-LEACH presents better network lifetime than LEACH for scenarios with more than two applications running simultaneously in the SSN.

Considering ***the memory consumption*** in bytes for the sensor node, we noticed that the memory consumption of S-LEACH (29049 bytes (28.4 Kb)) was 37.8% higher than LEACH (21088 bytes (20.6 kb)). Although the memory consumption of S-LEACH presented an overhead in relation to LEACH, S-LEACH extends network lifetime of in relation to LEACH. It can be seen that the S-LEACH consumed a moderate amount of memory, because there are still 71.6, % of available RAM space. Additionally the external flash memory is completely available for the application.

D. Comparison between simulated and real nodes

In this section, the same scenario simulated using Solarium was implemented on a real sensor node platform loacted in a controlled environment (our research laboratory at UFRJ). We aimed to validate the results obtained from simulations by comparing them with the results obtained from a real WSN platform. In this case, the nodes were kept stationary and disposed on the floor.

The experiment on simulated nodes consumed less energy than the real experiment, since there was no interference on the simulated experiments. On average, the obtained network lifetime value for real experiments with 3 applications was 8.9 hours, with standard deviation of 0.30h, while in the simulated environment the obtained network lifetime average value was 9.75 hours, with standard deviation of 0.20 hours.

E. Discussion about of the accuracy of S-LEACH

In this section, we discuss that S-LEACH saves energy, preserves the data semantics as assures and enhances the data accuracy in SSNs making use of enhanced fusion techniques instead of traditional fusion techniques,. Traditionally fusion techniques were all designed for an application-specific network. This means that traditional fusion techniques process all the sensed data under the specific data semantics of a single target application. Recently, in works of [12] and [13] these traditional fusion techniques were adapted to the SSN scenarios in order to consider the distinct data semantics of each application. By data semantic we mean a pattern that describes an application and enables interoperability and integration between applications [13]. The issue of dealing with distinct data semantics of each application is particularly important and common in a distributed fusion system [13]; hereinafter we call this issue as application correlation in our work. If semantic differences were not taken into account, the fusion methods would produce unreliable results for different applications. Thus, by taking the semantics into account, it is possible to enhance the fusion techniques's accuracy. Besides the fact that the traditional fusion techniques process the sensed data under the specific data semantics of a single target application, they also assume that all the sensed data are weighted and handled equally and have the same data range. In SSN, fusion techniques need to consider that the same sensed data may have different degrees of importance for different applications and also different data ranges. Considering the aforementioned discussion, the authors in [12] and [13] argued that the existing traditional fusion techniques are not suitable to be used in SSN scenarios, since they were not conceived taking into account the SSN specific features. Therefore, there is a need for fusion techniques to deal with these features in order to achieve energy efficiency and reliable results in this emergent scenario. The authors in [12] and [13] proposed fusion techniques suited for SSNs (that we have used in the presented work) that are able to preserve the data semantics, to assure and enhance the data accuracy in SSNs since they combine information from multiple sensors, sources and applications to achieve inferences that are not feasible from a single sensor or source through probabilistic methods. The authors concluded that the data range and the weights assigned to applications (representing the relative priority assigned of each application) are extremely important in the fusion process. Since applications have different degrees of importance, it is intuitive to assume that their data have different degrees of importance. If an application less relevant but with a large data range has the same degree of importance of the other existing applications, it will lead to an error, i.e., a result which does not mirror the current situation of the environment. If we consider the data semantics, and weight it according to its importance, the fusion technique will return a result closer to reality. Therefore, we can say that S-LEACH making use the fusion techniques suited for SSN can guarantee and enhance the data accuracy.

V. Conclusion

In this paper, we have presented an algorithm tailored for SSN, called S-LEACH (Shared LEACH). S-LEACH is an extension of LEACH algorithm that allows the formation of clusters to serve multiple applications efficiently. The results of our experiments shows that S-LEACH increased the network lifetime of the experimented scenarios and does not have high memory consumption. As future work we direct the research and development of different WSN cluster-based routing protocols in order to be able to cluster nodes based on the applications requirements instead of a geographical position. In S-LEACH we have extended LEACH, that performs the clustering scheme based on geographical position of the nodes. We advocate that base the clustering in the application requirements better suits cluster-based routing algorithms for SSNs. Another direction for future work consists on designing a mechanism capable to decide which fusion technique should be applied, given a set of application running in the network. Through some strategies, such as computational intelligence, finding the most appropriate fusion technique is useful in environments where applications change quickly, such as the SSN scenario.

References

[1] Claudio M. de Farias, Luci Pirmez, Flavia C. Delicato, Wei Li, Albert Y. Zomaya, Jose N de Souza, "A scheduling algorithm for shared sensor and actuator networks," The International Conference on Information Networking 2013 (ICOIN2013), pp. 648-653, 2013 International Conference on Information Networking (ICOIN), 2013

[2] Yan Zhang, Laurence T. Yang, Jiming Chen; "RFID and Sensor Networks: Architectures, Protocols, Security, and Integrations"; Chapter 12 - Clustering in Wireless Sensor Networks; p334 ; ISBN 9781420077773; November 4, 2009 by CRC Press

[3] Heinzelman, W.R.; Chandrakasan, A.; Balakrishnan, H., "Energy-efficient communication protocol for wireless microsensor networks," System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on , vol., no., pp.10 pp. vol.2,, 4-7 Jan. 2000

[4] Hong, Jiman; Kook, Joongjin; Lee, Sangjun; Kwon, Dongseop; Yi, Sangho; "T-LEACH: The method of threshold-based cluster head replacement for wireless sensor networks"; Journal Article, Information Systems Frontiers, Springer US, 513-521, 1387-3326; 2009

[5] Loscri, V.; Morabito, G.; Marano, S., "A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH)," Vehicular Technology Conference, 2005. VTC-2005-Fall. 2005 IEEE 62nd , vol.3, no., pp.1809,1813, 25-28 Sept., 2005.

[6] Lindsey, S.; Raghavendra, C.S., "PEGASIS: Power-efficient gathering in sensor information systems," Aerospace Conference Proceedings, 2002. IEEE , vol.3, no., pp.3-1125,3-1130 vol.3, 2002

[7] AA Abbasi, M Younis; A survey on clustering algorithms for wireless sensor networks; Computer communications, 2007

[8] Bispo, K. a., Rosa, N. S., & Cunha, P. R. F. (2012). A semantic solution for saving energy in wireless sensor networks. 2012 IEEE Symposium on Computers and Communications(ISCC), 000492–000499. doi:10.1109/ISCC.2012.6249344

[9] Leontiadis, I., Efstratiou, C., Mascolo, C. and Crowcrof, Jt. 2012. SenShare: transforming sensor networks into multi-application sensing infrastructures. In Proceedings of the 9th European conference on Wireless Sensor Networks (EWSN'12), Gian Pietro Picco and Wendi Heinzelman (Eds.). Springer-Verlag, Berlin, Heidelberg, 65-81.

[10] S. Xiong, J. Li, M. Li, J. Wang, and Y. Liu, "Multiple Task Scheduling for Low-Duty-Cycled Wireless Sensor Networks," in INFOCOM '11.

[11] RAGHUNATHAN, V. et al. Energy-aware wireless microsensor networks. IEEE Signal Processing Magazine, v. 19, n. 2, p. 40–50, 2002.

[12] Farias, C.; Pirmez, L.; Delicato, F.; Carmo, L.; Wei Li; Zomaya, A.Y.; De Souza, J.N., "Multisensor data fusion in Shared Sensor and Actuator Networks," Information Fusion (FUSION), 2014 17th International Conference on , vol., no., pp.1,8, 7-10 July 2014

[13] de Farias, C. M., Pirmez, L., Delicato, F. C., Dos Santos, I. L. and Zomaya, A. Y.. 2012. Information fusion techniques applied to Shared Sensor and Actuator Networks. In Proceedings of the 2012 IEEE 37th Conference on Local Computer Networks (LCN 2012) (LCN '12). IEEE Computer Society, Washington, DC, USA, 188-191.

[14] Wilde, E.; Guinard, D.; e Trifa, V. Architecting a Mashable Open World Wide Web of Things, Institute for Pervasive Computing, ETH Zürich, Zürich, Switzerland, No. 663, February 2010

[15] Release 6.0 of the Sun SPOT SDK; Sun SPOT SDK Release Notes; Sun™ SPOT Programmer's Manual Release v6.0 (Yellow):New Rev 8 hardware : http://www.sunspotworld.com/docs/Yellow/ReleaseNotes.html;http://www.sunspotworld.com/docs/Yellow/SunSPOT-Programmers-Manual.pdf

A Software Update Method for Noisy Wireless Environment of WSNs

Hyeongrak Park, Hyeyeong Jeong, and Byoungchul Ahn
Dept. of Computer Engineering, Yeungnam University, Gyungsan, Korea

Abstract - *Wireless Sensor Networks have been applied to many places such as home health care, assisted living, environmental survey and so on. Nowadays, sensor nodes can perform many functions at the same time and include complex monitoring software. This paper presents an efficient software update in noisy wireless environment. After relay nodes are selected, two update methods are compared for energy consumption, update time and packet loss. Software update methods are measured in noisy environment using NS-2 simulators. In ideal environment, there are few difference in upgrade time, energy consumption and retransmission data sizes. At noisy environment, energy saving, upgrade time and retransmission data sizes are improved by dividing one file into several files. When a file is divided 3 small files, the total upgrade time reduced to 59% in the noisy environment. When a file is divided 4 small files, energy consumptions of smaller file transmission is reduced up to 74% compare with that of one file transmission. For severe noisy environment, it is much better saving of energy by dividing a file into smaller files.*

Keywords: Noise, Software update, relay, file division

1 Introduction

As semiconductor technology is advanced, sensor nodes of WSNs(Wireless Sensor Networks) are designed with small size, low power consumption and several sensors realized by one chip. Typically WSNs are composed of a lot of sensor nodes, which are deployed to monitor interest area. The sensor nodes are consisted of sensors, data processing, and communication parts. WSNs are applied to various areas such as home health care, assisted living, environmental survey and so on.

Since the computing power of sensor nodes has been increased, sensor nodes have enough power to perform several functions at the same time. Their software is programmed to operate several functions and its complexity is increased. As battery technology is advanced, the life time of sensor nodes is dramatically improved. And the deployment of solar cells makes sensor nodes operate permanently when they deployed in the fields. This means that their software of sensor nodes might be reprogrammed or added new functions. But it is very difficult to access a node one by one and update its software manually.

Most researches for software update are concentrated to methods of software update in ideal environment without noise. This paper presents an efficient software update method in noisy wireless environment. After relay nodes are selected, two update methods are compared for energy consumption, update time and packet loss.

2 Related work

There are some researches about software updates for sensor nodes. But they are focused on system management, not an update itself. Han *et al*. have presented a survey for software update. Energy efficient software update methods are described by comparing the other methods[7]. For efficient data transmissions, Intanagonwiwat *et al*. suggests direct-diffusion method[9]. But this research is focused on data aggregation and data transmission path. So it is not suitable for software updates since the data flow direction is reversed. For lower power consumption, Wei Ye *et al*. suggests S-MAC. S-MAC is an MAC level protocol using sleep cycle and clustering[10]. This protocol use periodic sleep and data bandwidth is very low. It can’t be applied for software data transfer since the data size of control software is much larger than that of the normal data.

Since control software contains execution code for a processor of sensor nodes, it is very important to maintain reliable data transfer. A method for reliable data transfer in WSNs is developed for 1:1 communication such as S-TCP and RMTS[3][4]. But 1:1 communication methods are inefficient to update many nodes for WSNs. If this method is used for re-programing sensor nodes of WSNs, each node must be updated first and retransmit the software update data to another node one by one. Therefore it is necessary to develop an efficient update method for sensor nodes with fast update time and small data retransmission.

Stephen *et al*. suggest a update software model for WSNs[8]. This model presents the theoretical approaches to update software. They have not provided simulation results or experiments to verify their model. They validate their model against three different systems, representing three classes of software update: static/monolithic updating(MOAP), dynamic/mobile agent-based updating (Mate) and dynamic/component-based updating(Impala).

For the update protocol, Kulkarnia and Arumugama have implemented a selective retransmission Go-Back-N scheme using TDMA protocol to provide reliable data transfer for WSNs. In order to save energy, TDMA scheme is very useful. Main feature of Infuse method is continuously sends data to the next node from a predecessor node[1]

3 Software update models

In this paper, all sensor nodes of WSNs are assumed to use the same hardware configuration such as a processor, its memory size and so on. It means that all sensor nodes use the same software version and their locations are static and distributed uniformly. Some assumptions for software update model described are given as follows:

① Wireless Sensor Network uses CSMA/CA based mesh structure ad-hoc network

② There is only one base station in the sensor network

③ All nodes use the same software

④ All nodes are fixed at one location

3.1 Software update time

The time to update the software of each node indicates the sum of the data transmission time, failure recovery time and reprogramming time for a node. T_{data} is data transmission represented by Equation(1).

$$T_{data} = \left(\frac{P_d + P_h}{bit\_rate} P_o\right) \times C \tag{1}$$

where, P_d : Data packet size
P_h : Packet header size
P_o : Wireless channel access time.

The whole error recovery time(T_{err}) is represented by Equation (2).

$$T_{err} = \left(\frac{P_d + P_h}{bit\_rate} P_o\right) \times C_f \times P_{err} + \left(\frac{P_c + P_h}{bit\_rate} P_o\right) \times C_f \times P_{err} \tag{2}$$

where, P_c : Control packet size
P_{err} : Packet error rate.

The update time for a node at the single hop(T_{step}) is formulized by Equation (3).

$$T_{step} = T_{data} + T_{err} + T_{up} \tag{3}$$

where, T_{up} : Update time after data reception
T_{data} : Transmission data time
T_{err} : Transmission error time.

Total update time(T) for all nodes in the sensor field is expressed by multiplying the number of update step and the time of update time of single hop. The update step is the distance between a base station to its farthest node divided by the radal transmission range of nodes. The update time for whole sensor node in the WSNs are expressed as Equation(4)

$$T = T_{step} \times \frac{l}{r} \tag{4}$$

where, l is distance from base node to its farthest node
r is radio radius of a node.

3.2 Energy model

The power consumption of relay nodes is calculated by the sum of receiving data and transmitting data, and is calculated as Equation (5).

$$\begin{aligned} J_{nr} &= J_r \ file\_size(1+e) \\ J_{ns} &= J_s \ file\_size(1+e) + J_{nr} \end{aligned} \tag{5}$$

where, J_r : the energy to receive data
J_s : the energy for transmit data
J_{nr} : the energy consumed by receiving nodes
J_{ns} : the energy consumed by relaying nodes
e : transmission error rate
$file\_size$: the size of software upgrade.

In Equation (5), J_{nr} and J_{ns} are the power consumption of receiving and transmitting data to other nodes. Some nodes located in duplicated radio area are received a few multiple times of data size of software. Therefore, the total energy consumption of all nodes is J in Equation (6).

$$J = (J_s + J_r \frac{N_t}{Area\_of\_field} \ \pi \ r^2 + e \ J_s) file\_size \ N \tag{6}$$

where, N_t : total number of nodes in a field
N : a number of relaying nodes.

In Equation (6), all parameters are fixed except e and N. It is very important to reduce transmission errors and the number of relaying nodes.

3.3 Upgrade algorithms for noisy wireless environment

In the paper, the following two methods are considered to figure out the noisy environmental performance.

① CHR(Cluster Head Relay) : This algorithm is proposed by Jeong[20]. This method uses cluster heads to update software. After cluster heads update software, they transmit update software to next cluster heads which are not upgraded.

② PFR(Partial File Relay) : This method uses the same algorithm of CHR. It divides one upgrade file into two, three and four small files and transmits them to next cluster heads. To find their performance and upgrade time, the basic transmission and reception method of the CHR upgrade algorithm is used.

3.3.1 CHR algorithm

In WSNs, all nodes must belong to one of clusters. Collected data are transmitted to the base station by cluster heads. When data is transmitted to the base station, cluster header information is added. In Figure 1, cluster head nodes, which are 5, 7, 9, 12 and 17 in clusters, send collected data to the base station. The base station searches node IDs of cluster heads among received data and selects these nodes as relay nodes. Before updating software, nodes to be updated will receive event data and broadcast them to their neighbor nodes. CHR algorithm is shown as Figure 2 and Figure 3. After receiving "*Relay-Start*" message, nodes prepare software update procedure.

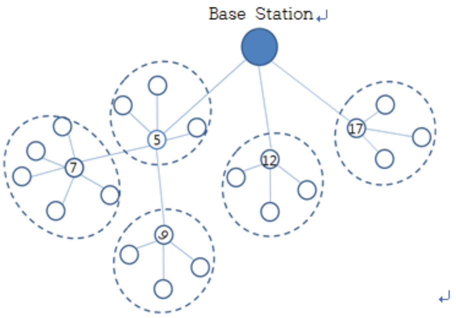


Figure 1. CHR method

Relay nodes broadcast "*Data-Start*" message and new software version to their neighbor nodes, and inform that software update transmission will begin. After software update is finished, updated nodes send "*Reprogram-Done*" message to their relay nodes.

Algorithm 1: CHR Transmission Node

```
Broadcast status information periodically
If (receive "Relay-Start" && exist proper neighbor node) {
    Send "Data-Start" to neighbor nodes.
    Send data
}
While(No. data receive nodes > 0) {
    If(receive "Reprogram-Done") {
         Select one node.
         Send "Relay-Start" to selected node.
         Decrease No. of data receive node.
    }
}
```

Figure 2. CHR Transmission Algorithm

Algorithm 2 : CHR Reception Node

```
Broadcast status information periodically
If (receive "Data-Start") {
    If (receive software ver. > stored software ver.) {
         Receive data and store it to memory
         Request missing or lost packet
         Reprogramming it-self.
         Send "Reprogram-Done" to data send node.
    }
}
If ( receive "Relay-Start") {
    Go to Transmission program
}
```

Figure 3. CHR Reception Algorithm

3.3.2 PFR algorithm

Before the software is updated, it operates as a normal sensor node by capturing data and transmitting them to its cluster node. When software needs to be updated, the server prepare updates file. The server divides update software into several small files to perform software update for nodes in the sensor networks. When the server sends out "*Relay-Start*" message and "*Data-Start*" message, relay nodes starts the update procedure. After nodes receive all packets for software update, they transmit "*Reprogram-Done*" message to relay nodes. The algorithms are shown in Figure 4 and Figure 5.

Algorithm 3: Data Transmission Node

```
Broadcast status information periodically
If (receive "Relay-Start" && exist proper neighbor node) {
     Send "Data-Start" to neighbor nodes.
     Send data[subfile_counter]
     for ( I=1; i<subfile_ counter)
     Send data[subfile_counter]
}
While(No. data receive nodes > 0) {
     If(receive "Reprogram-Done") {
          Select one node.
          Send "Relay-Start" to selected node.
          Decrease No. of data receive node.
     }
}
```

Figure 4. Sub-file Transmission Algorithm

Algorithm 4 : Data Reception Node

```
Broadcast status information periodically
If (receive “Data-Start”) {
    If (receive software ver. > stored software ver.) {
        Receive data and store it to memory
        Request missing or lost packet
        Reprogramming it-self.
        Send “Reprogram-Done” to data the send node.
    }
}
If ( receive “Relay-Start”) {
    Go to Transmission program
}
```

Figure 5. Sub-file Reception Algorithm

4 Simulations for performance analysis

4.1 Simulation environment

To find software update performance at noisy environment, NS-2 network simulator is used. The simulation environment is shown at Table 1. The total number of nodes is 100 sensor nodes and their locations are static and distributed uniformly within 400mx400m rectangular area. There is one base station to start software update. CSMA/CA 802.11 and 802.15.4 wireless standards are used for simulation. With simulations, data for energy consumption, update time and packet loss rate are collected.

Table 1. Simulation parameters

Parameters	Value
MAC protocol	CSMA/CA, Back-off Window 2~26 Slot Time=0.384ms, IFS=1.664ms
RF transmission radius	60m
Wireless bandwidth	250Kbps
Number of node	100
Transceiver power consumption	75.9mW(TX)/62.7mW(RX)
Distance inter-node	40m
Data packet size	128Byte
Data header size	8Byte
Software data size	128KByte
Data transmission rate	Wireless bandwidth 70%
Update time	1.5 Sec
Bit error rate(BER)	0 ~ 1.1x10-3

4.2 Simulation results and analysis

Each node is located uniformly and the node-to-node distance is 40m. Simulations are measured for energy consumptions, update time and packet loss rate by changing bit error rate. Upgrade software size is 128KB and the basic packet size 128B. This means that about 1000 packets are transmitted to other nodes. Figure 6 shows the energy consumptions. In Figure 6, Div-2, Div-3, and Div-4 mean that the upgrade file is divided into two, three and four respectively. The total software update time is shown in Figure 6. When one file is divided into several files, the total upgrade time is reduced dramatically. This means that errors during transmission are reduced and the number of retransmission is reduced. When a file is divided 3 small files, the total upgrade time reduced to 59% in the noisy environment. In the ideal environment, the upgrade time is very similar to each other. As noise is increased and the bit error rate is increased, the reduction rate of software upgrade time is proportional to the number of division.

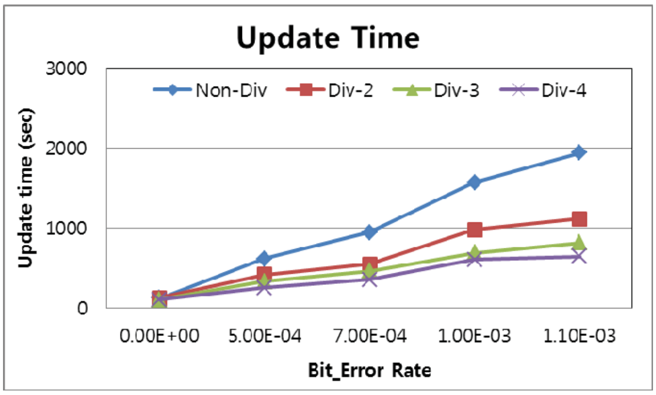


Figure 6. Software Update Time

The total energy consumption is shown in Figure 7. When a file is divided 4 small files, energy consumptions of smaller file transmission is reduced up to 74% compare with that of one file transmission. For severe noisy environment, it is very effective to divide a file into smaller files to save energy.

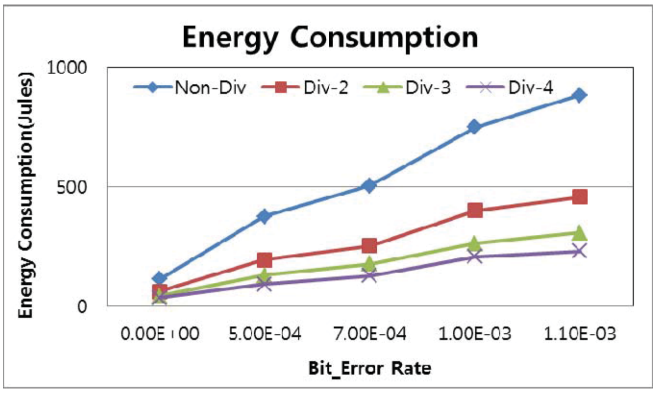


Figure 7. Energy Consumption

The total sizes of data retransmission are shown in Figure 8. When packets are lost or packets have errors, data should be retransmitted. When a file is divided into several files, the total data sizes of retransmission are reduced dramatically. When a file is divided 2 small files, the total data size of retransmission is the half of one file transmission.

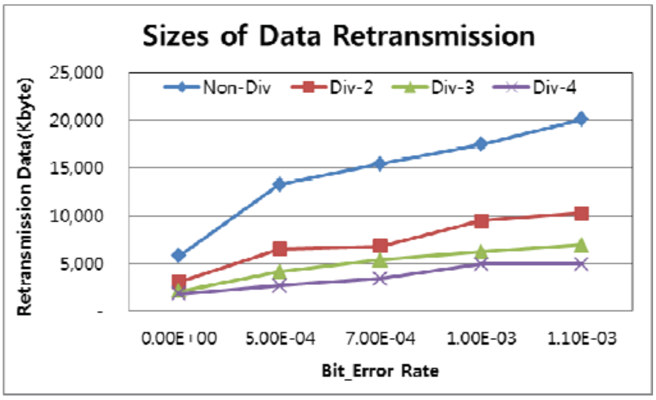


Figure 8. Data Retransmission

The data loss rate is shown in Figure 9. Div-2 shows more data loss than others. This data does correspond to Figure 6, Figure 7 and Figure 8. This means that bit errors of packets are dominant to data retransmissions. Since sizes of loss data are small, the recovery time of the loss data are small in noisy environment

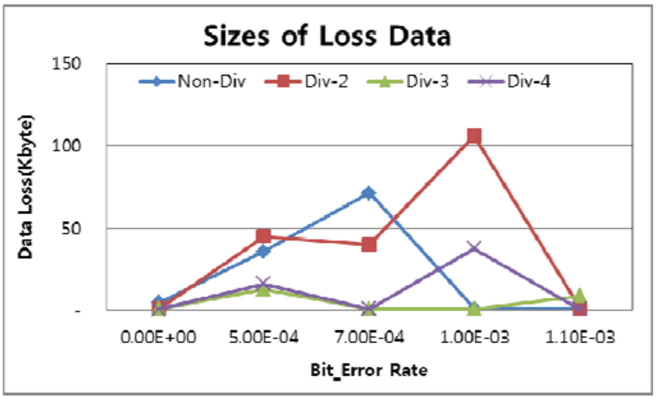


Figure 9. Data Loss

5 Conclusion

In this paper, software update methods are measured in noisy environment using NS-2 simulators. The bit error rates are considered 0, $5x10^{-4}$, $7x10^{-4}$, $1.0x10^{-3}$ and $1.1x10^{-3}$. In ideal environment, there are few difference in upgrade time, energy consumption and retransmission data sizes. In noisy environment, energy saving, upgrade time and retransmission data sizes are improved by dividing one file into several files.

When a file is divided 3 small files, the total upgrade time reduced to 59% in the noisy environment. When a file is divided 4 small files, energy consumptions of smaller file transmission is reduced up to 74% compare with that of one file transmission. At severe noisy environment, it is very effective to divide a file into smaller files to save energy.

Please address any questions related to this paper to Byoungchul Ahn by Email (b.ahn@yu.ac.kr).

6 References

[1] S. Kulkarnia, and M. Arumugama, “Infuse: A TDMA Based Data Dissemination Protocol for Sensor Networks”, Technical Report MSU-CSE-04-46, Dept. of Computer Science and Engineering, Michigan State University, 2004.

[2] T. Stathopoulos , J Heidemann , D Estrin, “A Remote Code Update Mechanism for Wireless Sensor Networks”, CENS Tech. Report #30, Centre for Embedded Networked Sensing, UCLA, 2003.

[3] Y. G. Iyer, S. Gandham, and S. Venkatesan, “STCP: A Generic Transport Layer Protocol for Wireless Sensor Networks”, Proc. of 14th International Conference on Computer Communications and Networks, pp.449-454, 2005.

[4] F. Stann, and J. Heidemann, “RMST: Reliable data transport in sensor networks”, Proc. of the First IEEE. 2003 IEEE International Workshop on Sensor Network Protocols and Applications, pp. 102 -112, 2003.

[5] W. Chen, P. Chen, W. Lee, and C. Huang, “Design and Implementation of a Real Time Video Surveillance System with Wireless Sensor Networks”, Proc. of Vehicular Technology Conference, 2008, pp.218-222. 2008.

[6] H. Wang , D. Peng , W. Wang , H. Sharif , H. Chen , “Image transmissions with security enhancement based on region and path diversity in wireless sensor networks”, IEEE Transactions on Wireless Communications, vol. 8, no. 2, pp.757-765, 2009.

[7] S. Brown and C. Screenan, “Software Update Recovery for Wireless Sensor Networks”, Proc. of International Conference on Sensor Networks Applications, Experimentation and Logistics (SENSAPPEAL) ICST 2009.

[8] C-C. Han, R. Kumar, R. Shea, M. Srivastavam, “Sensor Network Software Update Management: a Survey”, Intl. Journal of Network Management, no. 15, No. 4, John Wiley & Sons, pp. 283-294, 2005.

[9] C. Intanagonwiwat, R. Govindan, and D. Estrin “Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks”, Proc. of the 6th annual international conference on Mobile computing and networking(Mobicom '00), pp.56-67, 2000.

[10] Wei Ye, J. Heidemann, and D. Estrin, “Sensor-MAC (S-MAC): Medium Access Control for Wireless Sensor Networks”, Proc. of the 21st International Annual Joint

Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), vol.3, pp.1567-1576, 2002.

[11] M. L. Chiang, T. L. Lu, "Two-Stage Diff: An Efficient Dynamic Software Update Mechanism for Wireless Sensor Networks", Department of Information Management National Chi-Nan University, in IFIP Ninth International Conference, 2011.

[12] Y. Kwon, Ph.D. Dissertation, Yeungnam University, 2011.

[13] H. Noghabi 1, A. Ghazi askar 1, A. Boustani2, A. Moghani1, M. Zanjani, "Implementing a Greedy Chaun Routing Technique with Spread Spectrum on Gridbased WSNS", International Journal of Wireless & Mobile Networks (IJWMN) Vol. 4, No. 4, 2012.

[14] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, "Energy-Efficient Communication Protocol forWireless Microsensor Networks", IEEE proc, Hawaii International Conf. Sys. pp.1-10. 2000.

[15] J. Kulik, W.R. Heinzelman, H. Balakrishnan: Negotiation–Based Protocols for Disseminating Information in Wireless Sensor Networks. In: Wireless Networks, Vol 8, pp. 169-185, 2002.

[16] J.N. Al-Karaki, A.E. Kamal: Routing Techniques in Wireless Sensor Networks: A Survey. In: IEEE Wireless Communications, Vol. 11, no. 6, pp. 6-28, 2004.

[17] Zeenat Rehena, Sarbani Roy, Nandini Mukherjee : A Modified SPIN for Wireless Sensor Networks. In: IEEE, 2011.

[18] W. Dong, X. Liu, C. Chen, Y. He, G. Chen, Y. Liu, J. Bu, "DPLC:Dynamic Packet Length Control in Wireless Sensor Networks", in Proc.of IEEE INFOCOM, 2010.

[19] Y. Kwon, B. Ahn, "A Firmware Update Model for Wireless Sensor Networks", in The Korean Institute of Informations Scientists and Engineers, 2011.

[20] H. Jeong, B. Ahn. "Software UpdateMethod Using Clustering WSNs", in Proc of ICWN 14, pp.29-34, 2014.

A Framework for Improving the Performance of IoT Applications

Jose Sobral[1*], **Ricardo Rabelo**[1], **Douglas Oliveira**[2], **Jose Lima**[1], **Harilton Araujo**[3] **and Raimir Holanda**[4]
jose.v.sobral@ieee.org, ricardoalr@ufpi.edu.br,
doliveira2011@my.fit.edu, jcarloslimafilho@hotmail.com,
harilton.araujo@estacio.br, raimir@unifor.br
[1]Federal University of Piaui, Teresina, Piaui, 64049-550, Brazil
[2]Florida Institute of Technology, Melbourne, Florida, 32901, USA
[3]Estacio CEUT, Teresina, Piaui, 64046-700, Brazil
[4]University of Fortaleza, Fortaleza, Ceara, 60811-905, Brazil
[*] Corresponding author.

Abstract—*The Internet of Things is a paradigm that proposes the interconnection of objects and things collaborating for a common objective. Several works have used the integration of WSN and RFID for enable the data exchange between this objects in IoT applications. However, with this integration, the challenges of these technologies are expanded and new problems can emerge. This work proposes a framework to reduce the problems that arises from the integration of WSN and RFID into IoT applications. The proposed framework is formed by two components. The first is a RFID anti-collision protocol and the second is a mechanism to assist WSN routing protocols. The framework was compared with state-of-art protocols, and the results showed that, with the utilization of the proposed framework is possible enhance the network performance by providing a best structure for the IoT applications.*

Keywords: Internet of Things, Wireless Sensor Networks, RFID, Anti-collision, Routing.

1. Introduction

Internet of Things (IoT) is a multi-disciplinary domain covering a wide array of topics that go from purely technical (routing, requisition semantics) to a mix of technical and social issues (security, privacy, usability), as well as social and entrepreneurial subjects [1]. Applications of IoT, the existing and potential ones, are equally diverse: environment monitoring and personal health, track and control of industrial processes including agriculture, intelligent spaces, and intelligent cities are only a few examples of Internet of Things' applications. [2]

In particular, IoT should expand from technologies such as RFID (Radio Frequency Identification), which, for having object identification and location functionality, can be used in a great number of applications. At the same time, IoT may employ other relevant technologies, such as Wireless Sensor Networks (WSN), which make it possible to collect information about the environment in which they are inserted. [3]

In order for IoT to perform efficiently their application, it is necessary that the used network structure provides some key system level resources, such as devices heterogeneity, scalability, ubiquitous data exchange through proximity wireless technologies, energy-optimized solutions, localization and tracking capabilities, self-organization capabilities, semantic interoperability and data management, embedded security and privacy-preserving mechanisms. [4]

The most promising technologies for the IoT paradigm are RFID and WSN [1] [3]. Due to their specific limitations, these technologies are not able to individually support all mentioned required resources. As an alternative to this problem, the WSN and RFID integration can implement a structure capable of providing such resources. When developing this joint structure, it should be taken into consideration the specific issues of each technology, as well as problems derived from the integration, with the goal of supporting IoT applications.

Several works show solutions that integrate WSN and RFID for IoT applications [5] [6] [7]. However, these works do not consider the problems related to the integration of these technologies. Not considering these WSN and RFID integrate challenge for IoT network can degrade the applications performance.

Thus, this work proposes a framework able to improve the network performance that integrates WSN and RFID for IoT applications looking for the problems arising with this integration, well as the problems already existing in these technologies. The framework is composed by two components. The first component has the objective of improve the reading tags system, making it most fast and efficient. The second component aims enhancing the performance at message exchange between the nodes with sensing and reading capacities, reducing the packet loss rate and the energy consumption. With the utilization of the proposed Framework are expected:

- Improvement of RFID reading system performance;
- Reducing packet loss rate.
- Reducing the network nodes energy consumption;

The work is organized as follows: Section 2 present a short overview about the integration between WSN and RFID technologies. The section 3 describes the proposed framework. The results evaluation is discussed in Section 4, followed by conclusions in Section 5.

2. Integration of WSN and RFID for Internet of Things Applications

The Internet of Things (IoT) is the pervasive presence of a variety of things and objects (such as RFID tags, sensors, actuators, smartphones, and so on) around us that are capable of interacting among themselves cooperating towards the achievement of a common goal [3]. These objects and things are commonly exchanging information over the Internet.

According to [3], one of the key components to IoT are RFID systems, which are composed by one or more readers and many tags. Into RFID system, a reader sends commands and search for tags that are attached to objects or people. The tags answer the request from the reader with a unique identification number, which will be used in the application. The RFID system works as a real time data acquisition system. To explore its whole potential it is necessary that the acquired information be rapidly processed and forwarded to other systems that may use it. The collisions generated as a result of simultaneous answers from the tags is a critical issue in RFID systems [8]. Hereupon, IoT applications that use RFID systems should take into account collision related circumstances.

The WSNs have a crucial role in IoT scenarios [2], since they allow the development of applications capable of monitoring many variables in their insertion location. Generally, WSN have a great number of distributed sensor nodes, have energy, storage and processing constraints and should have auto configuration mechanisms in case of loss of communication and failure in the sensor nodes. WSNs tend to execute a collaborative function in which the sensors provide data, which is processed (or consumed) by special nodes, called base station or sink nodes. One of the main challenges in WSN is the energy consumption, considering that in most cases, the nodes are located in places of difficult access, which makes it unfeasible to change power supply batteries [9].

The sensors nodes can be enough to collect and transmit data from an environment. Yet, WSNs have limitations at exclusive identification of an object/person in certain types of applications, not representing an optimized solution in terms of efficiency and implementation cost. On the other hand, RFID systems are fast and convenient for object identification and introduces a simpler and cheaper approach for virtual identification and location of objects in the growing IoT paradigm [10].

The majority of IoT applications needs of object identification, in addition to data on the environment in which they are inserted [6]. Since WSN and RFID technologies can’t fill this need individually, it is necessary integrate both technologies to achieve the expected level of data collection (identification and environment data) required by most IoT applications.

According to [11] the integration of RFID and WSN technologies maximizes its benefits, creating new perspectives for a greater number of applications, and approximating the real world to academic researches. This happens because the result of this integration is a technology with extended capacity, scalable and portable with reduced costs.

In [12] four types of WSN and RFID systems integration are discussed. This integration can occur in the following ways: integrating tags with sensors (temperature, humidity, pressure and so on), integrating tags with wireless sensor nodes (tags with sensing and multi-hop communication capabilities), integrating readers with wireless sensor nodes (readers with sensing and multi-hop communication capabilities) and a set of RFID components and sensors integrated through application.

This work considers the way that the RFID readers are integrated to wireless sensor nodes. Thus, aiming to promote the integration between WSN and RFID, it is necessary a computational element capable of collecting data from the environment (temperature, humidity and pressure) and identify RFID tags. The computational element that integrates WSN and RFID are called reader-sensor (RS) node. In this work, RS nodes can have two different architectures, which will be presented next.

2.1 Integration With Software Defined Radios Architecture

The RS nodes with Software Defined Radio (SDR) architecture use just one communication interface to exchange messages switching between WSN and RFID elements. The SDR includes a transmitter in which the operational parameters like frequency, modulation strategy or maximum output potency can be changed using software without the need of changing any hardware component responsible for the radiofrequency [13]. In order to make it easier to understand, the RS nodes that use SDR radio will be called RS-SDR (Reader-Sensor Node with Software Defined Radio). With the use of this type of reader node it is not possible to communicate with sensor nodes and RFID tags at the same time. That means that when a reader node is collecting data about the tags it can not receive data from other sensor or reader nodes of the network.

2.2 Integration With Dual Radio Architecture

The RS node with Dual Radio architecture uses two independent communication interfaces. For instance, first interface can use a CC2420 radio for exchanging messages with the other sensor or reader nodes of the network. The second interface can use a CC1000 radio to communicate with the RFID tags on the objects. The use of two radios allows the reader node to communicate with the sensor nodes and the RFID tags. These nodes will be called RS-DR (Reader Node with Dual-Radio). In contrast to RS-SDR, the RS-DR are capable of performing simultaneous communication with the RFID tags and other sensor or reader nodes of the network. This is possible because the RS-DR utilizes two independents interfaces of communication.

3. Proposed Framework

This work proposes a framework to improve the performance of the network structure that integrates WSN and RFID for IoT applications. The framework proposed is formed by two components. The first component is referent to the way in which the data from RFID elements are collected for the application. In this component is presented a anti-collision protocol for RFID tags based on the EPC Class 1 Generation 2 standard [14]. The second component is responsible for the routing technique, a existent routing protocol is boosted with intelligent systems to improve the perform once.

The following subsections present the components of the proposed framework. In addition, each subsection shows a brief explanation of the problem that the component of framework aims minimize in order to provide the resources required for IoT applications.

3.1 Fuzzy Q-Algorithm

According to [8], the high amount of exchanged packets between readers and RFID tags can generate problems that affect the network scalability and also the quality of service required by the applications. Generally, the quality of service is affected by loss of network packets, which is commonly caused by packet collision. In order to reduce the packet loss caused by collisions, the RFID readers uses anti-collision protocols that coordinate the sending of answer messages of the tags. Some integration proposals of WSN and RFID not consider the collision problems at the moment of the reading of tags. A inefficient reading system can impact the performance of IoT applications. The delay in the reading process caused by the collisions can increase the energy consumption resulting in the premature death of RS nodes.

Thus, the proposed framework presents a component that is responsible for enhance the performance of the tags reading system held by the RS nodes. The Fuzzy Q-Algorithm (FQA) is a anti-collision RFID protocol based on EPCglobal UHF Class-1 Generation-2 (EPC C1G2) protocol and in the Q-Algorithm [14]. At EPC C1G2 the tags identification process consists in a inventory. A inventory is composed of several rounds, which in turn, are composed of several slots. At protocol, each time that a tag need to be identified, all identification process is performed, independent if the tag has already been identified or not. This repetitive message exchange can increase the time of the tags reading process, besides increase the collision chance. In addiction, EPC C1G2 used a Q-Algorithm for define the size of each round. The Q-Algorithm adjusts the slots quantity of each round based in a static parameter. This adjust form can hurt the performance of IoT applications where the quantity of tags at the reader range change quickly.

Considering the necessity of IoT applications, FQA utilizes Fuzzy Systems for define dynamically the parameter used for determinate the quantity of slots in a round. Furthermore, FQA seeks reduce the sending of redundant identification codes (EPC), which can affect the acting of the IoT applications by the increasing the time of the tags identification process.

In the FQA protocol, QUERY commands are used for start a reading process. Upon received a QUERY command the tag select randomly a slot count between 0 and $2^Q - 1$ where Q is a value between 0 and 15 sent inside of the QUERY command. QUERY_REP commands are sent by readers for the tags for decrease their slot count. When a tag has a slot count equals to 0, it sends a RN16 command for the reader. The RN16 contains a aleatory number of 16 bits that not change until the tags receive a special command from reader. Upon receive a RN16, the reader verify if already received the RN16 previously. If true, the reader sends a ACK with the RN16 and a marked flag. If false, the reader send a ACK with the RN16 and a unmarked flag. When the tag receives the ACK with the RN16 previously sent and the unmarked flag, it sends your EPC identification code. If a flag of the received RN16 by the tag is marked, the tag changes its state by ACKNOWLEDGED. When the reader receives the EPC code of the tag, the reader stores the EPC together with the RN16 previously sent by the tag. This stored information is used to verify if a tag was already identified. If most of one tags send a RN16 at same slot, occurs a slot collision and none of the tags can be identified. If none tags send a RN16 in a slot, this slot is empty. If the reader receives the EPC code with success, the slots is successfully. With the purpose of obtains a best adjust of the slots quantity at each round, FQA readers uses a variation of Q-Algorithm. In the FQA, Q_{fp} is a floating representation of Q. After each slot the value of Q_{fp} can be incremented or decremented based on c, where $0.1 < c < 0.5$. If occurs a slot collision, Q_{fp} is incremented in c. If occurs a slot empty, Q_{fp} is decremented in c. If occurs a slots successfully Q_{fp} no change. After it, Q_{fp} is rounded, if the value differs of Q a QUERY_ADJUST command is sent to adjust the Q value of the tags. At the end of each inventory, the value of c is adjusted dynamically by a Fuzzy System. The new c is based in the Q value of the last inventory.

With the utilization of FQA, is expected the efficiency increasing of the tags reading process and at same time, the reduction of this process. With the dynamic adjusts of c is expected to improve the tags reading rates, which are resulting of a appropriated Q value. Avoiding the sending of redundant EPC codes, is expected to reduce the time needed for the tags identification. In this way, is efficient intended to enhance IoT applications with an identification process.

3.2 Fuzzy System-Based Route Classifier

A important challenge related to the WSN seeking IoT applications is referent to the problem of node energy consumption, where the nodes located near of sink node or that compose the most used routes tends to exhaust their energy resources too early [15]. The premature death of the nodes can cause the rupture of the routes previously established, making with the network have to re-organize, causing higher energy consumption and bandwidth. Thereby, the route selection scheme need consider the residual energy of nodes and the quality of communication among them with the objective of reduce the energy consumption while, at same time, provide a load balance and a better distribution of the limited resources of network [16]. A problem that generally is not considered when developing a routing protocol for IoT applications that integrate WSN and RFID is the problem of the quantity of tags near of the RS nodes. The greater the quantity of tags in a node RS area, highest will be the time for read these tags.. With a high reading time, the energy resources of a RS node tends to exhaust quickly, thus, routes that have RS nodes with high tags density at their proximity should be avoided.

Thus, this component of the proposed framework present a mechanism based in Fuzzy Systems able to classify routes and assist routing protocols in scenario of IoT that integrate WSN and RFID, this mechanism is called Fuzzy System-Based Route Classifier (FSBRC). For assist the routing protocol the FSBRC considers information of several layers of the WSN and RFID. The informations used by FSBRC travels in specifics fields inside of control messages of the routing protocol or are collected together to the RFID elements that compose the network. This information are: energy level of route, hops number, LQI and tags density. They are used for the Fuzzy System of the FSBRC for determine the quality of each route.

Although the FSBRC can be utilized together with several multipath based routing protocols, in this work it is used together the Directed Diffusion (DD) routing protocol [17]. By means of their functioning, the DD can store different routes for a destination and use them in case of failure or for send redundant data. In this work, the proposed framework utilize the FSBRC component together of the DD protocol for provide the routing of packets exchange between nodes RS, sensors and sink. The DD protocol was selected for be one of the most utilized in WSN.

DD uses the FSBRC mechanism when receives a interest message. Following is described the step sequence executed by the RS node when receives a interest message:

- Begin: Having received a message at the physical layer, the node calculates the LQI value that indicate the communication quality between the sender and the receiver. LQI value is sent to the network layer together with the received message.
- Step 1: When receive the interest message on the network layer, the node initially verify if the LQI value calculated is highest of that the LQI value contained in the message. If true, the LQI value of the message receives the calculated LQI value.
- Step 2: the node collects the data of energy, hops number, LQI and tags density contained in the received message.
- Step 3: the node uses the values of energy, hops number, LQI and tags density for determinate the route quality using the FSBRC mechanism.
- Step 4: the node verifies if already have a entrance in their cache to the received interest, if true the cache is updated, verifying if is needed create a new gradient or update a already existent. Posteriorly the message is disposed.
- Step 5: if the interest was still not received, the node inserts the interest in the cache together with the informations of route quality and the ID of sender node.
- Step 6: the node verify if their residual energy is less than the contained in the message, if true the node changes the value of energy in the message with their value of residual energy, otherwise the value is not changed.
- Step 7: the node verify if their tags density is greater than the contained in the message, if true the node changes the value of tags density in the message with their value of tags density, otherwise the value is not change.
- Step 8: the hops number is incremented by one, the value of sender ID in the message receives the node ID and the message is re-sent in broadcast.

After the routes formation, whenever a node needing send data messages its shall select the route with highest quality among the available routes seeing the informations contained in their cache. In a fixed time interval the RS nodes exchange control messages for refresh the values of energy level, LQI and tags density. This make with than the network can change the sending path of messages case some route are losing quality. With the utilization of FSBRC mechanism of the proposed framework, it is expected improve the routing protocol performance lowering the energy consumption and decreasing the packet loss rate. In this way, like consequence of the presented benefits, it is expected attend the requirements of optimized energy solution and ubiquitous data exchange through proximity wireless technologies of IoT applications of most efficient form. In addiction, FSBRC can have their functioning optimized using the algorithm described in [18].

4. Results and Discussions

To evaluate the performance of the proposed approach, it used an extended version of Castalia simulator [19]. The simulations were executed two hundred times, aiming to obtain an margin error lower than 2% for the confidence interval of 95%. The application of IoT that was executed over the network is query-driven and aims to identify the tags that are in the reading area of the RS nodes, and also the local conditions (temperature, humidity and pressure) where the tags are. With exception of the sink node, all other nodes are deployed randomly and only tags have mobility. All the data collected by the RS nodes are sent to the sink node which is used as gateway to communicate with the Internet. The other parameters of simulation are shown in Table 1.

Table 1: Simulation Parameters

Parameter	Value
Simulation Area	50 m x 50 m
Time of Simulation	10 minutes
Base Station Location	(25,25)
Number of Nodes	20, 40, 60, 80, 100
Number of Tags	50
Initial Energy	100 Joules
Query Duration	200 seconds
Frequency of Sending Data	5 seconds
Mac Protocol	Tunable MAC

The evaluation of the proposed framework was carried by the comparing proposed framework against the state-of-art protocols. We simulate four main scenarios, described as follows:

- Scenario 1 (SC 1): the network uses RS nodes with DR architecture, routing protocol Directed Diffusion and anti-collision protocol EPC C1G2;
- Scenario 2 (SC 2): the network uses RS nodes with SDR architecture, routing protocol Directed Diffusion and anti-collision protocol EPC C1G2;
- Scenario 3 (SC 3): the network uses RS nodes with DR architecture and the proposed framework. The protocol used with the FSBRC is the Directed Diffusion.
- Scenario 4 (SC 4): the network uses RS nodes with SDR architecture and the proposed framework. The protocol used with the FSBRC is the Directed Diffusion.

The metrics used for evaluate the proposed framework are: query success rate (QSR), tag identification speed (TIS), packet loss rate (PLR) and average energy consumption (AEC). The QSR and TIS are obtained using the equations 1 and 2, respectively, where K is the number of inventories, Y_i is the number of query commands used during the inventory i, X_i represents the number of tags identified successfully during the inventory and T_i represents the length of inventory i in seconds. PLR and AEC are traditional metrics used to measure the performance of WSN. They are obtained using the equations 3 and 4, respectively, where N is the number of nodes, Ec_n is the energy consumed by each node n, Pr represents the quantity of packets received by the sink node and Ps_n represents the quantity of packets sent by each node n.

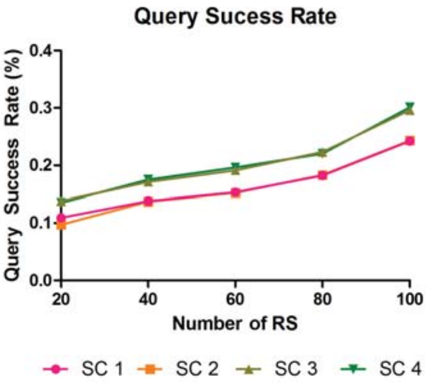


Fig. 1: Results for the metric of Query Success Rate.

$$QSR = \frac{\sum_{i=1}^{K} X_i}{\sum_{i=1}^{K} Y_i} \tag{1}$$

$$TIS = \frac{\sum_{i=1}^{K} X_i}{\sum_{i=1}^{K} T_i} \tag{2}$$

$$PLR = \frac{Pr}{\sum_{n=1}^{N} Ps_n} \tag{3}$$

$$AEC = \frac{\sum_{n=1}^{N} Ec_n}{N} \tag{4}$$

The Figure 1 presents the results of QSR to the evaluated scenarios. We can observe that the scenarios that use the proposed framework have an increase in the QSR metric when compared to scenarios that do not use the framework. The QSR is improved, on average, in 24% in the scenario that uses RS nodes with architecture DR and it improves 28% in the scenario where the RS nodes use SDR architecture. This improvement is justified due to dynamic definition of the parameter c. It is possible to adjust the size of the round in a more precise way, thus resulting in an increase of the success rate of each query. As benefit, the IoT application can collect data requested faster and more reliable. This way, the proposed framework can make the tracking and localization capability more efficiently and the ubiquitous data exchanged through proximity wireless technologies, which are important requisites of IoT applications.

The Figure 2 shows the results obtained in the simulations to the TIS metric. The results show that the scenarios that use the framework have a identification speed of tags bigger than the scenarios that do not use the framework. In the scenarios

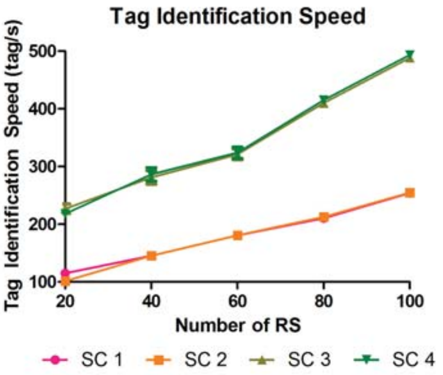


Fig. 2: Results for the metric of Tag Identification Speed.

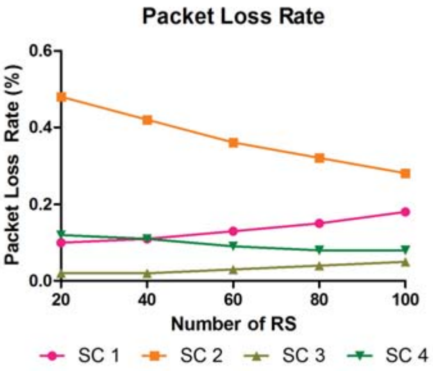


Fig. 3: Results for the metric of Packet Loss Rate.

that use RS nodes with DR architecture, the one that uses the framework achieve an superiority, on average, of 91% compared to the scenario that do not use the framework. In the scenarios that use RS nodes with SDR architecture the superiority in the use of the framework is, on average 97%. The use of the framework was capable of increasing in 115% the TIS in scenarios with twenty RS nodes with architecture SDR. Such results are justified due to fact that the framework avoids the sending of redundant identification codes, thus reducing the time necessary to finish the reading process. The speed in which the tags are identified can interfere in the power consumption and also in the rate of loss packets. RS-DR nodes that have low TIS can increase the time of the identification process thus resulting in an already bigger power consumption, because two communication interfaces were connected for a longer time. In RS-SDR nodes, low TIS value can affect the packet routing through the network, because when it is performing the identification process of the tags, the RS-SDR nodes are unavailable to receive packets from other nodes in the network. With the use of the proposed framework these problems are minimizes, thus making the IoT application may fulfill their goals in a more efficient way. Good values of TIS, just like QSR can be become more efficient the tracking and location capability and the exchange of data.

The Figure 3 presents the results obtained in the simulations for the metric of packets loss rate. The results show that, in the scenarios evaluated, the use of the proposed framework can reduce the rate of packets loss independently of the type of architecture used by the RS nodes. By means of the best route selection the proposed framework is able to avoid the loss of packets from broken routes, with poor quality of communication between the nodes or with high density of tags. By considering the power levels of each route, the framework avoid routes that have nodes with low quantity of power, that possibly can lead to the disruption of the route. The number of hops between the destination node and its source is also considered by the framework, by selecting routes with the smaller number of hops. By using LQI as parameter to qualify a route the framework proposed considers the quality of the links between the nodes that compose the route. Thus the routes with poor link quality will be avoided thus reducing the amount of packets lost. Another important factor considered by the framework in the moment of measuring the quality of a route is the tag density in the area of reader nodes. By considering these parameters the framework can avoid that the packets be forwarded to routes that have nodes which are overloaded with the process of reading the tags. This analysis is even more important in scenarios where the RS nodes use SDR architecture, once those nodes that are reading cannot forward packets to other nodes. This way, avoiding routes that have nodes in reading state is extremely important to reduce the rate of packets loss. Like benefits of this packet loss reduction, the network used by the IoT applications can provide a most efficient structure, enhancing the ubiquitous data exchanged through proximity wireless technologies.

The results of average power consumption metric in the nodes are presented in the Figure 4. The results show us that the scenarios that use the framework have an average consumption smaller when compared to the scenarios that do not use the framework. Independently of the architecture used by the RS nodes, the framework can reduce the consumption of power in 5%, on average. This happens because with the use of the framework o time spent in the process of tag identification is reduced, resulting in a smaller consumption of power during each identification process. In some applications, the low rate of packet lost that was mentioned before can also contribute to reduce the power consumption. With the reduction in the number of packets lost it also reduced the need to resend control messages, causing a small use of the interfaces of communication thus resulting in saving the power. Thus the proposed framework can reduce the average power consumption proportioning an optimized solution of power to IoT applications.

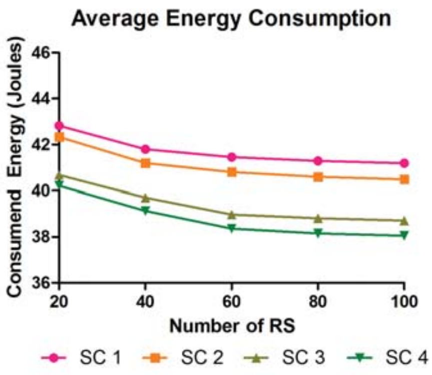


Fig. 4: Results for the metric of Average Energy Consumption.

5. Conclusions

IoT can be understood as the interconnection of different types of objects and things that communicates through a network infrastructure with the goal of cooperating for a common objective. In order for IoT applications perform their tasks, it is necessary that some requirements be provided by the network structure. However, these current technologies are rarely capable of providing all these resources. Looking for the required resources of the IoT applications, several works have proposed the integration of WSN and RFID. However, these works generally do not consider the problems that arise with the integration of the technologies involved. In this way, this work proposes a framework to offer a most efficient network structure for IoT applications taking in account the problems that are augmented after the integration of WSN and RFID.

Generally, it was observed that, for the scenarios assessed, the proposed framework on this work can enhance the performance of the network structure used by IoT applications. The framework was able to increase the query success rate and the tag identification speed, reduce the packet loss rate and decrease the energy consumption. These benefits are very important for IoT applications. Thus, with the utilization of the proposed framework it is possible to improve the network and provide, with efficiency, some required resources of the IoT applications, as: data exchange among ubiquitous technologies, optimized energy solution, localization and tracking capabilities and devices heterogeneity.

As future works, it is intended to perform simulations considering denser scenarios (greater number of elements) that better represent IoT scenarios. It is also intended to perform new studies on the components that integrate the framework with the purpose of searching improvements or new mechanisms to compose them.

References

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of things (iot): A vision, architectural elements, and future directions,” *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[2] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, and T. Razafindralambo, “A survey on facilities for experimental internet of things research,” *IEEE Communications Magazine*, vol. 49, no. 11, pp. 58–67, 2011.

[3] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[4] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, “Internet of things: Vision, applications and research challenges,” *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[5] F. M. Al-Turjman, A. E. Al-Fagih, W. M. Alsalih, and H. S. Hassanein, “A delay-tolerant framework for integrated rsns in iot,” *Computer Communications*, vol. 36, no. 9, pp. 998–1010, 2013.

[6] S. Rajesh, “Integration of active rfid and wsn for real time low-cost data monitoring of patients in hospitals,” in *2013 International Conference on Control, Automation, Robotics and Embedded Systems (CARE)*, Dec 2013, pp. 1–6.

[7] G. Yang, M. Xiao, and C. Chen, “A simple energy-balancing method in rfid sensor networks,” in *2007 IEEE International Workshop on Anti-counterfeiting, Security, Identification*. IEEE, 2007, pp. 306–310.

[8] D. K. Klair, K.-W. Chin, and R. Raad, “A survey and tutorial of rfid anti-collision protocols,” *Communications Surveys & Tutorials, IEEE*, vol. 12, no. 3, pp. 400–421, 2010.

[9] W. K. Seah, Z. A. Eu, and H.-P. Tan, “Wireless sensor networks powered by ambient energy harvesting (wsn-heap)-survey and challenges,” in *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on*. Ieee, 2009, pp. 1–5.

[10] A. Abahsain, A. E. Al-Fagih, S. M. Oteafy, and H. S. Hassanein, “Selective context fusion utilizing an integrated rfid-wsn architecture,” in *2013 IEEE Consumer Communications and Networking Conference (CCNC)*. IEEE, 2013, pp. 317–322.

[11] A. Mason, A. Shaw, A. Al-ShammaâĂŹam, and T. Welsby, “Rfid and wireless sensor integration for intelligent tracking systems,” in *Proceedings of 2nd GERI Annual Research Symposium GARS-2006*, 2006.

[12] H. Liu, M. Bolic, A. Nayak, and I. Stojmenovic, “Taxonomy and challenges of the integration of rfid and wireless sensor networks,” *IEEE Network*, vol. 22, no. 6, pp. 26–35, 2008.

[13] W. H. Tuttlebee, *Software defined radio: enabling technologies*. John Wiley & Sons, 2003.

[14] EPCglobal, “Epctm radio-frequency identity protocols class-1 generation-2 uhf rfid protocol for communications at 860 mhz–960 mhz version 1.2.0,” 2008.

[15] H. S. Ramos, A. C. Frery, A. Boukerche, E. M. Oliveira, and A. A. Loureiro, “Topology-related metrics and applications for the design and operation of wireless sensor networks,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 10, no. 3, p. 53, 2014.

[16] K. Machado, D. Rosário, E. Cerqueira, A. A. Loureiro, A. Neto, and J. N. de Souza, “A routing protocol based on energy and link quality for internet of things applications,” *Sensors*, vol. 13, no. 2, pp. 1942–1964, 2013.

[17] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, “Directed diffusion for wireless sensor networking,” *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, pp. 2–16, 2003.

[18] J. V. Sobral, R. A. Rabelo, H. S. Araujo, R. A. Baluz, *et al.*, “Automated design of fuzzy rule base using ant colony optimization for improving the performance in wireless sensor networks,” in *2013 IEEE International Conference on Fuzzy Systems (FUZZ)*. IEEE, 2013, pp. 1–8.

[19] A. Boulis, “Castalia: revealing pitfalls in designing distributed algorithms in wsn,” in *Proceedings of the 5th international conference on Embedded networked sensor systems*. ACM, 2007, pp. 407–408.

Prediction and Recovery based Smart Target Tracking Mechanism for WSNs.

Rida Nisar, Muazzam A. Khan, Nazar A. Saqib, Jahan Zeb
NUST College of Electrical and Mechanical Engineering,
National University of Sciences and Technology, Islamabad, Pakistan.
muazzamak@ceme.nust.edu.pk

***Abstract*—Wireless sensor networks encompasses a wide range of applications which not only covers a strong infrastructure of communication but also highlights the broader aspects of security, surveillance, military, health care and environmental monitoring. Among these applications of wireless sensor networks target tracking is very essential which is installed in the required areas of surveillance for tracking any attacker/intruder and habitat monitoring. Bandwidth and power consumptions are the two inevitable constraints for meeting the demand of localization and energy levels. This paper focuses on target tracking in WSNs using clustering and prediction based protocol. Base station acts as the cluster formation manger and indicates when the moving target is witnessed.**

***Index Terms*— Target tracking, intruder/attacker, clustering, prediction.**

I. Introduction

Wireless sensor network is an assembly of special transducers with a capability of communication as well as monitoring and recording conditions at diverse locations. These multiple detection stations are called senor nodes, each equipped with a sensor, transceiver and a power source. On the basis of sensed circumstances around, this specialized sensor is believed to act as electrical signal generator which entirely depends upon physical effects encountered. These sensors communicate via radio waves and differs from place to place from hundred to hundreds of thousands based on the requirement of the environment. Sensors deployed have small size, limited energy and low memory.

One of the most tremendous and vital applications of WSNs is target tracking that is very important issue with a wide spectrum of research and several applications. The concept is simple i.e. any moving target whether a person, vehicle or object can be tracked traversing in a WSNs with sensing capability of sensor nodes. Location and position of the moving object is constantly recorded, studied and compared each time unit with some reference point in order to declare the exact position and location of the target. It is similar to a feedback loop where a constant comparison is done between input and output. Therefore, it can be concluded that two critical tasks are involved in the object tracking management: 1) continuous supervision and 2) broadcasting, where continuous supervision is the monitoring of sensor nodes to keep a track of moving objects while broadcasting means to report that a moving object is detected to the relevant application running.

Clustering and prediction based protocols are presented in this paper, where the two critical parameters are distance and energy. In addition to these two parameters, the involvement of base station gives the true sense of tracking an intruder or a moving object in WSNs. Base station manages the prediction and clustering formation therefore it has maximum information about energy levels of each node in the network.

II. Related Work

WSNs can be categorized under two main headings i.e. hierarchical and peer-to-peer. Inside hierarchical network, a mesh oriented multi hop radio based connectivity among or between wireless nodes is employed. This is shown in Fig. 1.

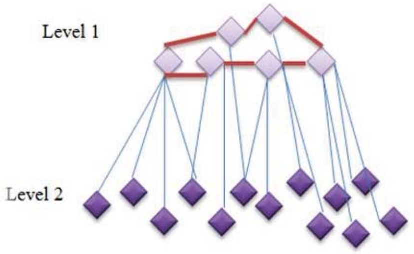


Figure 1: Hierarchal Network [2]

In this network the nodes of level 1 are backbone nodes while the nodes of level 2 are sensor nodes. Any event in the vicinity of the nodes can be detected as well as reported to the sink by the nodes, and the sink can further communicate to the outside world such as laptop, mobile etc.

The second category is peer to peer network in which, instead of a mesh based multi-hop radio connectivity there is flat single-hop radio connectivity among or between wireless nodes is deployed. This is shown in Figure 2.

Peer to peer or point to point network uses static routing on wireless networks. Every node is only able to communicate with its neighboring nodes and information can be exchanged between neighbors.

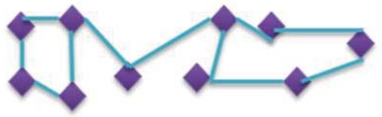


Figure 2: Peer To Peer Network

Hierarchal network can be further classified into four schemes as shown in Figure 3. These are Naive activation based tracking, tree based tracking, cluster based tracking and hybrid approaches [1]. In Naive activation based tracking scheme, every node of the network is in tracking mode all the time [6]. Each node sends its measured position of the target to the base station which further calculates and predicts the accurate position of the target on the basis of received information. This strategy is the simplest and offers best tracking results but has worst energy consumption as all the nodes are in tracking mode simultaneously, therefore reflects heavy computation and burden on base station. So this approach is not robust in case of link failure as well as in channel congestion scenario.

In tree based tracking protocols, the nodes in the network can be organized as a graph in which the vertices can be considered the nodes and the edges can be viewed as the connections, which indicate the communication between nodes in the network. Scalable Tracking Using Networked sensors (STUN) [2], [3] and Dynamic Convoy Tree-Based Collaboration (DCTC) [4] protocols are tree-based tracking approaches. In these schemes in addiction to nodes and vertices, a cost is also assigned to each communication. Leaf nodes are used for mobile target tracking and transmission of the collected data. Nodes then, send the information to sink through intermediate nodes.

Cluster based tracking approaches provides not only scalability in networks in both terms i.e. addition or removal of nodes but also it helps in better bandwidth usage. Wireless sensor networks divide the nodes into several clusters each having a boss or cluster head which is elected randomly or on specific criteria. Each cluster head is responsible for collecting data from the nodes in its cluster and send it to the base station Cluster based tracking can be categorized into further three types: i) Static Clustering, ii) Dynamic Clustering, and iii) Space Time Clustering.

In static clustering, clusters are formed at the time of network formation. The size of cluster, area it covers as well as the members under one cluster remains constant throughout the network. This simple approach has many drawbacks, such as it is not fault tolerant and any failure of cluster head renders all the nodes in the network useless. Due to the fixed parameters, this approach does not work well in the dynamic environment where a node is needed to be in awake mode or sleep mode [9].

Target Tracking	
Hierarchical	**Peer to Peer**
• Naïve activation • Tree based tracking • Clustering based ○ Static ○ Dynamic ○ Space-time • Hybrid	• Embedded filter based • Alternating direction based

Figure 3: Categories of Target Tracking in WSNs

Dynamic clustering is useful in several ways. Formation of clusters is purely conditional on the events happening in the surroundings [10]. Sensors which are near the active cluster heads are invited to become members of the cluster and respond to the cluster head. In this approach, nodes can belong to different clusters at different times. As, only one cluster is activated on the basis of maximum probability in vicinity of a target. Therefore, the contention at MAC level is completely mitigated and redundant data is suppressed. Information driven sensor querying (IDSQ) [11], is one of the examples of dynamic clustering. Space time clustering is the third type of clustering technique in WSNs. [12] proposed this architecture in which clusters of space time neighboring nodes are organized dynamically and the information is propagated around on the basis of local messages in the space time cluster. Cluster head estimates the track of the target by this collected data.

The infrastructure of hybrid methods relies on more than one type of target tracking schemes. Distributed predictive tracking (DPT) and Hierarchical prediction strategy (HPS) are the examples of hybrid methods. In DPT the issue of scalability is resolved using clustering based protocol and prediction based method is adopted for efficient energy solution [7]. In HPS, a cluster is formed using Voronoi division and next location to be targeted is predicted using least square method [8].

Peer to peer networks as shown in Figure 3 can be classified into two classes which are: embedded filter based consensus and Alternating-direction based consensus. The first technique is a two steps method in which every time on exchanging the information among nodes an update step is resided. In a nutshell, each step of the algorithm comprises an exchange of information exchange and update process. Gaussian random parameters for decentralized estimation was established in [13] by Delouille et al for stationary environments and dynamic case was considered in [14] by Spanos. The second method of peer to peer networks is Alternating-Direction based consensus; in this technique simplicity and flexibility is confirmed due to the presence of single hop communications among sensors. The algorithm makes sure that the mean square error is minimized. A degree of parallelization is achieved in alternating-direction based consensus. Sensor failure is handled by using a subset of "bridge" sensors, as described in [15-18].

III. Problem Statement

The main short coming of previous algorithms is that what if a target is there but it is missed by the nodes, i.e. target is lost due to several reasons. This paper designs a target recovery mechanism when a target is missed. Figure 4 shows the block diagram of steps involved in the target tracking algorithm.

IV. Methodology

This paper proposed a smart mechanism for target detection, prediction and recovery of a node is done using efficient energy clustering protocol.

A. Clustering

The main concept of clustering is to group the sensor nodes and collect information from neighboring nodes and send it to the base station. Information may include monitored environment or intruder tracking or an atmosphere where human access is not easily possible. Usually base station is outside the field area, which process the acquired data and reports to the user.

There are number of protocols for clustering having different pros and cons. This paper suggests the most energy efficient protocol known as LEACH (*low energy adaptive clustering hierarchy)* as proposed in [16]. This is TDMA (time division multiple access) MAC protocol which is designed for clustering in wireless sensor networks. The main focus of LEACH is pro-long the life span of a network by making it efficient in energy consumption. It is hierarchical protocol in which the sensor nodes collaborate or respond to cluster head and then it is the responsibility of the cluster head to further transmit the data to the base station. In every iteration of the protocol, there is a kind of polling mechanism in which each node waits and determines whether it will become the next cluster head or not. Nodes except cluster heads communicate via cluster heads in TDMA fashion. LEACH is an iterative process, in which each round begins with a set up phase, then the clusters are organized and nodes inform the specific cluster head to which head they belong then comes the data transmission phase where TDMA is used to exchange information between heads [17-20].

B. Target Detected

Target is detected using Received Signal Strength Indicator (RSSI). It calculates the distance between two sensors by measuring the power transmitted from sender to receiver. It is logarithmic ratio of power of received signal and reference power. It is known that power dissipates from a point source when it moves further and has inverse relation with distance, therefore, distance can be easily computed using this relationship. The main advantage of RSSI is low computational cost because most receivers are capable of estimating the received signal strength. Although in some cases inaccuracies of distance estimation is found but this can be improved by using RSSI with better transmission. The bigger the distance to the receiver node, the lesser will be the signal strength when it arrives. RSSI plays a significant role in deciding which link to use in order to make packet delivery optimal. Target localization can be optimized using other techniques showing best results.

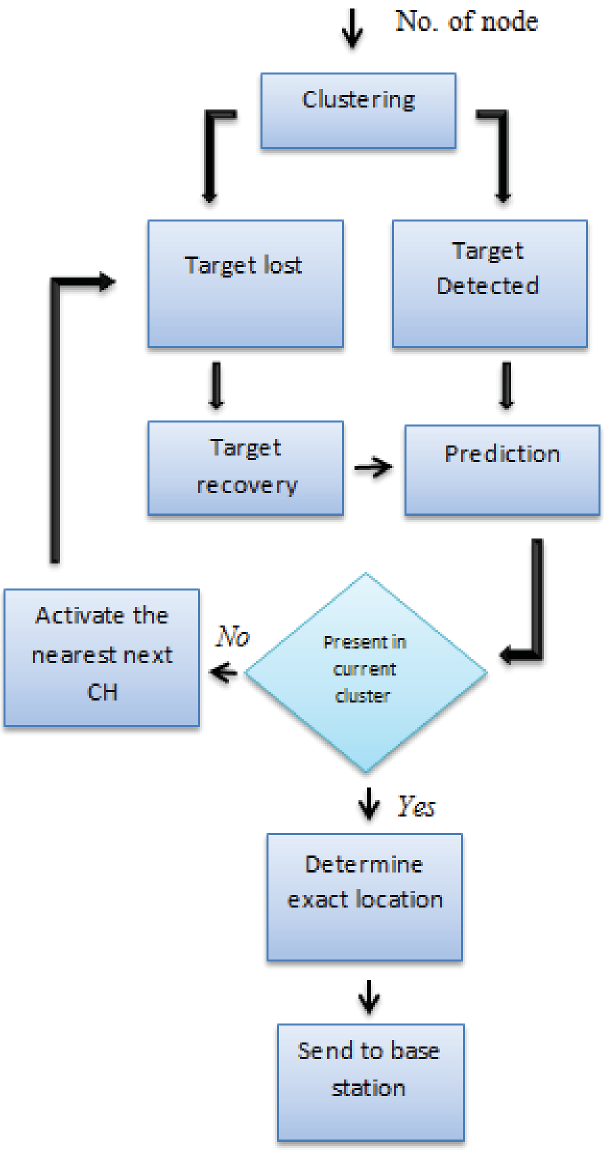


Figure 4: Flow chart for target tracking

C. Prediction

Next possible location of the target can be computed using prediction mechanism. As prior knowledge can help in determining the posterior position, similarly prediction method computes the next position of the target. This works in a linear fashion, for example at j^{th} instant, the co-ordinates of the target are(x^j, y^j), and at j^{th+1} instant, co-ordinates may shift to(x^{j+1}, y^{j+1}) which is calculated using two important parameters like target speed and direction. If the predicted location is with in current cluster then cluster head informs the base station, else the command is given to the next nearest cluster head in vicinity of predicted co-ordinates of the target. After handing over the command, the current cluster goes to sleep mode in order to save as much energy as possible.

D. Target lost

Sometimes, a network fails to track the target properly and target is lost which has several possible reasons, following three can be few possible reasons behind target loss:

1. Failure of nodes or network

Sensor nodes are usually battery operated and when it comes to battery, there is a chance of low battery which in turns results in failure of sensors in the network. Network fails due to heavy traffic and physical malfunctioning

2. Error in target detection step

There is a possibility that measured distance is incorrect as the target enters the cluster and moves very fast so localization can be incorrect in such a scenario. This can also happen due to uncertain changes in target's velocity and direction

3. Handing over the command

When the current location is estimated and found that target is not present in the current cluster, in such case, next command is handed over to next nearest cluster head. Due to energy issues, clusters involved in a network are not always in awake mode, this may result in target loss
To handle such cases where target is lost there must be recovery mechanism as in real time applications, this can be hazardous.

E. Recovery Mechanism

As mentioned earlier, cluster head is responsible for predicting next location of the moving target and it keeps on sending warning messages to the next nearest cluster head in case when target is out of the reach of current cluster. It waits for the acknowledgment message, but if fails to receive any, then broadcast the target lost message to the entire network. This procedure is well described in [17]
Immediate steps shown in Figure 5 are taken by current cluster when it realizes that by any means a target in the network is lost.

V. CONCLUSION

This paper presents different network schemes and their further classification according to characteristics and capabilities of sensor nodes. It highlights the concept of clustering of nodes in the network. One of the main constraints of WSNs is power limitation which can be reduced to some extent by using energy efficient protocol LEACH for clustering. Localization of target can be achieved by received signal strength indicator. The method can be improved on the basis of experimental results on a network simulator by using different distance measuring schemes with different energy efficient protocols. The main objective of the paper was to introduce a mechanism for target tracking in wireless sensor networks as well as it discovers some of the scenarios in which target is lost and can be recovered.

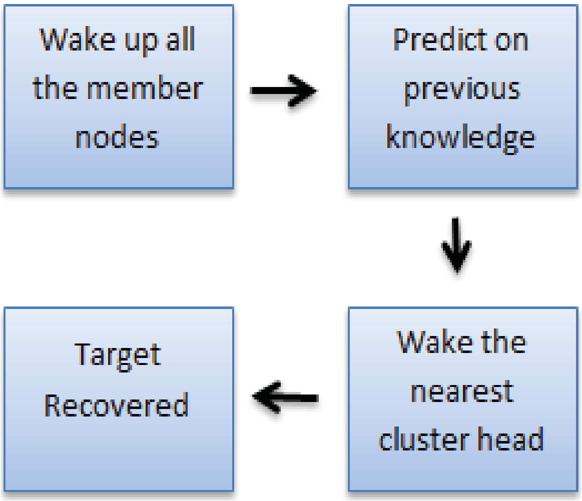

Wake up all the member nodes
Predict on previous knowledge
Wake the nearest cluster head
Target Recovered

Figure 5: Steps taken by CH when target is lost

REFERENCES

[1] S. Bhatti, J. Xu, "Survey of Target Tracking Protocols using Wireless Sensor Network", In Proc. Fifth International Conference on Wireless and Mobile Communications, IEEE, 2009

[2] C. Y. Lin and Y. C. Tseng, "Structures for In-Network Moving Object Tracking in Wireless Sensor Networks", Proceedings of the First International Conference on Broadband Networks (BROADNETS) IEEE, 2004

[3] H. T. Kung and D. Vlah, "Efficient Location Tracking Using Sensor Networks", Proc. IEEE Wireless Comm. and Networking Conf, Vol. 3, 2003

[4] W. Zhang and G. Cao, "DCTC: Dynamic Convoy Tree-Based Collaboration for Target Tracking in Sensor Networks," IEEE Transactions on Wireless Communications, Vol. 3, 2004.

[5] P. Azad and V.Sharma, "Cluster head selection in wireless sensor networks under fuzzy Environment" Vol. 3, 2013

[6] T.Hsieh, "Using Sensor Networks for highway and traffic applications", Vol.23, 2004.

[7] J. W. Bakal "Simulation of Target Tracking in Wireless Sensor Network", International Journal of Advanced Research in Computer Science and Software Engineering Vol. 4, February 2014

[8] Z. Wang et al, "Predicting Moving Targets in Hierarchical Sensor Networks" in Networking Sensing and Control, IEEE International Conference, 2008

[9] P. Marsch, "Static Clustering for Cooperative Multi-Point (CoMP) in Mobile Communications" in Communications (ICC), IEEE International Conference on , 2011

[10] F. Bajaber et al, “Dynamic/Static Clustering Protocol for Wireless Sensor Network” in Computer Modeling and Simulation, IEEE, 2008

[11] F. Zhao et al, “Information-Driven Dynamic Sensor Collaboration for Tracking Applications” in IEEE Signal Processing Magazine 2002

[12] S. Phopa et al, “Sensor Network based ;ocalization and target tracking through hybridization in the operational domains of beam forming and dynamic space-time clustering” in global telecommunications, IEEE Vol.5 2003

[13] D. Schizas et al, “Consensus in Ad Hoc WSNs with Noisy Links: Part I: Distributed Estimation of Deterministic Signals” in IEEE Transactions on Signal Processing, Vol. 56, 2008

[14] Y. Peng “Decentralized Estimation and Control of Graph Connectivity in Mobile Sensor Networks”, 2008

[15] I. Schizas et al, “Distributed LMS for Consensus-based in-network adaptive processing” in Signal processing, IEEE transactions on Vol. 57, 2009”

[16] R.Wendi et al, “Energy-Efficient Communication Protocol for Wireless Micro sensor Networks” Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000

[17] G. Mohini et al, “Cluster Based Target Tracking and Recovery Algorithm”IJASCSE, Issue 1, Vol 4, Dec 2012.

[18] Vahid Hosseini et al, “Designing a Clustering and Prediction based Protocol for Target Tracking in Eireless Sensor Networks (WSNs)” ACSIJ Advances in Computer Science: Vol 2, Issue 3, No 4, July 2013.

[19] T. Ali, Muazzam A. Khan et al, “Secure Actor Directed Clustering in WSANs” International Journal of Distributed Sensor Networks, Vol 2013, Sep 2013.

[20] Muazzam A. Khan et al, “An Efficient and Reliable Clustering Algorithm for WSANs” Circuits and Systems (MWSCAS), 2010 53rd IEEE International Midwest Symposium on , vol., no., pp.332,338, 1-4 Aug. 2010.

[21] Muazzam A. Khan et al, “A QoS Based Multicast Communication Framework for WSANs” International Journal of Innovative Computing, Information and Control, Vol 7, No 12, Dec 2011.

Achieving energy-balance with unequal clustering under single-hop transmission in wireless sensor networks

Zhang Jing[1,2], Liu Yanheng[1,2], Li Bin[3,*], Li Lin[1,2], Sun Geng[1,2]
[1]College of Computer Science and Technology, Jilin University. Changchun, Jilin 130012, China;
[2]Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University. Changchun, Jilin 130012, China.
[3]College of Mathematics, Jilin University, Changchun, Jilin 130012, China
* Corresponding author. E-mail addresses: lb@jlu.edu.cn

Abstract - In cluster-based protocols, cluster heads (CHs) located far from the base station (BS) always consume more energy than those nearby under single-hop transmission. These CHs reach their end of life earlier than others and limit the network lifespan. This study presents a new algorithm that attempts to achieve energy consumption balance of sensor nodes. An analysis is made on the basis of energy consumption difference of sensor nodes which is caused by diverse transmission distances. Furthermore, the optimal unequal cluster size is obtained according to the distance between node and the BS. The number of CHs has an increasing tendency along with increasing distance to the BS. By comparing our algorithm with LEACH, EECS and I-LEACH, the network lifespan is prolonged by up to 66.7%, 51.8% and 33.1% respectively, and the proposed algorithm balances the energy consumption of nodes effectively.

Key words: wireless sensor networks; unequal clustering; energy-balance; network lifespan

1 Introduction

As for WSNs, Real-time communication is essential in some applications, such as surveillance systems[1,2] and real-time patient monitoring[3]. Single-hop transmission clearly has minimum transmission delay. It is meaningful to prolong the lifespan of energy-limited WSNs under single-hop transmission. The Low-Energy Adaptive Clustering Hierarchy (LEACH)[4] is the first hierarchical architecture protocol. The cluster size of all clusters is equal. However, the many-to-one data transmission pattern in WSNs causes sensor nodes located far from the BS to consume more energy than those near the BS under single-hop transmission. The early death of nodes limits the lifetime of network. If the cluster size is unequal, then the CHs near the BS can extend their cluster sizes, and the CHs far from the BS can lessen their cluster sizes to conserve energy. The energy consumption balance of sensor nodes is thus achieved, and the lifetime of network will be prolonged finally.

Many LEACH-based protocols[5-8] have been presented to promote its performance. Most of improvement studies focus on CH optimization. The expected number of CHs is close to five percent of the number of nodes in LEACH, but it cannot be guaranteed that the number of CHs is always optimal. For solving this problem, an improved protocol based on LEACH presented in [5]. CHs are elected dynamically according to the residual energy of sensor nodes, which stabilizes the expected number of CHs with the energy information. The number of CHs is varied with the number of living nodes that is fixed in LEACH. The optimal number of CHs under different scenarios is presented in [6], the optimal cluster size of each cluster is also obtained, but each cluster has the same size. The reference information that residual energy, number of neighbors, and distance to the BS are used to generate a threshold for each sensor node in the CH selection process[7], but the number of CHs cannot be guaranteed around an optimal value. EECS[8] selected CHs in an autonomous manner through node competition with energy.

Sensor nodes with more residual energy become the final CHs. All clusters are also assumed to be the same size. All of above mentioned protocols focus on the aspect of CH optimization. Actually, the size of each cluster should also be optimal, because the transmission energy consumption cost by the CHs near the BS is significantly less than that far from the BS. Thus, the number of cluster members in the cluster near the BS should be increased, and the number of clusters far from the BS should be decreased. In this way, the balanced energy consumption of nodes will be achieved. The network lifetime will not be limited by the early death of nodes. Motived by these factors, we present a new method called achieving energy balance with unequal clustering (AEBUC) for WSNs.

The reminder of this paper is organized as follows. Section 2 describes the network and energy models used in this study. Section 3 presents the analysis on energy consumption of a cluster in the network and the computation of optimal cluster size for each node. Section 4 presents the AEBUC details, including setup and steady phases. Section 5 analyzes the performance of AEBUC against three other algorithms. Finally, Section 6 provides some conclusions.

2 System Model Assumption

2.1 Network Model

N sensor nodes are distributed randomly and uniformly on a monitored area with the length L and the width W. The BS is near the monitored area with a fixed location. The network parameters and conditions are assumed as follows:

1. All sensor nodes with a unique ID are homogeneous and static after deployment.

2. Each sensor node generates packets at the speed of k bits per second and sends packets to its CH.

3. Cluster heads deliver their packets to the BS with single-hop.

4. The transmission radii of sensor nodes are adjustable as illustrated in [9].

5. The approximate distance can be computed based on the received signal strength between two sensor nodes.

The network node density ρ is given by the following:

$$\rho = \frac{N}{L \times W} \tag{1}$$

2.2 Energy model

The energy consumption of sensor nodes generally consists of three parts: transmission energy consumption $E_{t,}$ receiving energy consumption E_r and sensing energy consumption E_s. E_s is frequently too small and is thus ignored, such as in [10-11]. Thus, E_t and E_r comprise the major portion of energy consumption in this study. E_t is defined in [12] as follows when sensor node i transmits a packet with k bits to node j:

$$E_t = \begin{cases} kE_{elec} + k\varepsilon_{fs} d^2, & if\ d < d_0 \\ kE_{elec} + k\varepsilon_{mp} d^4, & if\ d \ge d_0 \end{cases} \tag{2}$$

$$d_0 = \sqrt{\varepsilon_{fs} / \varepsilon_{mp}} \tag{3}$$

E_t is dependent on the transmission distance d between i and j. According to the value of d the propagation loss can be modeled as free-space model or muti-path attenuation model, where E_{elec} is the electronics energy, ε_{fs} and ε_{mp} denote the amplifier energy. Reception energy consumption E_r is defined as

$$E_r = kE_{elec} \tag{4}$$

Data aggregation is used in this study for highly correlated of sensed data packets. Cluster heads are assumed to aggregate the data received from its cluster members into a single length-fixed packet. Energy consumption for data aggregation is equal to $k*E_{DA}$(nJ/bit/signal).

3 Theoretical analysis

The energy consumption of a cluster in the network can be analyzed on the basis of the energy model described in Section 2.2. The transmission energy consumption of a cluster member is determined by the size of packet k and distance d between the CH and cluster member. This value can be expressed as follows:

$$E_{CM} = k(\pi\rho R^2 - 1) \cdot (E_{elec} + \varepsilon_{fs} \overline{d_{toCH}}^2) \tag{5}$$

where $\overline{d_{toCH}}$ is the average distance between the cluster member and its CH. It is assumed that the cluster is a circle with a radius R and the CH is in its center. Any node in this cluster can be expressed with coordinate of (x, y). Therefore, the expected distance, as well as $\overline{d_{toCH}}$ between any node and the center, can be expressed as

$$\begin{aligned} E(d) &= \overline{d_{toCH}} = 1/\pi R^2 \cdot \iint \sqrt{(x^2+y^2)}dxdy \\ &= 1/\pi R^2 \cdot \iint r^2 dr d\theta = 1/\pi R^2 \cdot \int_0^{2\pi} d\theta \int_0^R r^2 dr = \frac{2R}{3} \end{aligned} \tag{6}$$

Thus, the Eq. (6) can be expressed with the cluster radius R as

$$E_{CM} = k(\pi\rho R^2 - 1)\cdot(E_{elec} + \frac{4}{9}\varepsilon_{fs}R^2) \tag{7}$$

Equation (7) shows that the energy consumption of cluster members in a cluster is related to the size of packet k and the radius of cluster R.

The energy consumption of CHs consists of three parts, namely, reception energy consumption, aggregation energy consumption, and transmission energy consumption. Thus, E_{CH} can be expressed as follows:

$$\begin{aligned} E_{CH} &= k(\pi R^2\rho - 1)E_{elec} + k\pi R^2\rho E_{DA} + kE_{elec} + k\varepsilon_{fs}d_{toBS}^2 \\ &= k\pi R^2\rho(E_{elec} + E_{DA}) + k\varepsilon_{fs}d_{toBS}^2 \end{aligned} \tag{8}$$

Equation (8) shows that the energy consumption of CHs in a cluster is related to the size of packet k and distance d_{toBS} between the CH and BS. On the one hand, energy efficiency can be achieved by compressing the data packets, which are smaller packets with highly accurate information. On the other hand, distance affects energy consumption. Energy efficiency can be achieved by decreasing the number of cluster members in a cluster far from the BS. The size of the aggregation packet is k in this study which is the same as that used in [6]. Thus, the relationship among energy consumption, cluster size, and distance to BS is distinct. The energy consumption of a cluster can be expressed as follows:

$$E_{cluster} = E_{CM} + E_{CH} \tag{9}$$

The energy consumption of a cluster is determined by the cluster radius R and distance d_{toBS} between the CH and BS. The distance between the CH and BS is definite once a network is deployed. Thus, the cluster radius R is the only parameter that affects the energy consumption of a cluster. Equation (7) shows that more energy is consumed by a cluster with bigger radius and a cluster farther from the BS. Thus, a tradeoff value of cluster radius R and distance d_{toBS} between the CH and BS can be determined to balance the energy consumption of each cluster. The lifespan of the network will be prolonged accordingly.

The maximum and minimum distances between the CH and BS are definite once a network is deployed. Thus, Eq. (10) can be used to obtain the minimum cluster radius R_{min} and the maximum radius R_{max} with assurance of minimum energy consumption of a cluster.

$$\begin{cases} E_{cluster}^{MAXd_{toBS}}(R_{\min}) = E_{cluster}^{MINd_{toBS}}(R_{\max}) \\ \min(E_{cluster}^{MAXd_{toBS}}(R)) \end{cases} \tag{10}$$

$$\begin{aligned} & E_{cluster} = \min(E_{cluster}^{MAXd_{toBS}}(R)) \\ & = k(\pi\rho R^2 - 1)\cdot(E_{elec} + \frac{4}{9}\varepsilon_{fs}R^2) + k\pi R^2\rho(E_{elec} + E_{DA}) + k\varepsilon_{fs}d_{toBS}^2 \end{aligned} \tag{11}$$

where, E(R) is the energy consumption of a cluster. The size of cluster with the farthest distance to the BS should be minimum and vice versa. The optimal cluster radius of each sensor node can be obtained simply by using Eq. (11), and applying in cluster formation process.

4 AEBUC details

4.1 Setup Phase

The setup phase consists of CH selection and cluster formation. CH selection begins once the cluster size of each node is obtained, which is also named competition radius for each node to complete CH competition as in [8]. Several sensor nodes with a random value between 0 and 1 less than a threshold T are selected as candidate cluster heads (CCH) first. CCHs compete with their neighbors to select final CHs with higher energy.

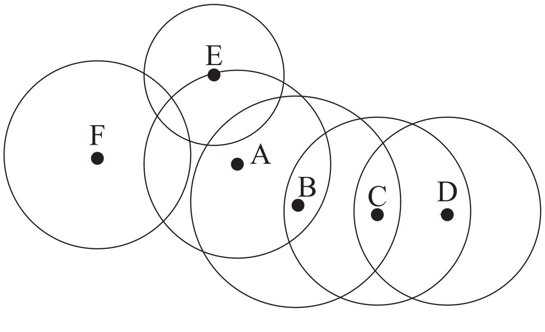


Fig.1 Node Competition Relationship

An illustration is shown in Fig.1 to describe the competition process. If the distance between CCH *A* and CCH *B* is less than the maximum radius of the two nodes, then nodes *A* and *B* are neighbors of each other. Figure 1 shows that CCHs *B* and *E* are neighbors of CCH *A*, but not CCH *F* because the distance between *A* and *F* is larger than the maximum radius of the two. CCHs *A* and *E* will quit competition if CCH *B* has the most residual energy among *A*, *E* and *B*. If the situation that the residual energy of CCH *C* is larger than CCH *B* and CCH *D* is greater than CCH *C* exist, CCH *C* becomes the final CH finally. The iteration competition is limited to two to avoid the spare distribution of CHs. Thus, the final CH *C* is excluded in the competition with CCH *D*.

Cluster formation begins following the CH selection. The final CHs broadcast short packets that contain the node id and distance to the BS d_{toBS} to notify normal nodes to be the final CH. Normal nodes select an ideal cluster to join to accomplish cluster formation. The ideal cluster means that the cluster member consumes less energy to achieve packet transmission, and the joined cluster member does not cause the CH far from the BS to be burdened with excessive load. Thus, the cost expressed in Eq. (12) is used as the metric to evaluate which cluster is the ideal one for each normal node to join in. Normal nodes store all received broadcast packets to compute for the cost of the CH and then select one CH with the minimum cost to join in.

$$\cos t(i) = c \cdot \frac{d^{i}_{toCH}}{MAX(d_{toCH})} + (1-c) \cdot \frac{d^{i}_{toBS} - MIN(d_{toBS})}{MAX(d_{toBS}) - MIN(d_{toBS})} \quad (12)$$

where d_{toCH} is the distance between cluster member and CH. MAX(d_{toBS}) and MIN(d_{toBS}) are the maximum and minimum distances between the CH that normal node can hear and the BS. The cost computation includes two parts, the former part is to find the closest cluster head to join in, and the later part is to join in the cluster that its CH is nearer BS. A tradeoff consideration is made in Eq. (12) using parameter *c*, which is used to balance the CH load. The setup phase is finished when all normal nodes determine their ideal CH.

4.2 Steady phase

All sensor nodes are in the process of data packet transmission in the steady phase. Normal nodes transmit packets to their CHs based on the TDMA schedule allocated by their CHs. Cluster heads collect all cluster member packets, compress these packets into one packet, and finally transmit it to the BS directly. The network enters the setup phase again when the entire process is completed. Thus, the network lifespan is split into rounds. Each round is made up of setup and steady phases.

5 Simulation results

We simulate our algorithm and three other cluster-based algorithms (i.e., LEACH, EECS, and I-LEACH) with MATLAB. A performance comparison is conducted on these algorithms. Several characteristics of the four algorithms are listed in Table 1. Three types of CH selection methods among the four algorithms are considered in terms of load balancing, except for LEACH. The cluster sizes are the same in the three other algorithms but are unequal in AEBUC. The cluster size ranges from 25 to 28 in AEBUC, whereas it is set to 26 in EECS.

Table 1 Characteristics of the four algorithms

Protocols	CH selection	Load balancing	Cluster size or K_{opt}
LEACH	Rotation without any information	No	20
EECS	Competition	Yes	26
I-LEACH	Rotation with energy, distance information	Yes	20
AEBUC	Competition	Yes	25~28

5.1 Parameter Setting

The parameters used in the simulation for the four algorithms are listed in Table 2.

Table 2 Parameter value

Parameter	Value
Monitor area	(100×100)
Node number	400
BS Location	(50,200)
Initial energy	0.5J
E_{elec}	50 nJ/bit
$\varepsilon_{\mathrm{fs}}$	10 pJ/bit/m^2
$\varepsilon_{\mathrm{mp}}$	0.0013 pJ/bit/m^4

Parameter	Value
d_0	87m
E_{DA}	5 nJ/bit/signal
Packet size	400bits

The values of parameters *T* and *c* used in AEBUC are obtained through a large number of experiments with the parameters set as Tab.2. The optimal value of *T* and *c* can be obtained when the network lifetime reach the maximum. The number of CCHs will increase with *T* increasing .The outcome by varying of the value of *T* depicted in Fig. 2 shows that the network lifespan reached the maximum value when *T* is set to 0.2. The cost for CH competition is added as the number of CCHs increases when the value of *T* is extended. However, the number of CCHs decreases and becomes insufficient for CH competition when the value of *T* is diminished to 0.1, which also decreases the network lifespan.

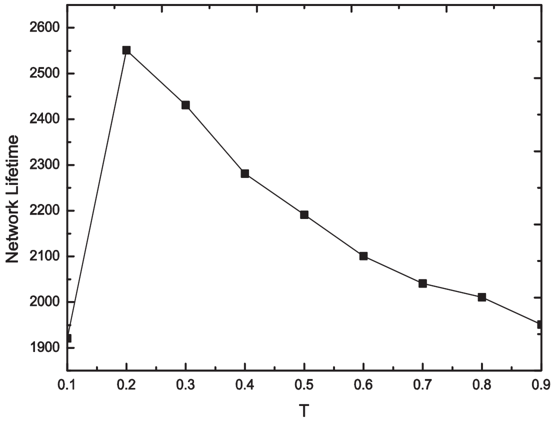


Fig.2 Experiment of Parameter *T*

The parameter *c* impacts on the transmission route selection. The setting of parameter *c* is also tested in Fig. 3 at the same time. The network lifespan reaches its maximum value when *c* is set to 0.9. Cluster members only consider the distance information to the BS for selecting CH when *c* is set to 0, which results in a large number of cluster members in the cluster near the BS. Such CHs will exhaust their energy fast. In contrast, it considers the distance information to CH only when *c* is set to 1.0, which merely results in a large number of cluster members in some cluster. Thus, the network lifespan is larger when *c* is 1.0 than when *c* is 0.

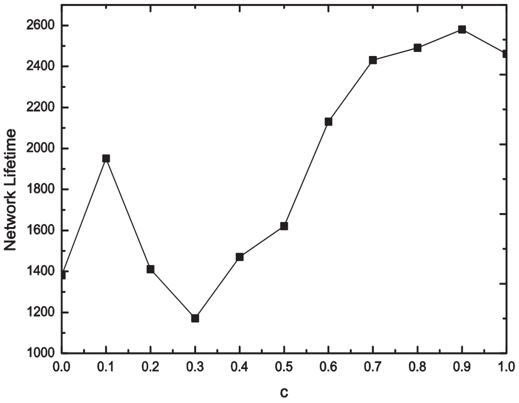


Fig.3 Experiment of Parameter *c*

5.2 Performance comparison

Energy efficiency is an important matric to evaluate the algorithm performance. Several metrics are defined below to accomplish the performance comparison among four algorithms.

(1) Network Lifespan (NL). NL is considered as the round when the first node is dead in this study. The later the death of the first node appears the better the performance of energy consumption balance.

(2) Average Residual Energy (ARE). ARE reveals the average residual energy of nodes which can be expressed as

$$ARE = \frac{total\ residual\ energy\ of\ nodes}{number\ of\ nodes} \tag{13}$$

(3) Energy Balance Factor (EBF). EBF is used to measure the energy consumption balance of nodes that is defined as standard deviation of the residual energy of nodes.

$$EBF = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(E_i - E_{avg})^2} \tag{14}$$

Where E_i is residual energy of nodes, and E_{avg} is average of residual energy of nodes. The performance of energy consumption balance is better when the value of EBF is smaller.

First, the ARE of all CHs are compared among the four algorithms in Fig.4. AEBUC clearly performs better than other three algorithms because an unequal cluster size is enabled, which provides the CHs far from the BS with smaller cluster members than that near the BS. This scenario achieves the energy consumption balance in AEBUC. I-LEACH considers both energy information and distance information, which makes the location of the selected CH superior to that in EECS and LEACH.

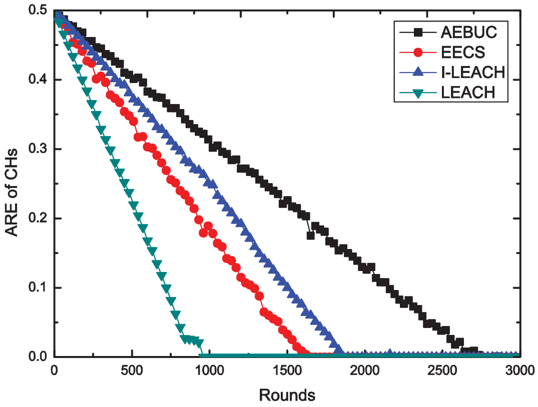


Fig.4 Average residual energy of cluster heads

Figure 5 shows the performance of energy consumption balance of the four algorithms. Fluctuation is shown in each algorithm, but it is clear in EECS and weak in AEBUC. The performance of another two algorithms is between them. Although load balance is considered in EECS, it focuses on the information of the distance to the BS to balance the energy consumption of CH when cluster members compute the cost to select CH. This condition increases the difference of cluster member energy consumption. Because many cluster members will join in the CH far from itself. The cost is computed in the opposite way in AEBUC that the information of the distance to CH is more important. An unequal CH distribution that in the region far from the BS distributes more CHs than the region near the BS also contributes to the energy consumption balance. The CH rotation mechanism is adopted in LEACH and I-LEACH, whose fluctuation is less than EECS.

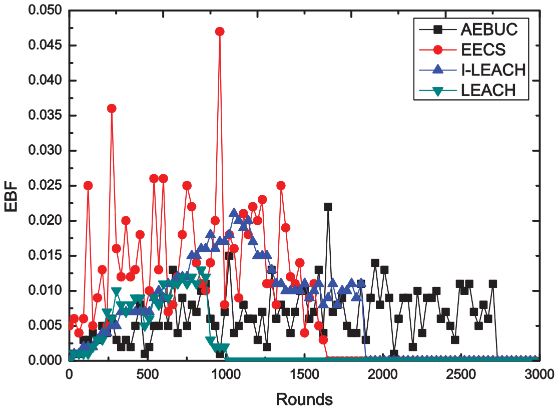


Fig.5 Energy Balance Factor

The number of CHs is statistically illustrated in Fig.6. The downtrend is shown in four algorithms because the number of dead nodes increases gradually. Figure 6 also clearly shows that the number of CHs in I-LEACH is larger than that in the three other algorithms because the threshold used in I-LEACH for CH selection changes for each node. The number of CH has no significant difference in EECS and AEBUC. These algorithms also exhibit slight variation trends.

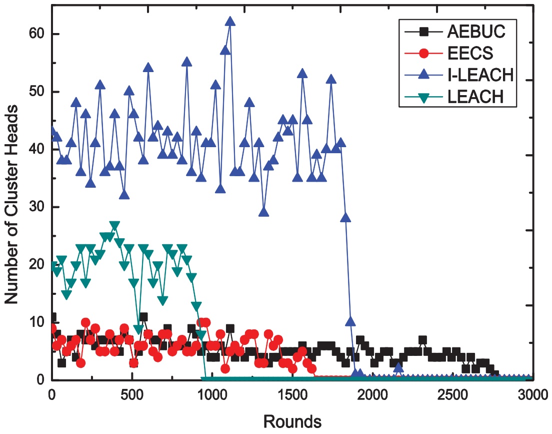


Fig.6 Number of Cluster Heads

Finally, NL is observed in Fig.7. AEBUC clearly prolongs NL effectively. Compared to LEACH, EECS, I-LEACH, AEBUC delays the death of the first node by up to 66.7%, 51.8% and 33.1%, respectively. The transmission distance has a large effect on the transmission energy consumption under single-hop transmission in cluster-based algorithms as shown in Eq. (2). An unequal cluster size generates more CHs in the region far from the BS than that near the BS. Therefore, AEBUC has better performance on energy consumption balance than the three other algorithms, which attributes to extend the network lifespan.

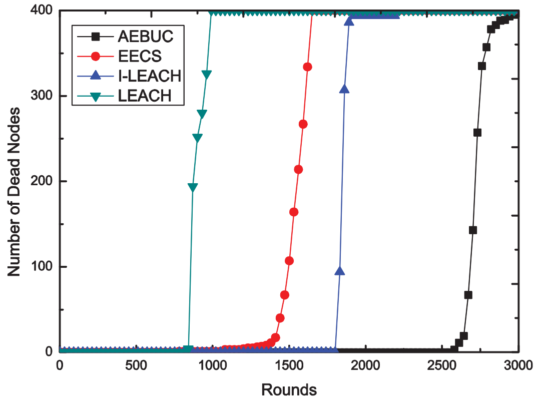


Fig.7 Number of Dead Nodes

6 Conclusion

WSNs have been developed and applied in diverse fields for many years. Solving the issue of unbalanced energy consumption is always a hot topic. This study presents a new solution called AEBUC that partitions the network into unequal sizes to balance energy consumption and prolongs the NL effectively. The problem is depicted by Eq. (10), based on which the maximum and minimum cluster sizes can be obtained, furthermore, the optimal cluster size of each node can be obtained. The final CHs are selected after the CH competition process. The network partition is finished when each normal node determines its ideal CH. We simulate AEBUC and three other algorithms with MATLAB. The simulation results show that our algorithm prolongs the lifespan of network by up to 66.7%, 51.8% and 33.1% when comparing to LEACH, EECS and I-LEACH, which also decreases the difference of the energy consumption of sensor nodes.

Acknowledgements

This work was supported in part by National Nature Science Foundation under Grant 61073164 and 61373123. The authors express sincere appreciation to the editors and the anonymous reviewers for their helpful comments.

7 References

[1] D.S. Ghataoura, J.E. Mitchell, G.E. Matich, Networking and Application Interface Technology for Wireless Sensor Network Surveillance and Monitoring, IEEE Communication Magazine, vol. 49, Issue 10, pp. 90-97, Oct. 2011.

[2] Edison Pignaton de Freitas, Tales Heimfarth, etc. Cooperation among Wirelessly Connected Static and Mobile Sensor Nodes for Surveillance Applications, Sensors, vol. 13, Issue 10, 12903- 12928, Sep. 2013.

[3] Wang, Xiaonan; Le, Deguang; Cheng, Hongbin; Xie, Conghua. All-IP wireless sensor networks for real-time patient monitoring. Journal of biomedical informatics, vol. 52, pp. 406-17, Dec. 2014.

[4] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan. An application-specific protocol architecture for wireless microsensor networks, IEEE Trans. Wireless Commun. vol. 1, Issue 4, pp. 660–670, Oct. 2002.

[5] Aimin Wang, Dailiang Yang, Dayang Sun. A clustering algorithm based on energy information and cluster heads expectation for wireless sensor networks, Computers and Electrical Engineering, vol. 38, pp. 662-671, May, 2012.

[6] Navid Amini, Alireza Vahdatpour, Wenyao Xu, Mario Gerla, Majid Sarrafzadeh. Cluster size optimization in sensor networks with decentralized cluster-based protocols, Computer Communications, vol. 35, pp. 207–220, Jan, 2012.

[7] Zahra Beiranvand, Ahmad Patooghy, Mahdi Fazeli. I-LEACH: An Efficient Routing Algorithm to Improve Performance & to Reduce Energy Consumption in Wireless Sensor Networks, 2013 5th Conference on Information and Knowledge Technology, pp.13-18, May, 2013.

[8] Mao YE, Chengfa LI, Guihai CHEN, and Jie WU. An Energy Efficient Clustering Scheme in Wireless Sensor Networks, Ad Hoc & Sensor Wireless Networks, vol. 3, pp. 99-119, Apr. 2006.

[9] Hill, J., Szewczyk, R., Woo, A., Hollar, S., Culler, D. E. and Pister, K. S. J. System Architecture Directions for Networked Sensor. International Conference on Architectural support for Programming Languages and Operating Systems, pp. 93–104, Nov. 2000.

[10] Huan Li, Yanlei Liu , Weifeng Chen , Weijia Jia , Bing Li , Junwu Xiong, COCA: Constructing optimal clustering architecture to maximize sensor network lifetime. Computer Communications, vol. 36, pp. 256-268, Feb. 2013.

[11] Hakan Bagci, Adnan Yazici, An energy aware fuzzy approach to unequal clustering in wireless sensor networks, Applied Soft Computing, vol. 13, pp. 1741-1749, Apr. 2013.

[12] Wendi Beth Heinzelman (2000), Application-Specific Protocol Architectures for Wireless Networks, PhD thesis, Massachusetts Institute of Technology, June, 2000.

SESSION

SECURITY, PRIVACY, AND RELATED SYSTEMS AND ALGORITHMS

Chair(s)

TBA

Private Data Protection in a Continuous Nearest Neighbor Query

Charles Asanya and Ratan Guha
Department of Electrical Engineering
and Computer Science
University of Central Florida
Orlando, Florida 32816

***Abstract*—With the help of location-aware mobile devices users can issue query and receive nearest point of interests from a location-based services provider. User private information is needed to personalize the service. If this information is compromised user's privacy can be exposed. In this paper, we propose idea to protect user private information in location-based services continuous nearest neighbor query focusing on moving query moving object. Some existing solutions works like snapshot query and therefore inefficient. Our proposal combines Voronoi tessellation and Hilbert curve order to create nearest neighbor relationship. We treat object location as a function of time and introduce transition and update time. We implement a database that scales to object size, and then execute double private information retrieval protocol on the server to periodically return to the user the exact nearest point of interest through a path with acceptable performance without revealing any user private information. Our complexity analysis and experimental evaluation of the network transmission time using ns-3 simulator show improvement over previous technique.**

***Keywords*—Voronoi Cell; Hilbert Order; Transition Time; Update Time; Quadratic Residuousity Assumption.**

I. INTRODUCTION

Location-based services (LBS) allow mobile device users to receive the nearest point of interest (POI) to their location within a spatial network. With the help of geographical information of the mobile device LBS providers are able to provide location information to subscribers. For instance; store locator application allow users to quickly request and find nearest store location with the help of location intelligence [1]. To effectively provide and customize the service, LBS providers need the location and data profile of the user. However, users will be reluctant to use the service if they believe that doing so will expose their private information. Consider a situation where Alice is searching for a specialized treatment health center. To correctly answer her query Alice has to disclose her location and information desired. This disclosure can reveal Alice's ailment, which may cause her embarrassment, unwanted attention or reprisal. Alice's query can be answered in of two ways known as snapshot nearest neighbor query (SNNQ) or continuous nearest neighbor query (CNNQ). SNNQ executes one time response to a user query and does not update user as user or the POI location changes, while CNNQ allows LBS server to continuously update the user with the nearest POI as the user and POI changes their positions.

Several techniques have been proposed to prevent user private information from been disclosed or exploited by adversaries like proposals of [2] and [3]. However, they only addressed snapshot query, and does not support moving query. As stated in [4] and [5], it is inefficient and infeasible for a moving object to issue query at every point of the line segment as in snapshot query, therefore solution that works like snapshot will be inefficient. Several of the techniques also rely on the honesty and integrity of the server or third party. In [6], [7], [8] and [9], techniques were proposed for a moving query in LBS, but they did not consider security and privacy of the user

In this work, we address privacy issue related to CNNQ. This type of query can be divided into; (1) moving query static object (MQSO); (2) moving query moving object (MQMO); and (3) static object moving query (SOMQ). The focus of our work is on MQMO which is a query where user and the POI locations are changing. Fig. 1, shows MQMO where user u in location l_1 has p_1 as the nearest POI. When u location changes to l_3, the nearest POI is still p_1 due to simultaneous movement of both objects. Private CNNQ shall be able upon a single query request anonymously update u with the nearest p as u or p location changes until query is specifically terminated or u goes outside LBS coverage area.

Our technique combines Voronoi Diagram with Hilbert Curve order to isolate objects. It provides transition and update time that periodically allows database to return only a single nearest POI to a user. To eliminate third party or the reliability on the server honesty we employed cryptographic technique as in [2], [10] and [11] using private information retrieval (PIR) protocol based on the intractability of a large prime which allows user to retrieve information from a database without revealing the exact information retrieved. The database D as in Fig. 2 is of size n and is represented as a square matrix of m bit string. If Alice wants to retrieve the value represented by bit D_i; to preserve privacy, she sends an encrypted query $E(D(q(i))$ together with transition time, where E is used for encryption. The server at each transition time responds with a value $r(E(D(q(i))))$ as in Fig. 3, which allows Alice to compute D_i.

Our contributions to private MQMO-CNNQ are as follows:

First, we propose plane partitioning and continuous fractal space filling curve based on Voronoi diagram and Hilbert curve order to create nearest neighbor relationship through a path.

Second, we propose to treat object location as a function of time to create pre-determined transition and update time where nearest neighbor changes.

Third, we provide experimental evaluation with respect to complexity and network transmission time using ns-3 simulator to show a system with promising performance that can scale for different environments.

The rest of the paper is organized as follows; Section II presents the preliminaries. Section III discusses related work, while section IV presents the design framework. The implementation is presented in section V. Analysis and Results are discussed in section VI, while conclusion and future work are stated in section VII.

II. PRELIMINARIES

Continuous nearest neighbor is defined as the nearest POI in the path of a moving using at each point of a segment as the user moves along the segment. In this paper, we consider moving query

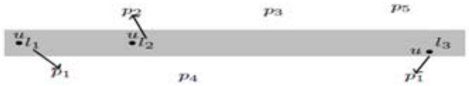


Fig. 1. Moving Query Moving Object CNNQ

(user) moving object (POI). Security and privacy in CNNQ demands that the private information from a user query as it moves within a spatio-temporal network where differences exists between data object distances be not disclosed to any party as user continuously receives update on its nearest POI.

A. Modelling Moving User and POI Locations

Our model needs to account for the objects location at each time interval. A spatial network (SN) consists of user and POI objects. Assuming they are statically or dynamically restricted within the edges of the SN. If v is the velocity of a moving object at a location $l \in SN$, then the object (user) position at time tick T_i can be represented by;

$$l_u(T_i) = f(v * t_{i-1}) \tag{1}$$

Consequently object (POI) location can also be represented by;

$$l_p(T_i) = f(v * t_{i-1}) \tag{2}$$

where i represents the $i - th$ location of user u and POI p at time tick T_i, while T_{i-1} is the previous location.

B. Problem Definition

The problem statement in a MQMO-CNNQ is such that user u and POI p located in location l restricted to area A at time T and individually moving with its own velocity v is at location l' at a future time T'. If p represents the nearest POI to u in $l \in A$ at time T, then as the objects moves form $l \rightarrow l'$ at time $T \rightarrow T'$, private MQMO-CNNQ shall continuously return to u the static or dynamic $p \in A$ whose distance (d) from $u \leq \forall p \in A$ as u or p moves within the spatio-temporal network until query termination without disclosing l, l' or p to another party.

C. Solution Overview

The strategy here is to create nearest neighbor relationship between the POIs in a path by utilizing Voronoi diagram and Hilbert curve order. The distances between the POIs are averaged and represented as time intervals. Location of the objects within the spatial network is treated as a function of time within the time interval, and in between this time interval is the point where nearest neighbor will change. This will allow the server to continuously update user with the nearest POI within the path with only the initial query while still protecting user privacy.

We partition an area with Voronoi diagram as in [10] using the set of POIs, and then super-impose regular grid of size $\sqrt{n}$ x $\sqrt{n}$ over the diagram. For every cell c in the grid determine all Voronoi cell intersecting it, then bind the corresponding POI to c. Map all the points in the bounded cell to Hilbert curve order as in [2]. Assign Hilbert value to all the points in c, and transform these values which represent set of points in a multi-dimensional space into records in a database. Determine the transition and update time interval where the nearest neighbor will change using user velocity and elapsed time.

After initialization, user sends the query message together with the transition time to the server. Execute PIR protocol on user request and then create array of size k (k is the number of POI in a user path of travel) in ascending order, and have the pointer point to the first element. Update the array at every update time interval and shift the pointer by one. At each transition time interval send to the user the element pointed to by the pointer.

III. RELATED WORK

Several solutions including [12] have been suggested to prevent query from divulging user private information. The proposal partitions area into cells using Voronoi graph. The contents of each cell form cooperative group with centered common server (CS). User sends request to CS. The CS takes its own location and a neighbor CS location that the user is heading to as two anchor point and chooses other continuous anchors in the line segment between them. The CS organizes users to form cooperative k-anonymity group according to their moving trend and sends request to LBS removing user identity and actual location. Then CS sends each snapshot query of continuous query request with an anchor which the user has not yet passed. The problem with this technique is that it issues unnecessary query at each segment of the anchor point. It also depends on third party which may not be trustworthy and can provide single point of failure. It also depends on other users having similar moving pattern.

A design that issue query based on pre-determined uniformly distributed static point was proposed in [13]. It finds the vertex of a cloaked area; it then finds the nearest static point to the vertex and then finds the nearest object to the static query point as an approximate answer for the vertex. It repeats the same process when the nearest neighbor changes. This technique works like snapshot query and therefore inefficient. The query response also will be approximate and not exact nearest neighbor. Also, privacy depends on third party anonymizer. Chow and Mokbel in [14] propose k-sharing region. A cloaked spatial region contains at least k users, and the region is also shared by at least k of these users. Query is issued based on the cumulative cloaked region which will stop adversary from linking query to a specific user. However, it does not say how a user continuously receives nearest neighbor from LBS. It will also be difficult to maintain exactly k users in the entire cloaked region; therefore query may never be answered.

We propose idea that combines Voronoi diagram and Hilbert curve order to create nearest neighbor relationship along a path. We then create transition and update time where nearest neighbor changes to eliminate the need for continuously issuing query. Server runs PIR protocol of [15] on user request and creates array data structure. At each update and transition time interval server updates the array and pops the element pointed to by the pointer and sends to the user.

IV. FINDING DISTANCES BETWEEN NEAREST NEIGHBOR

To find distances between moving objects are challenging. A brute force approach can be used to calculate the distances. This approach is effective but inefficient. Work in [9] modelled distance as directional weighted graph $G(V, E)$, where V is the vertex, and E is a single directed edge. If E is the empty set that contains no vertex and R is the set for all the single directed edges, then all elements $E \in R$ corresponds to a single directed edge. If we assign the edges weight $w(v_i, v_{i+1})$, then the length of a path $P \forall E \in R$ is;

$$P = \sum w(v_i, v_{i+1}), \tag{3}$$

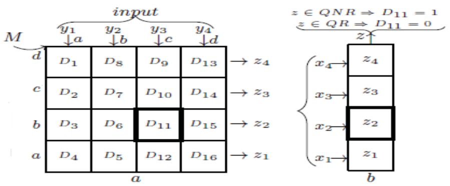


Fig. 2. Matrix representation of database

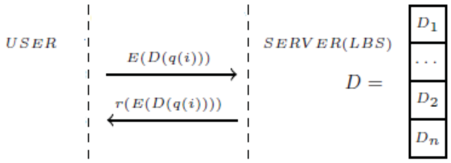


Fig. 3. User encrypted request and server encrypted response

We use this approach to model distances between objects. We represent the vertices as the nodes (POIs), and the edges as the distances between the POIs. The distance between two POIs is the number of edges in a shortest path connecting them. The shortest distance $d(n_i, n_j)$ between two nodes (POI) is therefore defined as min w(P) for all paths between n_i, n_j, where n_i and n_j are the $i-$th and $j-$th node respectively. If a user is located between path n_i, n_j, then, the nearest POI to the u is the POI with minimum w(P) from the u. The w(P) is treated as time interval between two POIs.

V. Indexing Continuous Nearest Neighbor Query

The result of the query through P contains a set $< S_i, T_i >$ tuples as described in [16], where S_i is the ordered list $(p_1, p_2, \cdots, p_k)$, and k is the number of objects in the path. T_i represents the time interval during which S_i is the nearest POI to user u through the path. POIs p are ordered by their increasing distance to u, i.e.,p_i is the closest POI to u, while p_{i+1} is the next closest POI to u, and p_k is the $k-$th POI closest to u in the interval T_i, T_{i+1} and T_k respectively. The returned sets through P in the interval or segment $\bar{se}$ satisfy the conditions; $\cup_i T_i = [t_s, t_e]$ and $T_i \cap T_j = \emptyset, \forall i \neq j$ where $[t_s, t_e]$ is the query interval divided into non-intersecting sub-intervals n. $Ti = [t_s, t_1]$, and $T_n = [t_{n-1}, t_e]$ for $(t_s < t_1 < t_{n-1} < t_e)$. The time stamps $(t_1, t_2, \cdots, t_{n-1})$ are known as the split or the transition time interval where the nearest POI to u will change.

A. Private CNNQ Framework

We use Voronoi diagram and Hilbert curve order that scales and align with database to isolate objects and then run PIR protocol to secretly return a single nearest object to a user at each transition time interval.

1) Voronoi Diagram: As shown in Fig. 4,Voronoi diagram allows space to be divided into regions [17] and [18]. It uses the nearest-neighbor rule. We implemented a Network Voronoi Diagram (NVD) described in [17]. Aurenhammer [19] defined Voronoi diagram in terms of dominance of two distinct points. We modelled the network as weighted graphs where the intersections are represented by nodes of the graph, and roads are represented by the links connecting the nodes. The weights are the distances of the nodes that represent the time it takes to travel between the nodes assuming a constant velocity.

2) Hilbert Space Curve: It is a space filling curve that preserve spatial proximity [20]. Hilbert space partition is described in [21]. As shown in Fig. 5, Hilbert curve visits every point in n-dimensional grid space exactly once without crossing itself [5]. It is used for mapping multi-dimensional space to one-dimensional space while still preserving locality [2]. That is, if two POIs are close in the 2-D space, they are likely to be close in the Hilbert ordering as well; therefore, nearest neighbor to a user is the point of interest whose Hilbert value is closest to that of the user.

3) Private Information Retrieval(PIR) Protocol: PIR is a protocol that allows a query to secretly retrieve nearest neighbor. We implemented the flavor introduced in [15], and use it to retrieve user requested information. It is based on the premise that a function bounded in a polynomial time is not able to allow database of size n to differentiate between a query for the $i-th$ bit and a query for the $j-th$ bit $\forall(1 \leq i, j \leq n)$. PIR relies on Quadratic Residuosity Assumption used in [22], which states that, it is computational hard to find the quadratic residue in modulo arithmetic of a large composite number $N = q * q'$, where q and q' are large primes. If N is a natural number, Define

$$\mathbb{Z}_{\mathbb{N}}^{*} = \{x | 1 \leq x \leq N, gcd(N, x) = 1\} \tag{4}$$

Let the quadratic residuosity predicate be defined as $Q_N(y) = 0$ if $\exists x \in \mathbb{Z}_{\mathbb{N}}^{*}$ such that $x^2 = y \mod N$ and $Q_N(y) = 1$ otherwise. If $Q_N(y) = 0$ (i.e. y is a $square$ y), then y is said to be quadratic residue (QR), and if $Q_N(y) = 1$ (i.e. y is a $non-square$ y), then y is said to be quadratic non-residue (QNR).

If

$$\mathbb{Z}_N^{+1} = \{y \in \mathbb{Z}_{\mathbb{N}}^{*} | (\frac{y}{N}) = 1\}, \tag{5}$$

is true, then half of the numbers in $\mathbb{Z}_N^{+1}$ are $\in$ QR, and half are $\in$ QNR. If q and q' are large enough $\frac{k}{2}$ bit prime, for every constant c and a family of computational bounded polynomial circuit $C_{k_0}(y)$ there exist an integer k_0 such that $\forall K > k_0$

$$Pr_{y \in \mathbb{Z}_N^{+1}}[C_{k_0}(N, y) = Q_N(y)] < \frac{1}{2} + \frac{1}{k^c} \tag{6}$$

If (6) holds, and for large enough k_0, the probability of differentiating between a QR and QNR is not better than guessing. Server will not be able to determine if $y \in QR$ or $y \in QNR$

4) structure of the Database: Similar to [10] and [11], space partitioning is aligned with the database structure in order to provide the desired privacy. The database is of size n and arranged as a square matrix $M = \sqrt{n}$ x $\sqrt{n}$ indexed by X_i, where $i = 1, \cdots, n$. The space is mapped into grid G with cells c. Data objects representing POIs are contained in the space. Each $c \in G$ enclose set of POI. Each entry in the database indexed by X_i represent the POI in c. Each X_i can contain multiple POIs indexed by X_{i_j}. All objects have equal number of bits; otherwise it may be exploited to expose user identity or interest.

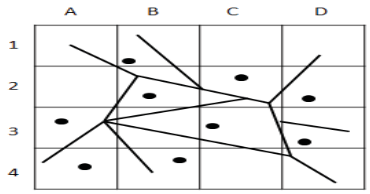


Fig. 4. Grid imposed on Voronoi Tessellation with 9 POI

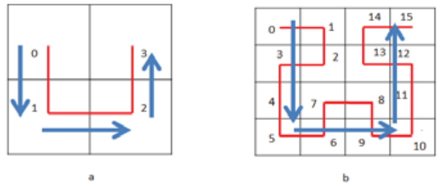


Fig. 5. Hilbert curve of (a) order 1 and (b) order 2

VI. IMPLEMENTATION

The steps for private MQMO-CNNQ of our design are threefold. We assume that POIs of each segment are pre-determined.

First, server super-impose Voronoi diagram over area, and then map the area to a $\sqrt{n}$ x $\sqrt{n}$ grid. All the POI in the area belongs to a cell. The contents of the cell represents the two endpoints of a query line segment $\bar{se}$, where $\bar{se}$ contains $p_{1\to k}$ (k is number of points in $\bar{se}$) that satisfies dominance of [19]. Each c now contains k POIs which is the answer set returned after all the points that intersect a Voronoi cell are bounded to c.

Second, all the points $p_i \in X$ for $i \in [1, \cdots, k]$ in the interval I are mapped to Hilbert Curve order as in [2] and partially in [23] such that a one to one correspondence between the subintervals of I and sub-squares of Q satisfy adjacency and nesting conditions which determines a continuous function w_f that maps I onto Q [21]. All the points are assigned Hilbert values. For a path with 9 POI; c will contain 9 POIs, and all POIs are assigned Hilbert value H based on their distance to one another. For example, points p_1, p_2, p_3, p_4 have H values of 50, 150, 200, and 85 respectively. Points with closer values are closer in space.

Third, we transform these values which represent set of points in a multi-dimensional space into records in a database akin to [24]. Queries on the records are queries on these sets of points which are now represented by Hilbert values. We use B+tree structure for the geometric data storage. Fig. 6, depicts B+ tree for k=4. Each key value is greater or equal to the values in the left node which are Hilbert values. The index key is the maximum Hilbert value in the segment concatenated with the cell id, so that when a user identifies its cell, the user can then use the value to determine which query set belongs to its path. The cells are identified by the cell id, and each cell contains k POI in $\bar{se}$ representing the path of travel by a user. The cells are padded if necessary to prevent server inference from number of k transferred. Stored with the POIs are the weights w which represent the time interval it takes to travel between adjacent POIs known as transition time t_t. We also partition the time axis into different time duration, where each time tick is the update time t_u, i.e., the time the server updates database information. At each $t_u \approx t_t$, server computes l'_0. If there is no change in velocity, there will be no update from user. Server will use the stored value to determine object location.

A. Processing Query

The following steps will be followed for a user searching for the nearest POI:

Step 1;Server creates nearest neighbor record following the procedure described above and then builds B+ tree geometric database storage of the POIs.

Step 2; Initialization: User in location l initiates a query to obtain nearest POI along a path of interest. Server sends the grid, the key (Maximum Hilbert value), k (number of POI), and w represented by the average time interval t_{avg} for $(t_s < t_1 < t_{k-1} < t_e)$ for each $\bar{se}$.

Note: It is assumed that the distance between POI and t_{avg} can be averaged such that t_{avg} represents the transition time for all p from $p_1, p_2, \cdots, p_k$ for a known constant velocity v. If d_i is the distance where the nearest neighbor changes from p_i to p_j for $i \le j \le k$, server computes average time for a query object travelling at constant velocity v as

$$t_{avg} = \frac{\sum_i^{k-1} \frac{d_i}{k-1}}{v} \tag{7}$$

where k is the number of objects in the path.

Step 3; Requesting POI: User receives the grid together with the parameters and identifies the cell it belongs to. Focusing on that cell, user use its velocity to find the average distance d_{avg} where the nearest neighbor change will occur. The transition time t_t at any velocity v will then be calculated by the user as $t_t = \frac{d_{avg}}{v}$ as shown in algorithm 1 . Using the key, user then request all the POI contained in the set left of the key. For instance, if key of interest to a user is 60, it will then request the values 40, 50 and 60, Fig. 6. With a database s = $\sqrt{n}$, user randomly generates modulus $N = q * q'$ with a query message $y = ([y_1, y_2, \cdots, y_s]^k)$ and $x = ([x_1, x_2, \cdots, x_s]^k)$ each targeting the $k-th$ element in the segment $\bar{se}$ such that $y_b \in QNR$, while $\forall j \neq b, y_j \in QR$ and $x_a \in QNR$, while $\forall r \neq a, x_r \in QR$. User then sends the query message together with the t_t to the server. Let the matrix column multiplication be represented as,

$$z_r = \prod_{j=1}^{s} w_{r,j} \tag{8}$$

and row as,

$$Z_\alpha = z_{r_{r,j}} \tag{9}$$

where output z_r is masked by input x_r to produce Z_α. $w_{r,j} = y_j^2$ if $M_{r,j} = 0$, otherwise if $w_{r,j} = y_j$, $M_{r,j} = 1$ for all $j = 1 \to s$. Also, $z_{r_{r,j}} = x_r^2$ if $M_{r,j} = 0$, else $z_{r_{r,j}} = x_r$ if $M_{r,j} = 1$ for $r = 1 \to s$.

Step 4; Using Algorithm 2 server runs PIR protocol on the user request, and computes Z_α using equations (8) and (9). Server then creates array of size k as shown in Fig. 6. Initially, the pointer points to j=0. At each t_t interval server sends to the user the value at j=0. Upon receipt of a nearest neighbor, user determines the validity of the nearest neighbor by using its position and a flag with a value of true or false, i.e. server updates the user with the nearest neighbor at the user prescribed interval, in order for it to be the nearest neighbor user must have reached the transition point which will result in setting the flag to false, otherwise if the flag is true (meaning user has not reached the transition point and therefore, previous POI is still valid) user caches the nearest neighbor without decrypting it and use the previous nearest neighbor since it has not reached the transition point where the nearest neighbor will change. The flag will ensure that any deviation by the user from the expected travel pattern will not result in false positive. User computes equation (10) to find nearest POI.

$$\left(Z_\alpha^{\frac{q-1}{2}} = 1 \; mod \; q_1\right) \wedge \left(Z_\alpha^{\frac{q'-1}{2}} = 1 \; mod \; q_1\right) \tag{10}$$

If (10) is true then $Z_\alpha \in QR$, else $Z_\alpha \in QNR$. At each subsequent $t_u \approx t_t$, server updates the database and re-creates the array with current order of object location for $\bar{se}$. It will again in ascending order create array of Z_a of size k and shifts the pointer to j+1, and then send the value at j+1 to the user at the corresponding t_t. The process continues until the last k is sent or termination of the query.

Algorithm 1: (User creating query message)

```
Input: grid $G$, $k$, $keys$, $t_{avg}$
Output: $t_t$, $Y_{1\to k}$, $X_{1\to k}$
//output transition time and query messages with modulus $Y_k$ and $X_k$
Procedure:
for each $c \in G$
    if $u \in c$
        $d_{avg} = t_{avg}$ x $v$;
        $t_t = \frac{d_{avg}}{v}$; //$t_{avg}$ and $v$ is the average transition time of the user cell and user velocity respectively.
generate $Y_{1\to k}$ and $X_{1\to k}$;
return $t_t$, $Y_{1\to k}$ and $X_{1\to k}$;
```

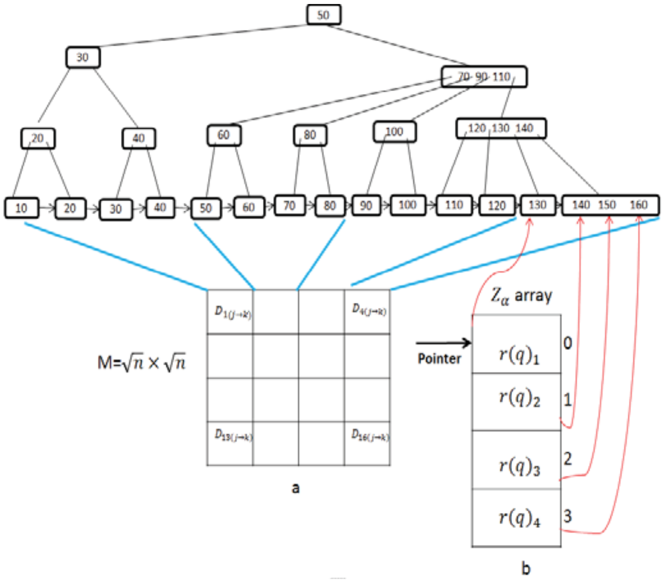


Fig. 6. B+ -tree representation of database matrix and Z_α array for k=4 with pointer at $Z_\alpha[3]$.

VII. Analysis and Results

We present the results of our technique with respect to network transmission time and complexity. We evaluate the transmission time using different k and n values for both user and database server. We then implement the design of [12] and compare the network response time with our technique.

Algorithm 2: (Server responding to user query)

```
Input: $n, R_{tree}, t_t, Y_{1\to k}$ and $X_{1\to k}$
//$n$ is the number of segments when mapped into grid.
//$R_{tree}$ is the data structure representation of the database
//$t_t$ is the user transmission interval
//$Y_{1\to k}$ and $X_{1\to k}$ is the user message and encryption modulus
Output: $arrayZ_\alpha[k]$
//$k$ is the number of POI in user route
Procedure:
for $i = 1 \to \sqrt{n}$
    for $j = 1 \to \sqrt{n}$
        for $s = 1 \to k$
            do $Z_{ijs} = \prod w_{ijs}$; //encrypt row and column
for $i = 1 \to k$
    do $arrayZ_\alpha = Z_{ijs}$
for $i = 1 \to k$
    if $time = t_t$
        return $arrayZ_\alpha[i]$;
        $i++$;
```

A. Experimental Setup

The algorithm was developed in C++ with little modification to ns-3 python script. The server and client experiment was conducted on ns-3 simulator running on Ubuntu 9.04 Linux variant operating system with intel-quad-core processor operating at 2.66GHz with 256MB RAM. The server communication with the access point runs a CSMA protocol with data rate of 100Mbps and 6500ns channel delay. Client communication is Wi-Fi 802.11a standard with a stream data rate of 54Mbps. The payload is 1450 bytes. The Access Point (AP) shares the same physical level and channel attributes as the client Wi-Fi device which we default to ns-3. The AP is stationary and the client is mobile and randomly wonders around at a random speed in a bounded box defined by x and y coordinates. We varied the grid size from 16 up to 225. For the objects, we use Sequoia points in [25], which contains 62,556 California place names. We use query size of 100 Bytes and object size of 10kB. The packet transmission interval was set to one second. We use 768 bits for the encryption modulus.

B. Transmission Time

First, we present the transmission time for different values of k as n changes from 16 upwards to 225, and then analyze the cost impact of n. In Fig. 7, and Fig. 8, we show the time for server to transmit the entire packet for different values of k and n. It shows no significant difference for different values of k when n is small. At k = 5, we saw proportional increase in transmission time; however as n gets large we observe an increase in the proportionality of the time increase compare to lower n. In Fig. 8, we consider the transmission cost for three different values of k and n to determine the impact of k. We made the following observations; First, for small n = 16, the impact to transmission time is less significant with increase in k. Second, for an increase in n = 64, k takes a sharp increase in the transmission time at k = 10. Third, for an increase in n = 144, transmission time increase is very minimal. What this means is that in the first observation, for area of high concentration of k, large number of k can be accommodated in a grid to make n small with little or no impact on the transmission cost.

For the second observation, when n increases up to 64, $k > 10$ will have significant increase in transmission cost. Therefore it will be more desirable to make n smaller, for instance n = 16 and k >10 will be more desirable. In the third observation, for larger n = 144, the proportionality of the increase is less for $k > 10$ than for $k <$ 10. This means that the design will scale well for large area with large concentration of k. Fig. 9 and Fig. 10 show client transmission time for different k and n. The result indicates negligible increase in transmission time for different k when n is small. We saw large increase at n =144 and greater. The query response time results shown in table I indicates an increase with increase in length of modulus. This offers privacy and efficiency tradeoff opportunity for a user. For instance, a user searching for nearest gas station may have minimal privacy requirement and for efficiency opt for smaller modulus than a user looking for gambling center who may prioritize privacy over efficiency and therefore choose larger modulus.

C. Complexity

The server and client communication and computation complexity is analyzed and compared with [10] as shown in Table II. As already stated, the database is of size $\sqrt{n}$ x $\sqrt{n}$, and each cell content is $k * m$ bit long. The length of the modulus is P bits, and c(P) is the computation function for the P bits. Our scheme shows improvement over [10] in the user communication when $\sqrt{n} > k$ which is likely always the case. Our technique has downside in the server computation caused by the double PIR. Nevertheless, the reduction in the server transmission and user computation cost due to the double PIR makes up for the downside. Since $\sqrt{n}$ is always greater than k, our scheme will always perform better.

D. Comparison with Previous Technique

We conducted experiments to compare our design with the design of [12]. We did not consider offline activities, that is, events that happen prior to user declaring intention for the nearest neighbor; such as creating a square grid and super-imposing Voronoi diagram over

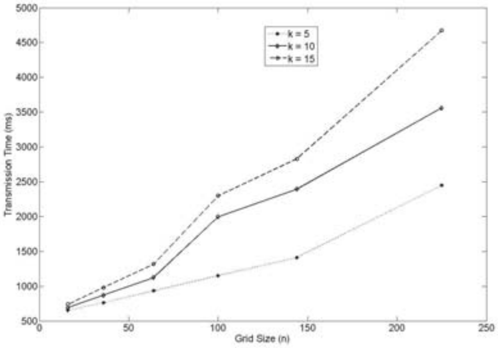


Fig. 7. Server Transmission time for three different k as n changes from 16 up to 225

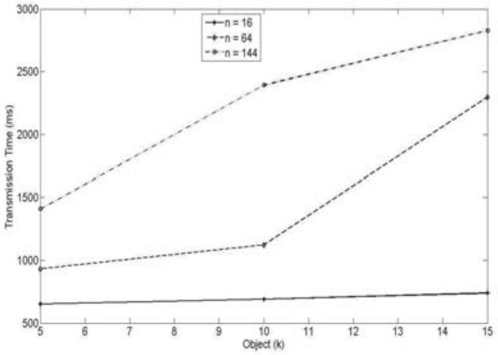


Fig. 8. Server Transmission time for k = 5, 10, 15 and n = 16, 64 and 144

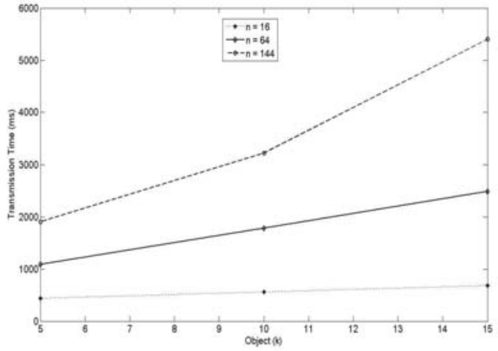


Fig. 9. FClient Transmission time with 96 Bytes Modulus for k = 5, 10, and 15

the grid and mapping all the POI to Hilbert value. In the case of [12], these activities include using Voronoi to divide road network into cell with centered Central Server (CS). We implemented both design in ns-3 simulator. We implemented [12] based on author description and as we understand it.

The design as described in [12] involves a three-way process. It partitions the entire area into many cells centered with different CSs using Voronoi graph. A user receives several broadcasting messages from neighboring CSs, and determines the cell she is heading to, and then computes the distance between each CS and herself, then register to CS with minimum distance from her. Registration information includes user location, destination and history velocity. After successful registration to a CS, a user will report her latest locations periodically to the CS and drop broadcasting message from it until the user gets out of the cell or registers to next cell. CS makes moving direction prediction for the user using user history velocity. CS then organizes k users to form a cooperative k-anonymity group according to their moving trends. It takes its own location and a neighbor CS location which the user is heading to as two elementary anchors and chooses other continuous anchors in the line segment between them. CS picks an anchor from anchor sequence to replace user's location and send a snapshot query of a continuous query request with k queries to LBS Provider (LSP). The LSP uses the anchor point to perform incremental nearest neighbor search (INN), and returns its results to the CS. The CS then refines the result and then returns k-POI to the user.

For comparison; we measure the response time (i.e., from initialization to the time user receives the POI). In both designs, we assume that a user is travelling through the same path with the POIs the same distance apart. We compare the time for the initial response and then subsequent responses until query terminates. For ns-3 client and server module we use same parameters described above. We compare [12] with three different lengths of modulus that we used for encryption in our design and show the results in table I. For the initial request, [12] performs better as the size of modulus increases. However, our design performs better throughout the life of the query as user continuously receives nearest POI. We believe that one of the contributing factors to the high cost of [12] is due to the fact that the design is basically a snapshot query rolled into continuous query. Similar operation is performed throughout the segment, while in our case, after initial operation server queues the encrypted k-POI, and sends a single POI at every transmission interval without any user action. Also, another downside is the k-anonymity requirement. There are chances that a query may never be answered if it did not satisfy k-anonymity. Also user privacy is at the mercy of the CSs.

VIII. CONCLUSION AND FUTURE WORK

We propose and evaluate a privacy protecting technique for location-based services moving query moving object continuous nearest neighbor query. We combine Voronoi tessellation and Hilbert curve to map object from a 2-d space to 1-d space without altering the proximity. The technique allows user to determine the interval at which to receive the next nearest neighbor from the server. It also allows user to authenticate server response. We conducted evaluation on the user and server transmission time, and the results show effective and efficient performance that can scale to environment. We also perform complexity analysis and compare with [10], the results for the most part also show improvement over [10]. Finally, we

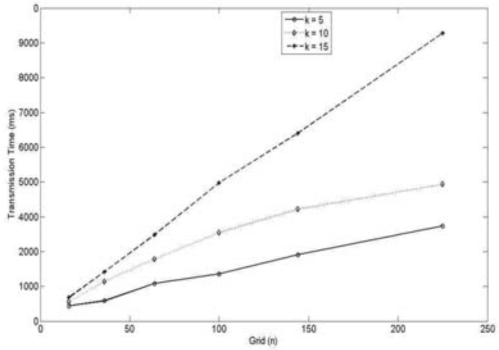


Fig. 10. Client Transmission time with 96 Bytes Modulus and n from 16 up to 225

TABLE I. INITIAL AND SUBSEQUENT RESPONSE TIME USING DIFFERENT MODULUS IN OUR WORK COMPARE WITH PREVIOUS TECHNIQUE

Size of k, n and N	Our Design, Initial (s)	Our Design, Subsequent (ms)	Previous Design, Initial (s)	Previous Design, Subsequent (s)
k=10, n=144, N=768bits	3.536	0246	4.283	4.283
k=10, n=144, N=1024bits	5.130	0360	4.283	4.283
k=10, n=144, N=2048bits	9.104	0448	4.283	4.283

TABLE II. COMMUNICATION AND COMPUTATION COST COMPLEXITY

Type	Previous Scheme	Our Scheme
User Communication	$O(Pm\sqrt{n})$	$O(2Pkm)$
User Computation	$O(c(P)m\sqrt{n})$	$O(c(P)m)k$
Server Communication	$O(Pm\sqrt{n})$	$O(Pkm)$
Server Computation	$O(c(P)km\sqrt{n})$	$O(2(c(P)km\sqrt{n}))$

conduct experiments on the response time of the design and compare with [12]. The results show performance improvement over [12].

For our future work, we hope to address the computational bottleneck involved in continuous nearest neighbor query by introducing parallel programming utilizing graphical processing unit (GPU) and compute device architecture (CUDA). We also hope to evaluate power consumption of the mobile device for our design.

References

[1] R. Goodrich. (2013, Oct.) Location-based services: definitions & examples.

[2] C. Asanya and R. Guha, "Anonymous retrieval of k-nn poi in location based services (lbs)," in *In Proc. WORLDCOMP International Conference on Security and Management, SAM '13*, Jul.22-25, 2013, pp. 294–300.

[3] Y. Huang and R. Vishwanathan, "Privacy preserving group nearest neighbour queries in location-based services using cryptographic techniques," in *Proc. GLOBECOM'10*, Dec. 6-10, 2010, Conference Publication, pp. 1–5.

[4] B. Zheng, W.-C. Lee, and D. L. Lee, "On searching continuous k nearest neighbors in wireless data broadcast systems," in *IEEE TRANSACTIONS ON MOBILE COMPUTING*, vol. 6, mar 2007.

[5] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "Search continuous nearest neighbors on the air," *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, MOBIQUITOUS 2004*, pp. 236–245, Aug.22-26, 2004.

[6] H. G. Elmongui, M. F. Mokbel, and W. G. Aref, "Continuous aggregate nearest neighbor queries," in *Journal on Advances of Computer Science for Geographic Information Systems, Vol. 17*, jan 2013, pp. 63–95.

[7] Y. Tao, D. Apadias, and Q. Shen. (2002, Aug.) Continuous nearest neighbor search.

[8] J.-L. Huang and C.-C. Huang, "A proxy-based approach to continuous location-based spatial queries in mobile environments," in *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, Vol. 25*, feb 2013, pp. 260–273.

[9] Y. Xiaolan, D. Zhiming, and J. Jing, "Moving continuous k nearest neighbor queries in spatial network databases," in *IEEE Computer Science and Information Engineering, 2009 WRI World Congress on, Vol. 4*, apr 2009, pp. 535–541.

[10] G. Ghinita *et al.*, "Privacy queries in location based services: Anonymizers are not necessary," pp. 121–132, Jun.9-12, 2008.

[11] R. Vishwanathan, "Exploring privacy in location-based services using cryptographic protocols," Ph.D. dissertation, Univ. of North Texas, May 2011.

[12] C. Ma, C. Zhou, and S. Yang, "A voronoi-based location privacy-preserving method for continuous query in lbs," Oct.7, 2014.

[13] C.-Y. Chow, M. F. Mokbel, and W. G. Agref, "Casper* query processing for location services without compromising privacy," pp. 1–45, Dec 2009.

[14] C.-Y. Chow and M. F. Mokbel, "Enabling private continuous queries for revealed user locations," 2007.

[15] E. Kushilevitzr and R. Ostrovsky, "Replication is not needed: Single database, computationally-private information retrieval," in *Proc. IEEE 38th Annual Symposium on Foundations of Computer Science*, Oct.20-22, 1997, pp. 364–373.

[16] Z. Qingsong, L. Yansheng, and Z. Yanduo, "Predictive continuous nearest-neighbor query processing in moving-object databases," in *Proc. IEEE International Conference on Wireless Communications, Networking and Mobile Computing, WiCom 2007*, 2007, pp. 3019 – 3022.

[17] O. Atsuyuki, B. Boots, and K. Sugihara, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley and Sons Ltd, 2000, vol. 2.

[18] Z. Geng, K. Xuan, W. Rahayu, D. Taniar, M. Safar, M. L. Gavrilova, and B. Srinivasan, "Voronoi-based continuous k nearest neighbor search in mobile navigation," in *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, vol. 58,, 1997, pp. 2247–2257.

[19] F. Aurenhammer and R. Ostrovsky, "Voronoi diagrams -a survey of a fundamental geometric data structure," in *ACM Computing Surveys*, vol. 23,, Sep.3, 19917, pp. 345–405.

[20] H. ling. Chen and Y. in. Chang, "Neighbor-finding based on space-filling curves," in *Information Systems*, vol. 30,, 2005, pp. 205–226.

[21] R. J. Nicholas, "Hilbert-type space-filling curves," 2001.

[22] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*, 2008, vol. Springer-Verlag.

[23] W.-S. Ku, L. Hu, C. Shahabi, and H. Wang, "Query integrity assurance of location-based services accessing outsourced spatial databases," in *Advances in Spatial and Temporal Database*, vol. 5644. Springer Berlin Heidelberg, 2009, pp. 80–97.

[24] T.-T. Tan, L. Davis, and R. Thurimella, "One dimensional index for nearest neighbor search," in *Advances in Spatial and Temporal Database*. ResearchGate, aug 1999.

[25] Devsaran, "Sequoia points," Jan.15, 2015. [Online]. Available: http://www.chorochronos.org/?q=node/58

[26] H. Shin, V. Atluri, and J. Vaidya, "A profile anonymization model for privacy in a personalized location based service environment," in *Proc. IEEE ICMDM'08*, Newark, NJ, Apr.27-30, 2008, pp. 73–80.

[27] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," in *Journal on IEEE Transactions on Mobile Computings*, vol. 7, Jan. 2008, pp. 1–18.

[28] ——, "Location privacy in mobile systems: A personalized anonymization model," in *Proc. ICDCS International Conference on Distributed Computing Systems*, Columbus, OH, 2005, pp. 620–629.

[29] M. Gruteser and D.Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *In Proc. ICMSAS 1st International Conference on Mobile Systems, Applications and Services Mobisys'03*, 2003, pp. 31–42.

[30] M. Kohlweiss, S. Faust, L. Fritsch, B. Gedrojc, and B. Prenee, "Efficient oblivious augmented maps: Location-based services with a payment broker," in *Proc. ACM 7th international conference on Privacy enhancing technologies PET'07*, vol. 4776, 2007, pp. 77–94.

[31] M. Kolahdouzan and C. Shahabi, "Alternative solutions for continuous k nearest neighbor queries in spatial network databases," in *Proc. IEEE 38th Annual Symposium on Foundations of Computer Science*, vol. 9. Kluwer Academic Publishers, 2005, pp. 321–341.

Hybrid Watchdog and Pathrater algorithm for improved security in Mobile Ad Hoc Networks

N. Soganile[1], **T. Baletlwa**[2], **and B. Moyo**[2]
[1,3] Department of Computer Science and Information Systems, University of Venda, Thohoyandou, Limpopo, South Africa
[2] Department of Computer Science, Botho University, Gaberone, Botswana

Abstract-*Mobile hosts have a new wireless networking paradigm, called ad hoc networks. These networks do not rely on fixed infrastructures, like traditional wireless networks. To sustain connectivity hosts rely on each other for all communication transactions. The unique properties of ad-hoc networks are promoting an increased trend towards their adoption. The military and other security sensitive operations were the main and first application areas of ad hoc networks, but the trend has shifted now, businesses are relying on them. Mitigating the vulnerability of these networks to security attacks remains the main design challenge for these types of networks. A serious problem common in mobile ad-hoc networks (MANets) is the black hole attacks. The challenge of detecting and eliminating blackhole attacks in mobile ad-hoc networks, has led to poor network performance. This paper focuses on vulnerabilities of MANets and in particular it looks at the blackhole attacks. We present the design of an improved Watchdog and Pathrater algorithm called Radical Watchdog and Pathrater Algorithm for detecting and eliminating Black hole attacks. We carry out an extensive literature review and also make an analysis of the existing techniques for detection and elimination of Blackhole attacks in MANets. Our work culminates with the design of a RWP algorithm for improved security in MANets.*

Keywords: Mobile Networks, Black hole, MANets, Security technique, Intrusion detection,

1 Introduction

Technology is expanding every day, forcing a change in communication trends. Mobile Ad-hoc networks are a new paradigm of wireless communication, for mobile hosts. Unlike traditional networks, Mobile Ad-hoc networks do not rely on any fixed infrastructure, or any centralized control, such as base stations, or mobile switching centres. The mobile nodes communicate using a wireless network [1]. According to [2] the Mobile Ad-hoc network hosts are mobile and flexible, and they communicate with each other within radio range, through direct wireless links, or multi hop routing. Due to its mobility and portability in wireless communication, it introduces data security threats, and security attacks. Das [3] states that routing protocols in Mobile Ad-hoc networks are there to set up the most suitable path, between the source and destination, with minimum overhead, and minimum bandwidth consumption, so that packets are delivered in a timely manner. These suitable paths are known as “mini hops”. “Routing protocols are usually engaged to determine the routes, following a set of rules that enable two or more devices to communicate with each other” [4]. In MANets routes are enabled in between the mobile hosts, using multi hop, as the transmission range of wireless radio is limited. The mobile hosts are responsible for passing through packets over Mobile Ad-hoc networks, and they are not aware of the topology of the network. Routing plays an important role in the security of the entire network. The mobility and portability in Mobile Ad-hoc networks introduces security threats and security attacks. A change in topology means that security will have to be accessible, as nodes may be mobile, entering and leaving the network [5]. Mobile Ad-hoc networks are vulnerable to attacks that can be categorized into two types: Passive attacks and Active attacks, where active attacks can further be subdivided into internal and external attacks. Mobile Ad-hoc networks routing protocols are exposed to different types of attacks, Black-hole attacks, being the most serious type.

[6] states that “Black-hole attack is a specific type of attack, where a malicious node injects false route replies to the route requests it receives, by advertising itself as having the shortest path to a destination”. Therefore, the fake replies can be counterfeited to deflect network traffic through the malicious node, for eavesdropping, or simply to attack all traffic to it, in order to perform a Denial of Service (DoS) attack, by dropping the received packets. Although the performance of Mobile Ad-hoc networks under Black-hole attacks can be improved, by accurately detecting and eliminating Black-hole nodes, there are however indications which specifically suggest that the inaccurate detection and elimination of Black-hole attacks in Mobile Ad-hoc networks, can result in poor network performance [7].

1.1 Problem Statement

The increasingly developing trend of information and communication technology has not only provided our world with unequalled rewards, but has correspondingly created a conducive environment for manifold security challenges. Ad-hoc networks for instance, though a new and innovative wireless networking paradigm, are yet cheap prey to malicious attacks, due to their portability and mobility. This security weakness places huge demand for effective and accurate techniques, for detecting and eliminating threats such as Black-hole attack, to guarantee satisfactory performance in MANets. The concern here is to analyse existing security techniques in MANets, and suggest an approach to more effectively detect, and eliminate black hole attacks.

2 Ad-hoc networks

Subir [8] states, "Ad-hoc networks are a collection of wireless mobile nodes, which form a network without the use of any fixed infrastructure, or centralized control such as a base station". Mobile nodes are free to move randomly, and organize themselves, therefore the networks wireless topology may change rapidly and unpredictably. The birth of Ad-hoc networks is informed by the desire to improve upon the operational efficiency, and indeed performance of mobile networks. Eliminating the need for huge infrastructure, and centralized controls in Ad hoc networks is one of the brains behind their exploit [9]. These networks are classified into First generation (used for different military scenarios e.g. Packet radio networks), Second generation (used for the same purpose as the first generation ad-hoc networks system, but included further advancements, such as Global mobile information Systems, Near term Digital Radio (NTDR)), Third generation ad-hoc network systems, are also known as commercial ad-hoc network systems e.g. Bluetooth, ad-hoc sensor networks etc [10]. Such networks introduce redundant communication paths, which tend to improve fault tolerance for the network. Additionally, data packets are given the ability to "hop" from one user to another, thereby effectively extending the network coverage area; a solution to overcome non-line of sight (LOS) issues [9]. Computer devices and other devices operate in a peer-to-peer mode, in wireless topology without any access points, or centralized control such as base stations, as all devices communicate directly with other devices [11]. However, the deployment of mobile application required fast and widespread changes in topology, which possess a huge challenge to the mesh topology obtainable in ad hoc networks.

This gave rise to the different categories of ad hoc networks. Ad-hoc networks can be categorized into static and mobile ad hoc networks.

2.1 Mobile Ad-hoc Networks (MANets)

According to [12], Mobile Ad-hoc Networks have attracted most attention from many researchers world over. Deng [13] highlights that "Mobile Ad-hoc Networks are a collection of mobile nodes, which communicate with each other, and are connected to each other through wireless links". These mobile nodes are not fixed to any centralized control, like base stations or mobile switching centres, and there is no restriction on the nodes to join or leave the network, therefore the nodes join or leave freely. In the work by Kumar [14], all mobile agents or devices in MANets, are described as being "autonomous".

As stated earlier, mobile ad hoc networks are innovative offshoots of the traditional ad hoc networks. The additional challenge of extreme network flexibility, placed upon ad hoc networks by mobile applications, requires a technology which ensures quick and accurate update of communication routes. Moreover, MANets have the capability to do so. "MANets are self-forming, self-maintained, and self-healing, allowing for extreme network flexibility" [9]. MANets can be implemented as self-contained networks, or linked up to the internet, or private networks. The second style of setup is called hybrid MANets. In MANets, a network of self-configuring routers, connected via wireless links, is created forming a random topology. Furthermore, because such routers are highly mobile, and can quickly organize themselves at random, the topology of the wireless network changes rapidly and unpredictably. In a mobile ad hoc network there could be a great variety of mobile devices participating as autonomous nodes, either sending or receiving data from others within the network, as shown below.

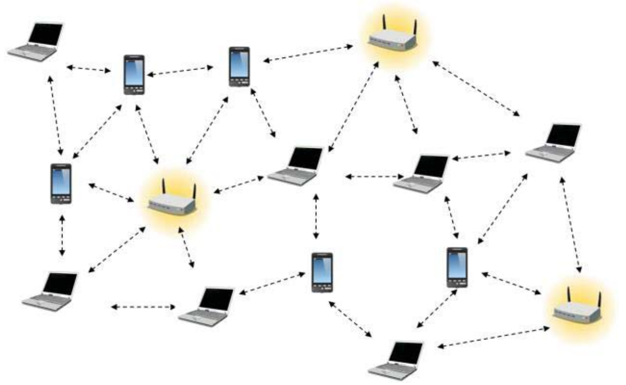


Figure 1: Mobile Ad-hoc Network

2.2 Features of MANets

The mobile ad hoc network has a number of features enumerated by [12] to include the following:

- Constantly changing topology, due to the continuous motion of nodes
- Unreliability of wireless links between nodes
- Lack of incorporation of security features, in statically configured wireless routing protocol, not meant for ad hoc environments
- Absence of Centralized Management

The above features, according to [12], are the reason why Mobile Ad-hoc networks are more prone to suffer several security challenges, caused by the malicious behaviors of nodes, other than the traditional wired networks.

2.3 Challenges of MANets

The common challenges faced by MANets is summarized by the following list:

- Limited bandwidth:
- Dynamic topology
- Mobility-induced route changes
- Battery constraints
- Security threats

2.4 Vulnerabilities in MANets

Every network is vulnerable in one way or the other, to one form of attack or the other. In the case of MANets, their vulnerability to attacks is higher than on the wired networks. In [15], some of the major vulnerabilities are listed as Lack of Centralized Management, No Predefined Boundary, Cooperativeness, Limited Power Supply, and Adversary inside the Network. The last weakness bothering security, forms the core of our work. As part of the multiple security concerns in MANets, we consider Black-hole attacks on MANets.

2.5 Black-hole Attack

The mobile nodes within MANets can freely join, and leave the network at any time [16]. This flexibility also introduces a security challenge, where a malicious node can pretend to be a legitimate member of the network, for purpose of compromising the security of the nodes. It is hard to detect that the behaviour of the node is malicious. Thus, this attack is more dangerous than an external attack [16]. The Black-hole attack actually falls under the category of attacks known as Network Layer Attacks.

The basic idea behind this kind of attack is that the intruding node injects itself into the active path from source to destination, or to absorb network traffic[16]. Technically, in a black-hole attack, the malicious node claims to have an optimum route to the node, whenever it receives route request (RREQ) packets, and sends the response packet (REPP) with highest destination sequence number, and minimum hop count value, to the originator node, whose RREQ packets it wants to intercept.

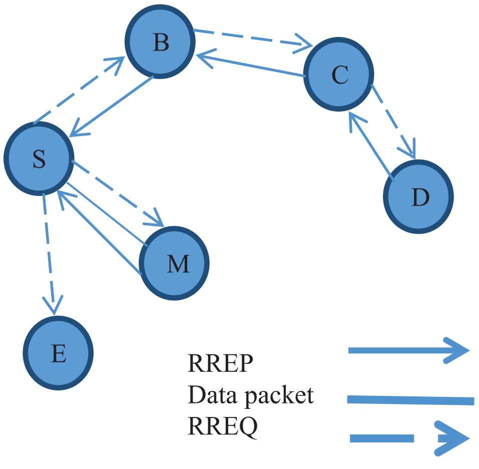


Figure 2: Black-hole Attack [16]

Looking at figure 2 above, node “S” wants to send data to node “D”, the destination node. It first initiates the route discovery process. The malicious node “M” immediately sends a response to source “S”, when it receives the route request. If the reply from node “M” reaches the source first, then the source node “S” ignores all other reply messages, and sends packet via route node “M”. As a result, all data packets are consumed, or lost to malicious node. This can lead to a security breach of confidentiality, integrity, and availability. So, by implication, in black-hole attack, a malicious node uses its routing protocol to advertise itself as having the shortest path to the destination node, or to the packet it wants to intercept the network packets [17].

2.6 Types of Black-hole Attack

There are two types of black-hole attacks described in ad hoc on-demand vector routing protocols (AODV in order). These are internal and external black-hole attacks.

2.6.1 Internal Black-hole attack

In this type of black-hole attack an internal malicious node fits itself in between the routes of a given source, and destination, as shown in figure 3b below. As soon as it gets the chance, this malicious node makes itself an active data route element [17]. At this stage, it gains the capability of conducting attacks against the network. This is an internal attack, because the node itself, at this point, belongs to the data route. Internal attack is more vulnerable to defend, because of difficulty in detecting the internal misbehaving node.

2.6.2 External Black-hole attack

In the case of external black-hole attacks, a malicious node simply stays outside of the network, and denies access to network traffic, or creates congestion in the network, or disrupts the entire network. External attack can become a kind of internal attack, when it takes control of a legitimate internal node and controls it maliciously to attack other nodes in MANets. The figure 3a below demonstrates an external black-hole attack.

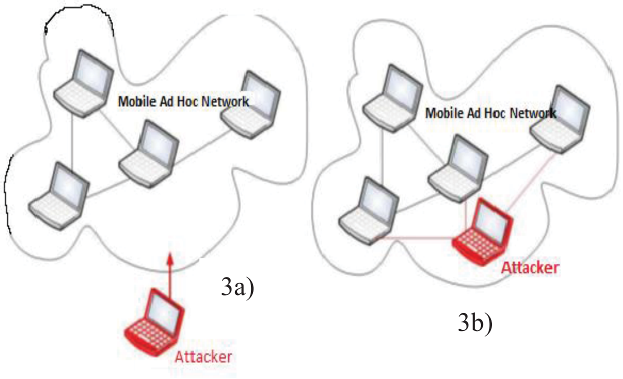


Figure 3: External and internal attacks [17]

2.7 Effects of Black-hole Attack on MANets

The cost of security breach in information communication cannot only be measured in monetary teams, because the reputation, integrity of organizations, and even the lives of its staff, could also be at risk. This is so, because in the event of security compromise, following a Black-hole attack, all three fundamental components confidentiality, integrity, and availability, which make up information security are violated. Black hole attack creates an artificial packet end-to-end delay, by misleading the source node into discarding responses from the legitimate node, while on the other hand keeping the legitimate node waiting for a response. This could have negative implications on bandwidth, and overall network performance. Throughput is also affect since it depends on the real time data being transmitted through the network. In a simulated experiment by [17], it was shown that throughput is higher in the absence of black-hole attack. [17] also highlights that e of data transmitted through the network, is a function of the number of nodes. Therefore, the presence of an illegitimate node adds to the existing network load. In addition, in order to frustrate the entire network, the malicious node tries to intercept all other messages within the network, thereby consuming more bandwidth.

3 Related work

3.1 Security techniques in MANets

Security always implies the identification of potential attacks, threats, and vulnerabilities of a certain system. In information systems, security is often defined in the context of being able to ensure confidentiality, integrity, and availability of network resources [18]. This fundamental requirement of computer security, as stated above is also valid when protection of correct routing behavior is to be considered, in any type of network [19]. This section surveys some security schemes, which have been deployed, or proposed to deal, with the attacks described in the earlier sections.

3.2 Wired Equivalent Privacy (WEP)

The security system in WLANS, based on 802.11 standard, consists of a data encapsulation technique called wired equivalent privacy (WEP), and an authentication algorithm called shared key authentication [19].

The weakness of WEP algorithm is that it can easily be broken. Introduced as part of the original 802.11 standard ratified, it was intended to provide data confidentiality, comparable to that of a traditional wired network. WEP, recognizable by 10 or 26 hexadecimal digits, was at one time widely in use, and often the first security choice presented to users by router configuration tools, but it has been observed to have several problems identified with 802.11 securities.

3.3 Watchdog and Pathrater Techniques

The Watchdog and Pathrater technique were introduced by [20] to improve throughput in a MANets, by identifying misbehaving nodes, which trick other nodes, by agreeing to forward the packets without ever doing so. The security model is made up of two components, the watchdog and Pathrater. While the watchdog is used to identify misbehaving (malicious) nodes, initiated by a Replica server, Pathrater helps routing protocols avoid these nodes, by removing them, and creating a new path. The watchdog occurs in every node in the network. When a node forwards a packet, the nodes watchdog component verifies that the next node in the path also forward the packet. The only way a watchdog can do this, is by listening in a promiscuous mode, to the next node's transmission. If the next node does not forward the packet, it is said to be a malicious (mischievous) node, and has to be reported. This is done by sending an alarm message to the other nodes on its friends list. When the nodes accept the alarm message, they check it, and change the status of the accused node, only if the alarm source is trusted, or a number of trusted nodes accused the same node. The previous status of the node is also maintained, where its structure contains server node ID, destination node ID, hop count and drop packets.

3.4 Intrusion Detection and Response Mechanism (IDRM)

IDRM security model for MANets [21] was first presented by Zhang and Lee. In its architecture, all the nodes take part in the intrusion detection, and response mechanism. Zhang and Lee worked with the basic assumption, that the user and associated program activities are observable, and under a cooperative distributed system.

The intrusion detection and response mechanism, by [21] comprised two key modules – a data collector and local detector. The data collection mechanism present in every node, gathers streams of real-time audit data, from various sources. Whereas the local detector analyzes the local data traces, gathered by the local data collection module, for evidence of anomalies.

4. The Design of the Radical Watchdog and Pathrater Algorithm

4.1 Overview of RWP algorithm

The Radical Watchdog and Pathrater algorithm (RWP) is a hybrid technique built on the ideas of Watchdog and Pathrater algorithm introduced by [20]. The RWP is achieved by extending the capabilities of the Watchdog and Pathrater to including a radical deletion of the malicious nodes up to the second hop in the network. The watchdog component on the malicious node detects the misbehaviour of node and then sends warning signals to its neighbouring nodes including the source node. On receiving this signal the source node stops using the malicious node as the link to its destination. All the first hop nodes including the source node delete the malicious node from their friends list and further send the warning signal to the second hop nodes. The second hope nodes respond by flagging the malicious node for future data transactions with it. Any attempt by the malicious node to send RREQ will be met with denial of access and deleted from the friends list. Once the deleted, a warning signal is sent to the next hope. The process will continue until the node is certified as clean by the watchdog component, the watchdog sends a clearance certificate to all its neighbouring nodes causing all the nodes to unflag the malicious node.

4.2 Radical Watchdog and Pathrater algorithm

The algorithm steps of the RWP are:

1. Source node sends route request (RREQ)
2. Node with shortest route sends a (RREP)
3. Watchdog detects that the Node is malicious
4. Watchdog sends an alarm message to the other nodes in first hop
5. Source node stops sending packets to malicious node
6. The nodes receiving the alarm delete the malicious node from friend list
7. And forward the alarm to the second hop nodes
8. On receiving the alert, those nodes on the second hop flag the malicious node, for future reference
9. Any request coming from the flagged node is not entertained and the node is deleted form the friends list, and further warning is send to the next hop nodes
10. Re-integrate the node when Watchdog sends a clearance certificate the node.

4.3 The Sequence diagram

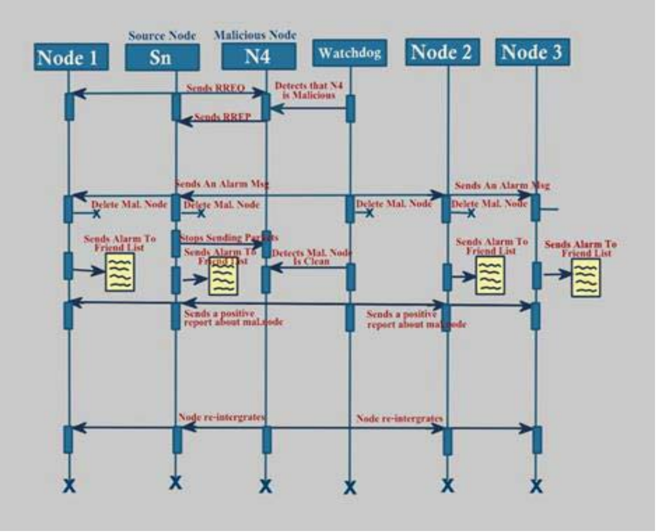

Source Node
Malicious Node
Node 1
Sn
N4
Watchdog
Node 2
Node 3
Sends RREQ
Detects that N4 is Malicious
Sends RREP
Sends An Alarm Msg
Sends An Alarm Msg
Delete Mal. Node
Delete Mal. Node
Delete Mal. Node
Delete Mal. Node
Sends Alarm To Friend List
Stops Sending Packets
Sends Alarm To Friend List
Detects Mal. Node Is Clean
Sends Alarm To Friend List
Sends Alarm To Friend List
Sends a positive report about mal.node
Sends a positive report about mal.node
Node re-intergrates
Node re-intergrates

Figure 4: Sequence diagram for Radical Watchdog and Pathrater

The sequence diagram in figure 4 above shows the interaction between objects in the sequential order. The node interaction, passing of messages and relevant alert signals is shown on the sequence diagram.

5. Conclusions and recommendations

Black-hole attack is one of the biggest concerns in Mobile Ad-hoc networks. It violates confidentiality, integrity, and availability; being the important three main components of information security. Black-hole attacks have a negative effect on the networks bandwidth, and the overall network performance is affected by the malicious nodes delaying the packets. The network throughput is reduced significantly. A number of researchers have made enormous contribution towards Black hole detection and elimination. It is hoped that the RWP algorithm presented in this paper adds to the list of techniques to handle the Black hole attack in MANets. RWP is an extension of the WP algorithm presented [20]. RWP algorithm is for detecting and eliminating the black-hole attacks in MANets. It controls the intrusion by the malicious node by deleting and sending warning alerts up to the second hop nodes. This extension ensures that the malicious node is eliminated not only within its surroundings but further across the hops in the mobile ad hoc network. Indications are that the RWP algorithm is an improved algorithm compared to the WP algorithm. Such claims need to be tested through simulation experiments.

Reference

[1] Weerasinghe, H., Fu, H. & Ieee, M., 2008. Preventing Cooperative Black Hole Attacks in Mobile Ad Hoc Networks : Simulation Implementation and Evaluation. , 2(3).

[2] Tanwar, S. & Prema, K. V, 2013. Threats & Security Issues in Ad hoc network : A Survey Report. , 2(6), pp.138–143.

[3] Das, R., Purkayastha, B.S. & Das, P., 2011. Security Measures for Black Hole Attack in MANET : An Approach. , 3(4), pp.2832–2838.

[4] Singh, T.P., Kaur, S. & Das, V., 2012. Security Threats in Mobile Adhoc Networks : A Review. , 2(1), pp.27–34.

[5] Singh, G., 2011. Security Threats and Maintaince in Mobile Adhoc Networks. , 2(3), pp.68–70.

[6] Subathra, P., Sivagurunathan, S. & Ramaraj, N., 2010. Detection and Prevention of single and Cooperative Black Hole attacks in Mobile ad Hoc Networks. , 6(March), pp.38–40.

[7] Shree, O. & Talib, M., 2011. Wireless Ad-hoc Network under Black-hole Attack. *International Journal of Digital Information and Wireless Communications (IJDIWC)*, 1(3), pp.591–596.

[8] Subir, kumar sarkar, Basavaaju, T. & Puttamadappa, C., 2008. *Ad Hoc Mobile Wireless Networks*, New York, London: Auerbach.

[9] Sumyla, D., 2006. Mobile Ad-hoc Networks.

[10] Bakht, H., 2011. History of mobile ad hoc networks Mobile ad hoc networks , pp.1–16.

[11] Lee, B. et al., 2004. *Wireless Hacking* C. Kloiber, ed., United states of America: Syngress.

[12] Li, W. & Joshi, A., 2011. Security Issues in Mobile Ad Hoc Networks - A Survey. , pp.1–23.

[13] Deng, H., Li, W. & Agrawal, D., 2002. Routing security in wireless ad hoc networks. *Communications Magazine, IEEE*.

[14] Kumar, B.P., Sekhar, P.C. & Bhushan, B.B., 2011. A SURVEY ON MANET SECURITY CHALLENGES AND ROUTING Abstract-. , 4(2), pp.248–256.

[15] Science, C. & Engineering, S., 2013. International Journal of Advanced Research in Study of MANET : Characteristics , Challenges , Application and Security Attacks. , 3(5), pp.252–257.

[16] Aarti, S.T., 2013. Study of MANET : Characteristics , Challenges , Application and Security Attacks. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(5), pp.252–257.

[17] Ullah, I. & Rehman, S.U.R., 2010. Analysis of Black Hole Attack on MANETs Using Different MANET Routing Protocols.

[18] Kovacich, G.L., 2002. *THE INFORMATIONSYSTEMS SECURITY OFFICER'S GUIDE: Establishing and Managing an Information Protection Program*,

[19] Venkatraman, L. & Agrawal, D.P., 2003. Strategies for enhancing routing security in protocols for mobile ad hoc networks. , 63, pp.214–227.

[20] Segio, M., Giuli,T.J., Lai, K. & Baker, M., 2000. Mitigating routing misbehaviour in Mobile Ad hoc Networks.

[21] Zhang, Y. & Lee, W., 2003. Intrusion Detection Techniques for Mobile Wireless Networks. , pp.1–16.

[22] Riso, B.G., Augusto, F. & Westphall, C.B., 2013. Wireless Communications : Security Management Against Cloned Cellular Phones Mirela Sechi Moretti Annoni Notare.

Anonymous Authorization Scheme Using ECC for RFID Privacy

Chung Sei Rhee
Department of Computer Science
Chungbuk National University Cheongju
Chungbuk, KOREA
csrhee@cbnu.ac.kr

Abstract

The development of mobile technique and the convenience of using mobile technique are hot subjects in many areas and a lot of researches have been done in mobile RFID reader technique. Mobile RFID and security problem in network are closely related. But all the security authentication algorithms of RFID are range problem between TAG and Reader. The range between Reader and back-end DB is composed by wired networks. Therefore, it is based on secure range. But we should consider information security and privacy problem in wireless range in the design of Mobile RFID reader. In this paper, we design a blind signature scheme based on weil-pairing finite group's ECC encryption scheme and propose the anonymous authorization scheme for mobile RFID reader.

Keywords: RFID, Privacy, Anonymous, Authorization, ECC

1. Introduction

The RFID(Radio Frequency Identification) is an automatic identification system, relying on storing and retrieving remotely connected data using device called RFID tag. A secure RFID system has to avoid eavesdropping, traffic analysis, spoofing and denial of as large read range and no line between any two location. There have been some and capable memory, physical tag memory separation, hash encryption etc .[1]

The RFID technique, however, causes the serious privacy infringement, such as excessive information exposure and user's location information tracking, due to the wireless characteristics since it is easy to be recognizable without physical contact between reader and tag while the tag information is sent.

Therefore, researches on the authentication protocol has been done to protect the information stored and solve the safety problem.

The security problems in RFID system are composed of three components such as tag, reader and back-end data base and the communication parts are also composed of tag, reader and back-end DB. Tag and reader parts are considered unsecure channel because of limited resources and wireless characteristics. But reader and back-end DB are considered as a secure channel since it use wired characteristics.[1] But the fast development of wireless communication and efficiency makes RFID reader research activity. [2]

Currently, encryption algorithm adopt RSA or ElGamel public key algorithm in mobile commerce. But these methods need lot of calculation and encryption/decryption speed is relatively slow, so these methods can't be applied to mobile RFID reader. We propose an eclipse Weil-pairing finite group anonymous authorization algorithm for mobile RFID reader. The proposed algorithm considers mobile RFID reader's operation ability and guarantee user's anonymity.

2. Related Works

Anonymity technique has been hot research subject since personal privacy protection problem occurs as a world-wide issues. Mix-network, electronic signature and anonymous certificate are known as main techniques for information protection.

2.1 Mix-Network based anonymous technique

Mix network is a technique which hides the web user′ internet use information.[3] Mix network receives encrypted message and re-encrypt the message and print it in random order. Therefore, it is impossible to find the correspondence between input message and output message. This method is applied to electronic mail, web-browsing and anonymity of pay system. [4]

1) Anonymizer

An anonymizer is a technique which hides information for web user′ internet usage.

2) Onion Routing

This is a technique hide data traffic contents using Mix-network. Onion routing developed as anonymity of packet on the network, it allow safe communication system as well as anonymity user′ communication contents.

3) Crowds

This is a technique hide contents of routing of HTTP traffic. To guarantee user's anonymity on the internet, it hides data traffic between server and sender.

4) Janus

This is a proxy server technique give anonymity both for client and server using URL encryption.

IP address and host name are hidden to generate anomymity.

5) TAZ

TAZ allows both integrity and security of data using encryption of URL and data stream. It solves the problem of Janus by encrypting data traffic.

2.2 Anonymous technique based on Electronic Signature

1) Group signature

Group signature was proposed by D. Chaum and Van Heyst in 1991.[5] member of group sign for the group and anonymity of signee is guaranteed.

Vehicle safety communication system proposed by department of transportation in America allows anonymity of user using group signature. Efficient signature requires short signature and short signature needs 200 bits.[6, 7]

2) Ring signature

Ring signature was proposed by Rivest, Shamir and Tauman in 2001 and it is very similar to group signature. The difference is no group administrator. Group signature has a group administrator, so signee is known but ring signature does not allow signee.[8,9]

3) Blind signature

Blind signature is developed as an electronic pay system and proposed by Chaum in 1981. This uses RSA encryption method. Blind signature allow complete anonymity and used for electronic voting system too.[10, 11]

2.3 Other Methods

1) Anonymous Credential

Anonymous Credential is called pseunym and proposed by Chaum in 1985.[12] System is composed by organization and user and organization knows user' pseudonym and issues credential but it does not know user' information. Therefore, user' privacy is protected. Recently, Anonymous credential was proposed by Camenisch and Lysyanskaya.[13]

2) Distributed Encryption method

Distributed Encryption method has been studied by Shamir in 1979 as a master key for secret key share problem. One holds master key, then key lost as well as cheating problem may happen. Therefore, it distributes key among many employees. If there are n employees, then there exists at least k employees to decrypts the key. This is called (n, k)-threshold.[14]

3. Our method

3.1 Anonymous encryption signature using Weil-paring

Anonymous signature is an effective signature method which does not show the message contents to the message signer.

Anonymous signature guarantee to protect personal privacy since it provides complete personal anonymity and un-tractability. In this paper, we propose an Anonymous signature method based on ID based public key encryption technique using Weil-pairing which is a super-singular elliptic curve. Signature is composed of three steps such as initialization, signature creation and signature verification.

1) Initialization step

Initialization step is the step where define the objects and create and register the parameters necessary to signature as shown in Table 1.

Table 1.

Initialization	Comment
G	GDH group using prime number Prime number l
P	Generator G
$\hat{e}$	bilinear weil-Pairing Function
H_1, H_2	Collision free Hash Function
id_x	ID of Signee
$t, r \in Z/l$	Random number
$W_x = H_2(ID_x)$	Public Key(signee)
$w_X = t \bullet W_X$	Private key(signee)
$P_X = tP$	Public
M	Signed Message

2) Signature Creation Step

We receive a signature using message exchange using the following algorithm.

Algorithm to create a signature.

Step1 : Signee select a random number $r \in Z/l$, then calculate $A = r W_X$ using personal public key W_X and send it to user.

Step2 : User calculate $h = H_1(M, A)$ using the received A and signed message M and return it to signee.

Step3 : Signee calculate $B = (r+h)w_x$ using encrypted message $h = H_1(M, A)$ and send B to user.

Step4 : User receive both A and B from signee, therefore get signed message $Sign_x(A, B)$.

3) Correctness of the signed the message

User verify signed message using bilinear function by calculating $\hat{e}(P_X, A + h W_X)$ and $\hat{e}(P, B)$ and compare two values.

Theorem 1. $\hat{e}(P_X, A + h W_X)$ and $\hat{e}(P, B)$ have the same value.

Proof : Since $P_X = tP$ and $A = r W_X$, we substitute it into $\hat{e}(P_X, A + h W_X)$ and get

$$\hat{e}(tP, r W_X + h W_X) \qquad (1)$$

Weil pairing has the property such that $\hat{e}(aP, bQ) = \hat{e}(P, Q)^{ab}$, we can change (1) as given below.

$$\hat{e}(tP, r\,W_X + h\,W_X) = \hat{e}(P, (r+h)\,W_X)^t$$
$$= \hat{e}\,(P, (r+h)t\,W_X) \qquad (2)$$

Since signee’s private key and public key’s relationship is $w_X = t\,W_X$, equation (2) is

$$\hat{e}(P, (r+h)w_X) \qquad (3)$$

Since $B = (r+s)w_X$, substitute it into (3), we get

$$\hat{e}(P, (r+h)w_X) = \hat{e}(P,B)$$

Therefore

$$\hat{e}\,(P_X, A + h\,W_X) = \hat{e}(P, B) \quad \text{q.e.d}$$

3.2 Reader Authorization using Anonymous Authorization

Authorization step in our model is composed of initialization for parameter and object, request for sign and creation as well as verification of created key. Initialization is similar to as we gave earlier, we just describe how to authenticate the signed message.

In order to guarantee the anonymity of reader’s user, signed message is transformed into anonymous signed message using hash function. Back-end sends an anonymous message to the reader then reader create sign and user authenticate it using signature authentication. This step is composed of five steps as shown in Figure 1.

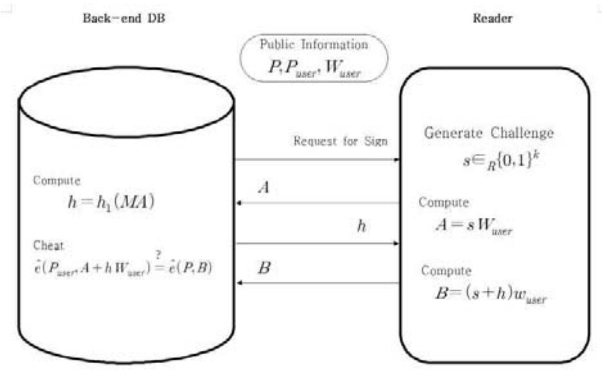


Figure 1. Reader Authentication technique using reader Anonymization

The Authorization step is given as follows.

1. Back-end DB requests the sign to reader for authentication.

   $DB \rightarrow\ Reader$: Request for sign

2. Reader select a random number $s \in Z/l$, and calculate $A = s\,W_{user}$ using random number and public key W_{user} and send it to back-end DB.

   Reader:

   Generate challenge $s \in_R [0,1]^l$

   Compute

   $A = s\,W_{user}$

   $Reader \rightarrow\ DB\colon A$

3. Back-end DB create reader related information M, and calculate hash value h using A and M. Then, it send hash value which contains anonymous message.

   DB:

   Compute

   $h = H_1(M,A)$

   $DB \rightarrow\ Reader : h$

4. Reader create a signed message B using hash value h, random number and personal private key and send it to back-end DB.

Reader:

Compute

$$B=(r+h)w_{user}$$

$$Reader \rightarrow DB: B$$

5. Back-end DB authenticates the anonymous message B related to reader using A, h and reader's public key.

We authenticate the correctness of reader whatever the reader.

4. Performance Evaluation

4.1 Analysis of Safety

1) The proposed method is based on weil-pairing anonymous authentication as well as one way hashing and an elliptic encryption tool. Therefore, it is safe against attack.

2) It is impossible to masquerade the reader. Normally, tag does authorization of the reader but strong authentication is impossible because tag gets limited resource effects. Therefore, present methods are weak to masquerade attack. The proposed authorization method authorizes based on super singular ellipse Weil-pairing authorization method. Therefore, there is no way to find t he random number s selected by reader and it is impossible to masquerade the reader.

3) Reader provide anonymity. The proposed reader authorization method encrypts using random number s, t and one way hashing h_1 using ID information, one can verify the validity of reader but can't trace the reader′ location.

4.2 Analysis of efficiency

The proposed reader authorization method encrypts using random number s, t and one way hashing h_1 using ID information, the operation is performed as an addition on eclipse. Therefore, key is relatively short and has strong safety.

We compare our method and anonymous techniques.

Table 2. Comparison of Anonymous signature methods

Method		Operation
Mix Network Based	RSA/ EIGamal	Exponent/ multiplication
Elecronic Signature Based	RSA/ EIGamal	Exponent/ multiplication
Proposed Method	ECC	Add Operation

5. Conclusion

RFID technique is used to many area and regarded as ubiquitous computing technique. But is relatively expensive, weak to security in mobile environment and privacy infringement.

Reader authentication method, we proposed here, has applied an eclipse curve discrete logarithm for Weil -pairing finite group encryption algorithm. Reader recognition information is anonymous by using one side hashing function. Therefore, anomynity of user is guaranteed. Our method also use ECC encryption technique, so it is faster than RSA or ElGamel public key.

References

[1] Yong Zhen Li etc, "Security and Privacy on Authentication Protocol for Low-cost RFID,", Proc of 2006 Internation Conference on Computational Intelligence and Security(CIS06), Part 2, pp. 1101-1104, Nov. 2006

[2] A. Juels "RFID Security and Privacy: A Research Survey,", In IEEE Journal on Selected Areas in Communications, Vol 24., No.2, pp. 381-394, 2006

[3] D. L. Chaum, " Untraceable Electronic Mail, Return Address and Digital Pseudonyms," Communications of the ACM, vol. 24, No. 2, 1981, pp. 84-88

[4] A. Serjantov, P. Sewell "Passive attck analysis for connection based anonym systems," International Journal of Information Security, Vol 4, No. 3, 2005, pp.171-180

[5] D. L. Chaum, V. Heyst, "Group Signature," Advances in Cryptography_Eurocrypt 1991, LNCS 547, pp. 257-265

[6] D. Boneh, X. Boyen and H. Shacham, "Short group signatures," Advances in Cryptology-Crypto 2004, LNCS 3152, pp. 41-55

[7] D. Boneh, X. Boyen, "Short signatures without random oracles," Advances in Cryptology-Crypto 2004, LNCS 3207, pp. 56-73

[8] R. L. Rivest, "Chaffing and Windowing: Confidentiality without Encryption," CryptoBytes 4(1), RSA Lab, 1998, pp. 12-17

[9] R. L. Rivest, A, Shamir, Y. Tauman, "How to leak a Secret," Asia-CRYPT 2001, LNCS 2248, pp. 552-565

[10] S. Brands, "Untraceable Off-line Cash in Wallets with Observers," CRYPTO'93, LNCS 773, pp. 552-565

[11] M. Bellare, C. Namprempre, etc, " The One-More_RSA-Inversion Problems and the Security of Chaum's Blind Signature Scheme,“ Journal of Cryptology, June 2003, pp. 185-215

[12] D. L. Chaum, " Security Without Identification: Tramsaction Systems to Make Big Brother Obsolute," Comm of ACM, 1985 pp. 1030-1044

[13] J. Camenisch, A. Lysyanskaya, "Signature Scemes and Anonymous Credentials from Bilinear Map," CRYPTO 2004, LNCS 3152, pp. 56-72

[14] A. Shamir, "How to Share a Secret," Comm of ACM, 22, 1979, pp. 612-613

Home Network Security: Beginner Versus Advanced

Lauren Evanoff[1]**, Nicole Hatch**[1]**, and Kanwalinderjit Gagneja**[1]
[1]Department of Computer Science, Southern Oregon University,
Ashland, OR, USA

Abstract - *The world of technology is an ever advancing realm where companies are constantly fighting to put out the latest and greatest product or service. The issue that arises with this is that adequate research and development is not being conducted. With technology rapidly changing and the use of the internet ever growing, the need for higher internet and network security is at an all-time high. With computer and network hackers and malicious malware and viruses, personal information is getting more and more difficult to secure. Within this paper we will show you how an average home user would set up their home network and the security settings and measures that should be taken to safeguard their personal information.*

Keywords: Security, wireless, standards, Ethernet, signal strength

1 Introduction

The connection for a home network is coming from ether dial-up, digital subscriber line (DSL), or a cable internet service provider (ISP). The link to the ISP is coming from the wall through a phone line or a coaxial cable to the modem. The modem is connected to a wireless router or directly to an end system using an Ethernet cable. End systems can include computers, printers, smart phones, tablets, game consoles, and smart televisions.

When setting up a network for the first time the user will need to first ask themselves, “What kind of network am I setting up, wireless or wired?” Wireless will be easier when setting up a small network with only a few computers, phones, and a printer, but may not give all the security benefits as wired does. Using a wireless setup gives more flexibility for connections, but leaves the network open to many potential security cracks. “Data is sent and received using radio waves eliminating the need for Ethernet cables. You can connect to the network from anywhere within range of the wireless router.” (Verizon, 2014)

1.1 Wireless or wired

Using a wireless setup has advantages and disadvantages. Advantages can include flexibility, visibility, cost, and easier after the setup is completed. There are no cables or cords to hide or possibly trip over, therefore giving the mobility to move around the room with the computer and not have to sit at a desk and having the ability to print from anywhere within range of the wireless router. If the location is going to have many visitors, a wireless network will be a better choice. Renting equipment is an option when receiving an internet service, which may be a better option for users that would prefer to not have to buy something new every time an upgrade or improved service is available.

Disadvantages can include lower connection speeds, disruptions and “dead spots” whereas a wired network would have fewer chances of these issues. Wireless connections “use 802.11g wireless networking which transmits data at 2.4 GHz with a speed of 54 megabits. The newer wireless standard of 802.11n is faster and has a longer range than 802.11g.” (Wilson & Fuller, 2001) More than likely, a wireless router and modem will be used when setting up a wireless network and they will be connected by an Ethernet cable; however sometimes the router and modem can be combined into one unit to simplify the installation. The router can send its signal up to 100 feet in all directions though there can be disruptions and dead spots due to walls and the layout of the home. (Wilson & Fuller, 2001) If the home owner has a two story house an upgraded router or a Wi-Fi router extender, which boosts signal strength, can be used to reach both floors. The location of the wireless router or extender has a huge effect on how strong the signal is throughout the home. Some older computers do not come with a built in wireless capability, so a wireless adapter will need to be installed. This will be plugged into a USB and will work like a built in component. (Netgear, 2014) In a wired network, end users will connect to the router using Ethernet cables. A wired setup is best for computers, printers, televisions and gaming consoles although cell phones and tablets do not have Ethernet capability.

As with wireless, a wired network has advantages and disadvantages as well. Advantages can include faster speeds with downloading and uploading data and an added layer of security. Wired networks provide more security for the user because they are not sending the information wirelessly but directly to the router or modem. The upload and download speed can depend on the ISP that the user has selected. The local cable ISP may offer up to 100 Mbps, while the local DSL offers around 12 Mbps. (Century Link, 2014)

When setting up more than one computer there are a few disadvantages to a wired network because there is more work that will go into it. The user will need to get longer Ethernet cables depending on how far apart the router and computers are and if the setup is going to have more than four computers then a switch or hub will be needed. (Wilson & Fuller, 2001) Over time this can become more costly of a setup. Wired networks are good when setting up a single room for the home office. If the home office will be in more than one location and is not a desktop computer, then a wired network may not be the way to go.

2 Setting up home network

When setting up a home network for the first time, the local ISP will come to the home and set it up or send everything to the home for the home owner to set up themselves. The box may include a modem, wireless router, power cords, an Ethernet cable and some instructions on how to complete the install. The instructions are going to show what settings work best for the home network. We are going to first show the security settings that are requested from the ISP 'Charter' for a new user right out of the box. Then we are going to show some of the advanced settings to make the home network stronger, along with some applications the user can do to make their information more secure.

The following will explain the procedures for setting up a typical wireless home network from the beginning, right out of the box. After calling Charter and requesting the service, we had to go down to Charter's local office and get an internet modem that we would be using to connect our home to Charter's internet service. The internet service we have chosen to go with has a download speed of 30 Mbps with an upload speed of 4 Mbps. (Charter Communications, Inc., 2014) We then received a box in the mail that included the router, router power cord, router stand and an Ethernet cable. The box also included a letter thanking us for choosing Charter and the instructions on how to set up the wireless router. The letter from Charter also explains that the router comes with a pre-set network name and network key. The letter tells us that all of the information for the router is printed on the router itself and that installation should take no more than 10 minutes. The letter also states that "Charter recommends that you do not change the pre-set Wi-Fi Network Name and Network Key" (Charter Team, 2014).

To start the installation process the instructions show us a picture on how and where the modem and router are to be put together. We started by attaching the coaxial cable to the cable wall outlet and then attached the cable to the modem. After attaching the power cord to the modem we then attached the Ethernet cable to the router connecting the two together. We then connected the power cord to the wireless router and put them in the location where we wanted it to be. For this home network the setup is going to be in the east corner of the living room and the house is single story and small so we do not need any router extenders.

The Netgear wireless router that came with our installation has two types of Wi-Fi bands that can transmitted, known as a dual band router, provides 2.4 GHz and 5 GHz bands. The difference between 2.4 GHz and 5 GHz is the signal range and the bandwidth you receive for your network. 2.4 GHz can reach further than 5GHz but if the user needs more bandwidth than 5GHz is greater than 2.4GHz. (Netgear, 2013) Most computers and wireless devices run on a 2.4 GHz, so for this home network we will be adding all the devices to the 2.4 GHz.

During set up when connecting the primary computer to the wireless router we are asked to type in the pre-set password. In this case the router came with the service set identifier (SSID) "MyCharterWiFic1-2G" and a password of "aquaticink138". These are both printed on the back of the router and on a sticker that we can attach anywhere else as needed. Netgear offers a user application interface that can be downloaded to your computer or smartphone to help run your network. We installed this on the primary computer that will be running the network and to an Apple smart phone. This application shows the status of the connection between the router and the computer, if the router is connected to the internet, a network map of all the devices that are connected to the network, router settings, a ready share link and parental controls.

When using the Netgear application we are able to change the user name and password settings. This user name and password is prompted when devices are attempting to connect to the network. We changed the SSID to "MyCharter" and the password to "Password123". By default the channel selection is set on auto and the wireless encryption option is set to WPA2-PSK. We did not change this setting, however, if we wanted our network to be open without a password being required to log on, we would choose "none" for the wireless encryption option. The router administration user name and password also came preset. The username is "admin" and the password is "password". Notice in the log in window that Netgear tells us what the password is. We were able to change the password to "Password123" but were unable to change the username from admin. Take note that by default both "remember password" and the IP address is already entered for us and we did not change this setting.

Not all the settings for the router are in the Netgear application. The application covers the simple settings for the router but to get into the administrative settings we need to use a web browser to log into the router itself. In the browser bar, type in the IP Address of the router, also known as the default gateway, where typically most routers are 192.168.1.1, but the default IP Address can also be found on the log in page of the router login. When putting the IP address into the browser bar a window pops up where we can put in the user name and password. In our case the username is “admin” and password “Password123”.

Some of the default settings are ‘Does your internet connection not require a login’, the default for this is no. There is also an option to change the type of internet address from ‘Dynamically from ISP’ which is automatically assigned from the ISP or a ‘Static IP Address’ which is manually set from the user. With our application we are using “get dynamically from ISP”. “…dynamic IP addressing allows for a new IP address to be assigned each time the user logs on...” (What Is My IP Address, 2014) Static IP allows the user to set and know what the IP address is at all times. This may be useful if the user is using the computer for gaming consoles to play online or if the user is using the computer as a server.

The default security setting can also be changed from the administrative location. Two default settings are to ‘enable SSID Broadcast’ and the wireless encryption which is set to WPA2-PSK [AES]. We are not changing any of these settings. The only one that we have changed is the SSID to “MyCharter”. From this location we can also set the password. This change can be done by both the application and the web browser link in the administrative location. We have set both 2.4 GHz and 5GHz to the same password but added 5G to the end of the user name to see each SSID.

3 Setting up Security

Now that you know the details of a beginner/low security network, let’s learn about the options the home user has to help protect their personal information and their network. There are many ways to accomplish this and we will discuss a few of them and provide some examples. To start with we will go over the different wireless encryptions available which are WEP, WPA and WPA2.

3.1 Wired Equivalent Privacy

Wired Equivalent Privacy (WEP) is an easily broken security algorithm for Institute of Electrical and Electronics Engineers (IEEE) 802.11 wireless networks. Introduced as part of the original 802.11 standard ratified in September 1999, its intention was to provide data confidentiality comparable to that of a traditional wired network (Wikipedia, 2014). Even though WEP was part of the original IEEE 802.11 standard it fell short from the needed requirements of wireless security. Many hackers would find this as a very easy encryption to hack leaving virtually all wireless networks at risk.

WEP would prove to have numerous flaws, one of which being it’s encryption process. WEP used RC4, which is a stream cipher that does not use the same traffic key twice. It also utilized a 24 bit initialization vector (IV) that is transmitted as plain text. The purpose was to prevent any repetition; however, 24 bits were just not long enough to make this happen on a busy network. After about 5000 transmitted packets, there was a 50% chance that the same IV would be repeated. So a potential hacker would not have to wait very long to catch this and give themselves access to the network. “In 2005, a group from the U.S. Federal Bureau of Investigation gave a demonstration where they cracked a WEP-protected network in 3 minutes using publicly available tools”. (Wikipedia, 2014)

3.2 Wi-Fi Protected Access

Wi-Fi Protected Access (WPA) was intended by the Wi-Fi Alliance as an intermediate measure to take the place of WEP pending the availability of the full IEEE 802.11i standard. WPA could be implemented through firmware upgrades on wireless network interface cards designed for WEP that began shipping as far back as 1999. (Wikipedia, 2014)

In April 2003, WPA was introduced as an interoperable security protocol based on draft 3 of the IEEE 802.11i amendment. WPA was designed to be a replacement for WEP without requiring hardware replacements, using a subset IEEE 802.11i amendment. (Wright, 2006). WPA implemented Temporal Key Integrity Protocol which generates a new 128 bit key for each packet. Having this prevented the types of attacks that affected WEP. Another addition to WPA is a message integrity check that would detect if messages had been altered during transmission.

In July 2004, the IEEE approved the full IEEE 802.11i specification, which was quickly followed by a new interoperability testing certification from the Wi-Fi Alliance known as **WPA2**. WPA2 is based on the Robust Security Network (RSN) mechanism, which provided support for all of the mechanisms available in WPA. As of March 2006, the WPA2 certification became mandatory for all new equipment certified by the Wi-Fi Alliance, ensuring that any reasonably modern hardware will support both WPA and WPA2 (Wright, 2006).

3.3 Passwords

When setting up both Wi-Fi and wired networks, a password can and should be used to connect to the router when accessing wireless connections and administrative settings. According to Network World, the four most commonly used passwords for 2013 were: ‘123456’, ‘password’,’12345678’, and ‘qwerty’ (Smith, 2014). Consumers have also been known to use their home phone number or their birthday as their password. These types of passwords leave the network open for someone to hack into it extremely quickly obtaining personal data to be used for identity theft and for personal information to be out in the open. A home user may not feel they have anything that a criminal would want, “I consider the detail of my life so boring to other people that I really couldn’t care less. I’ve got nothing to hide.” (Pogue, 2007)

Protecting the network is the first step in protecting your personal information.

A long-lasting disadvantage when using a password that is easy, common or simple is this leaves the network open for the opportunity of sharing un-wanted information and data without anyone's knowledge. A study done from November 22, 2010 to October 3, 2011 on 2,133 wireless networks, both consumer and corporate, identified the vulnerability the networks can have. The study found that among the consumer user 61% used WPA/WPA2 encryption, 19% used WEP encryption, 11% used default credentials, 6% had no security on their network and 3% used a hidden SSID. The study found that business networks they surveyed had 58% were secured, 22% unsecured and 20% were considered poorly secured. (Botezatu, 2011) The home user may want the network to be easy for them to access, but also need to recognize how easy it is for someone that does not belong on their network to get in.

Another disadvantage of simple passwords, are using passwords that have words from the dictionary in them. A very common attack on networks is a 'dictionary attack'. "A dictionary attack is a technique used to breach the computer security of a password-protected machine or server. A dictionary attack attempts to defeat an authentication mechanism by systematically entering each word in a dictionary as a password." (Janalta Interactive Inc, 2014) Dictionary attacks work well because many users insist on using a complete word from the dictionary for their password because it is easy to remember, however without having special characters with them, this makes the network easier to penetrate.

There are many steps the home user can take to make their home network more secure and the first and easiest step is to create a strong password. Strong passwords consist of a minimum of 7 characters, containing numbers, symbols, and upper and lower case numbers. Do not use your name or the names of members of your family. Do not use the word "password" or variations of that word and do not use your street address. Also do not use the same password for all of your accounts or use complete words from the dictionary. Creating a password with multiple symbols, numbers, upper and lower case letters in a random order is the best way to have the strongest password possible.

Examples of Strong Passwords:

!e2X6ty#$fm%_RsJOk 07gfE#s&kw@PN$gswVW

3.4 Security setup over router

Having a wired home network instead of wireless is also very beneficial when it comes to security. Without a wireless network then your home network is much less susceptible to being infiltrated. However, if wireless must be used, there are multiple settings in the home router that can be changed to increase the level of security. An important setting to change is to not allow administrative access over a wireless connection. Most routers come with this setting defaulted to wireless access enabled but should be deselected (Appendix B1). However doing this requires the user to be physically connected to the router itself to make any future administrative changes.

Changing your router to not allow guest access is another change that can be made (Appendix B2). Disabling this "feature" helps prevent unwanted access and keeps a person from having two options when trying to infiltrate your network. This way there is only one password that needs to be created and memorized and just one less door into your network. For a home network guest access is not needed due to the fact that you can choose who connects to your network. However, for a small business owner they can provide customers free internet access using the guest access. In this situation having guest access would be beneficial to providing good customer service, however it could leave your network more open to a possible attack.

Next is choosing the wireless encryption that you wish to use, WEP, WPA or WPA2. Many routers offer all three in different variations (Appendix B3). You might ask yourself, why would WEP and WPA even be offered if WPA2 is the latest and best option? Well some devices still require the older encryptions so they still leave them as options, however most modern day devices will use WPA2. Some routers have the option to use both WPA and WPA2 to help work with older devices as well.

One feature that many routers offer is a MAC address access list (Appendix B4). With this feature you can deny access to certain MAC addresses or allow a certain amount of MAC addresses to have access. At first glance this would seem like a very secure option for your home network, however many critics will say that this is one of the dumbest ways to secure a wireless LAN because all someone needs to do is watch an authorized person go in and the they can forge a MAC address with that person's address. The MAC address is just a 12 digit long HEX number that can be viewed in clear text with a sniffer (Ou, 2007).

Another feature that many new routers have is the device list. With this list you can see what devices are connected to your network or have been connected to your network in the past. You can click on them and look up the MAC address of the device to figure out if it is one of yours, and if it is you can label it as such so for future reference when you look at it, you will know which devices are yours (Appendix B5). If you see a device that is unknown to you, you can then attempt to track down the MAC address (maybe it was your friend's computer) and if you cannot, then you can take appropriate measures, such as changing your password.

Another security feature that most routers have is built-in firewalls (Appendix B6). "A firewall is a software or hardware-based network security system that controls the incoming and outgoing network traffic by analyzing the data packets and determining whether they should be allowed through or not, based on applied rule set" (Wikipedia, 2012). Assuring that the router firewalls are enabled can help secure your network traffic and prevent unwanted traffic from entering your network from the internet. Many home computers will have built in firewalls as well that will assist in securing the computer itself, but the firewalls in the router helps prevent malicious activity from even reaching the computer.

3.5 Virtual Private Network

Now that we have gone over strong passwords and typical home router settings, let's look at what the consumer market can offer the average home user. The first item we will discuss is a Virtual Private Network or VPN. A VPN provides an encrypted "tunnel" into the internet from your personal computer out to the internet. So if there is anyone watching your network they will be unable to view what you are doing since it is encrypted and they would have to crack the encryption itself before being able to view your information. Using a VPN service will provide the ability block the users IP address allowing them to browse the internet anonymously. This would prevent the ability of websites and internet services from tracking your web browsing habits (Private Internet Access). This can also bypass an internet service provider's firewall that blocks peer-to-peer networks like uTorrent. So there are many different uses for using a VPN, but a more secure internet browsing capability, in the home or public networks like at your local coffee shop, is a very enticing reason.

There are many different providers out there offering VPN service so the user has different options on what they want to get out of the service and how much they are willing to pay. There are some services that are free, however I wouldn't recommend those because free things in this world, like on the internet, will usually come with something else as well, like an abundance of advertisements. So I would recommend paying for an actual service. The VPN service I chose is available to numerous operating systems as well from Android on smart phones, to Windows, UNIX/Linux, and Mac OS (See Appendix B7). This offers various different users the ability to have this service making it easier on consumers.

3.6 Private Internet Access

One provider that is fairly inexpensive and very user friendly is Private Internet Access (PIA). We recently went through the process of setting up their VPN service on our personal computer and we are very satisfied with the product thus far. One key aspect that really caught our attention was that they offered one of the lowest prices but also the payment options they offered. The user can use a credit card like most online purchases, but they offer a much more secure option for method of payment, and that is a gift card. They have many different gift cards that are acceptable to include: Wal-Mart, Starbucks, Home Depot and Best Buy.

PIA has made the purchase and installation process very simple. If the user chooses to use a gift card they can click on the link for it, input the card information, then a window pops up telling them what the balance on the card is and how many days of access it will give them. If the user chooses to use it, they will enter their email address, which is the only piece of personal information required, then the new username and password is sent to you. After that, downloading and installing is a breeze with detailed instructions and the program virtually installs by itself. After the installation is complete, the VPN will connect automatically, if the user chooses, when they log into their computer, and they can even choose which region or country they would like to "connect" too.

3.7 Encrypt the hard drive

Another option to safeguard the files on your computer is to encrypt the hard drive. There is actually a free download to make this possible from truecrypt.org. With this you can completely encrypt your computer's hard drive or just a small portion of it that you can designate the size of (Appendix B8), this can be used on external hard drives as well. This option can be very useful if you constantly carry an external hard drive with you that has sensitive information on it such as: court documents, bank statements, job applications, even family photos. If you were to lose it and it is not protected, then the person that finds it now has access to all of that information contained on that hard drive. But using TrueCrypt can prevent that from happening. When setting it up you can even use triple the amount of encryption using 3 different encryption algorithms with 256 bit keys making it virtually un-crackable (Appendix B9).

Another feature of TrueCrypt is that you can have multiple "partitions" on a single hard drive. With this you can separate your files into groups, encrypting them differently and creating new passwords for all of them. Doing this takes protecting your personal information and important files to the extreme and you could be rest assured that your files are secured. Breaking a 256 bit key encryption is difficult enough but to do that 3 times with 3 different algorithms would take someone months to do. The encryption process is done as you work as well. "Files are automatically being decrypted on the fly (in memory/RAM) while they are being read or copied from an encrypted TrueCrypt volume. Similarly, files that are being written or copied to the TrueCrypt volume are automatically being encrypted on the fly (right before they are written to the disk) in RAM" (TrueCrypt).

Once the user downloads the program and goes through the steps to set up their "partitions" TrueCrypt is fairly simple to use. When TrueCrypt is open, you find the file source that you wish to access, click mount, enter your password, and then the file storage is mounted (Appendix B10) and the drive that the file storage is mounted to is shown (Appendix B11). Once this is completed, the user now has full access to their files; they can modify them, delete them or even add more. When the user is finished, just simply select the drive that you wish to dismount, click dismount, and that's all. Now you do not have access to your files and they are once again secure. These files cannot be accessed until the file storage is mounted back to the drive and the password is entered. Otherwise the would-be attacker is left to attempt to crack the encryption that has been placed on the files.

4 Conclusions

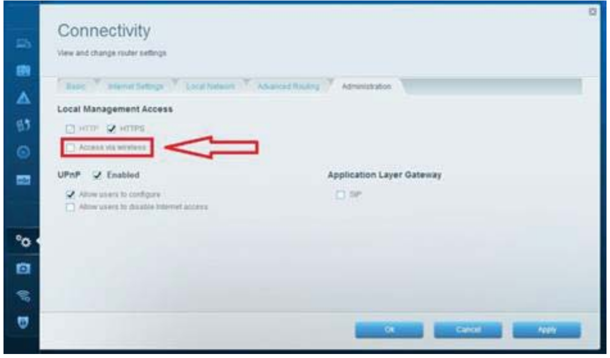
The internet was originally not designed with security in mind and therefor upgrades in cyber security will always been ongoing. Computer hackers and malicious viruses and worms will always be around and constantly evolving to the new changes to security. It is nearly impossible to completely protect your computer and home network, however in this paper we have shown you different ways to make the user's computer and home network less susceptible to attack. Every user has the freedom to choose what they want to do when it comes to their own cyber security and we hope that we have provided enough information to educate users on different security measures that can be taken.

5 References

1. Botezatu, B. (2011, October 11). *25 Percent of Wireless Networks are Highly Vulnerable to Hacking Attacks, Wi-Fi Security Survey Reveals*. Retrieved from Hot for Security: http://www.hotforsecurity.com/blog/25-percent-of-wireless-networks-are-highly-vulnerable-to-hacking-attacks-wi-fi-security-survey-reveals-1174.html
2. Case, L. (2010, May 11). *PC World*. Retrieved from The Ultimate Guide to Home Networking: http://www.pcworld.com/article/196049/the_ultimate_guide_to_home_networking.html
3. Century Link. (2014). Retrieved from Century Link: http://www.centurylink.com/home/specialoffers/?pid=P_601820177&PSTN=Q&siclientid=8652&sessguid=1f452c42-73ba-4e82-af4e-8a59c49c5c63&userguid=1f452c42-73ba-4e82-af4e-8a59c49c5c63&permguid=1f452c42-73ba-4e82-af4e-8a59c49c5c63
4. Charter Communications, Inc. (2014). *Internet*. Retrieved from Charter: https://www.charter.com/browse/internet-service/internet#internet-features
5. Charter Team. (2014). Letter.
6. Janalta Interactive Inc. (2014). *Dictionary attack*. Retrieved from Techopedia: http://www.techopedia.com/definition/1774/dictionary-attack
7. Netgear. (2013, November 27). *Netgear Support*. Retrieved from What is the difference between 2.4 GHZ and 5 GHz?: http://kb.netgear.com/app/answers/detail/a_id/24246/~/what-is-the-difference-between-2.4-ghz-%26-5ghz%3F
8. Netgear. (2014). *Networking*. Retrieved from Netgear: http://www.netgear.com/home/products/networking/
9. Ou, G. (2007, April 2). *The Six Dumbest Ways to Secure a Wireless LAN*. Retrieved from ZD Net: http://www.zdnet.com/blog/ou/the-six-dumbest-ways-to-secure-a-wireless-lan/43
10. Pogue, D. (2007, January 4). How secure Is Your Wi-Fi connection? Retrieved from http://pogue.blogs.nytimes.com/2007/01/04/04pogue-email/?_php=true&_type=blogs&_php=true&_type=blogs&_r=1
11. Private Internet Access. (n.d.). Retrieved from Private Internet Access: https://www.privateinternetaccess.com/pages/how-it-works/
12. Smith, M. (2014, January 21). *Top 25 most commonly used and worst passwords of 2013*. Retrieved from NetworkWorld: http://www.networkworld.com/community/blog/top-25-most-commonly-used-and-worst-passwords-2013
13. TrueCrypt. (n.d.). *Documentation/Introduction*. Retrieved from TrueCrypt: http://www.truecrypt.org/docs/
14. Verizon. (2014). *Support*. Retrieved from Verizon: www.verizon.com/Support/Residential/internet/highspeed/networking/setup/questionsthree/85834.htm#
15. What Is My IP Address. (2014). *Whats is my IP address*. Retrieved from Dynamic IP vs Static IP: http://whatismyipaddress.com/dynamic-static
16. Wikipedia. (2012, Fedruary 24). *Firewall*. Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Firewall_(computing)
17. Wikipedia. (2014, February 24). *Wi-Fi Protected Access*. Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Wi-Fi_Protected_Access
18. Wikipedia. (2014, February 23). *Wired Equivalent Privacy*. Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Wired_Equivalent_Privacy
19. Wilson, T. J., & Fuller, J. (2001, April 30). *How Home Networking Works*. Retrieved from How Stuff Works: http://computer.howstuffworks.com/home-network.htm
20. Wright, J. (2006, September 11). *Explaining WPA2*. Retrieved from NetworkWorld: http://www.networkworld.com/columnists/2006/091106-wireless-security.html

Appendix B

Appendix B1



Connectivity

View and change router settings

Local Management Access

□ HTTP ☑ HTTPS

□ Access via wireless

UPnP ☑ Enabled

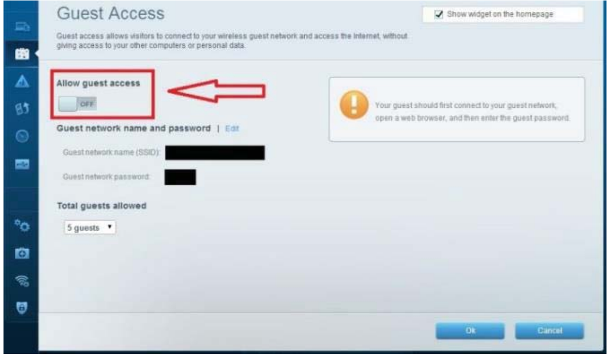
☑ Allow users to configure

□ Allow users to disable Internet access

Application Layer Gateway

□ SIP

Appendix B2



Guest Access

☑ Show widget on the homepage

Guest access allows visitors to connect to your wireless guest network and access the Internet, without giving access to your other computers or personal data.

Allow guest access

OFF

Your guest should first connect to your guest network, open a web browser, and then enter the guest password.

Guest network name and password | Edit

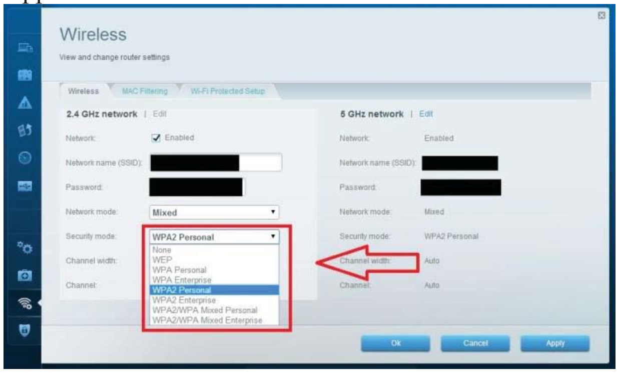
Guest network name (SSID): [redacted]

Guest network password: [redacted]

Total guests allowed

5 guests

Appendix B3



Wireless

View and change router settings

2.4 GHz network | Edit

Network: ☑ Enabled

Network name (SSID): [redacted]

Password: [redacted]

Network mode: Mixed

Security mode: WPA2 Personal

None
WEP
WPA Personal
WPA Enterprise
WPA2 Personal
WPA2 Enterprise
WPA2/WPA Mixed Personal
WPA2/WPA Mixed Enterprise

Channel width:

Channel:

5 GHz network | Edit

Network: Enabled

Network name (SSID): [redacted]

Password: [redacted]

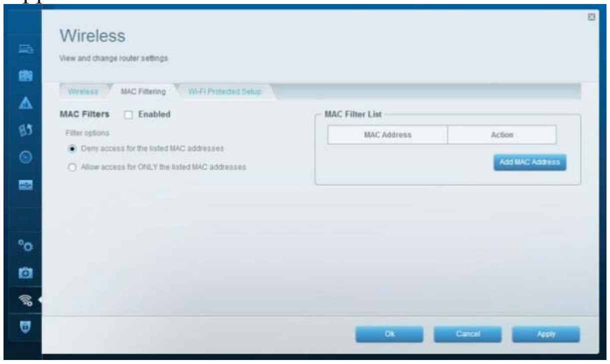
Network mode: Mixed

Security mode: WPA2 Personal

Channel width: Auto

Channel: Auto

Appendix B4



Wireless

View and change router settings

MAC Filters □ Enabled

Filter options

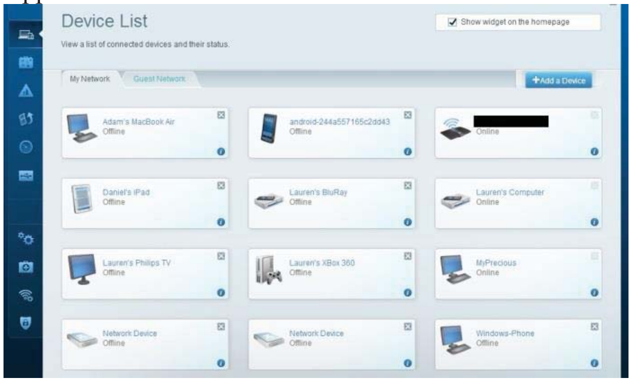
● Deny access for the listed MAC addresses

○ Allow access for ONLY the listed MAC addresses

MAC Filter List

MAC Address	Action

Appendix B5



Device List

☑ Show widget on the homepage

View a list of connected devices and their status.

Adam's MacBook Air
Offline

android-244a557165c2dd43
Offline

[redacted]
Online

Daniel's iPad
Offline

Lauren's BluRay
Offline

Lauren's Computer
Online

Lauren's Philips TV
Offline

Lauren's XBox 360
Offline

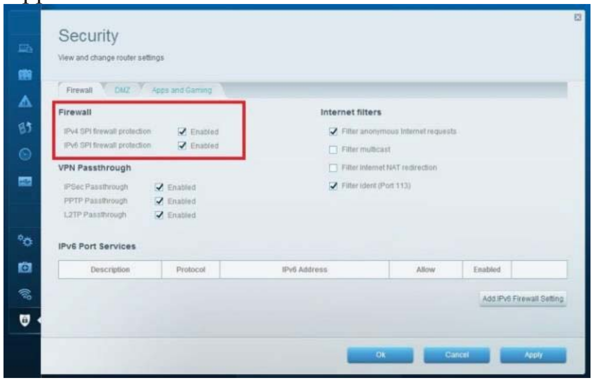
MyPrecious
Online

Network Device
Offline

Network Device
Offline

Windows-Phone
Offline

Appendix B6



Security

View and change router settings

Firewall

IPv4 SPI firewall protection ☑ Enabled

IPv6 SPI firewall protection ☑ Enabled

VPN Passthrough

IPSec Passthrough ☑ Enabled

PPTP Passthrough ☑ Enabled

L2TP Passthrough ☑ Enabled

Internet filters

☑ Filter anonymous Internet requests

□ Filter multicast

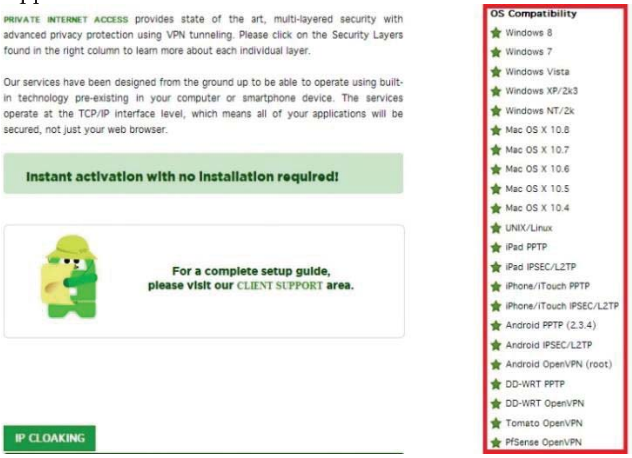
□ Filter Internet NAT redirection

☑ Filter ident (Port 113)

IPv6 Port Services

Description	Protocol	IPv6 Address	Allow	Enabled	

Appendix B7



PRIVATE INTERNET ACCESS provides state of the art, multi-layered security with advanced privacy protection using VPN tunneling. Please click on the Security Layers found in the right column to learn more about each individual layer.

Our services have been designed from the ground up to be able to operate using built-in technology pre-existing in your computer or smartphone device. The services operate at the TCP/IP interface level, which means all of your applications will be secured, not just your web browser.

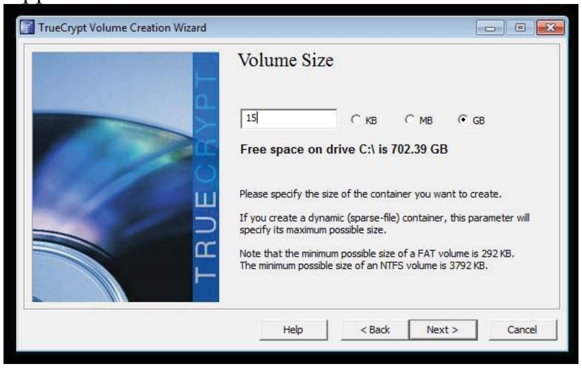
Instant activation with no installation required!

For a complete setup guide, please visit our CLIENT SUPPORT area.

OS Compatibility

- Windows 8
- Windows 7
- Windows Vista
- Windows XP/2k3
- Windows NT/2k
- Mac OS X 10.8
- Mac OS X 10.7
- Mac OS X 10.6
- Mac OS X 10.5
- Mac OS X 10.4
- UNIX/Linux
- iPad PPTP
- iPad IPSEC/L2TP
- iPhone/iTouch PPTP
- iPhone/iTouch IPSEC/L2TP
- Android PPTP (2.3.4)
- Android IPSEC/L2TP
- Android OpenVPN (root)
- DD-WRT PPTP
- DD-WRT OpenVPN
- Tomato OpenVPN
- PfSense OpenVPN

Appendix B8



TrueCrypt Volume Creation Wizard

Volume Size

15 ○ KB ○ MB ● GB

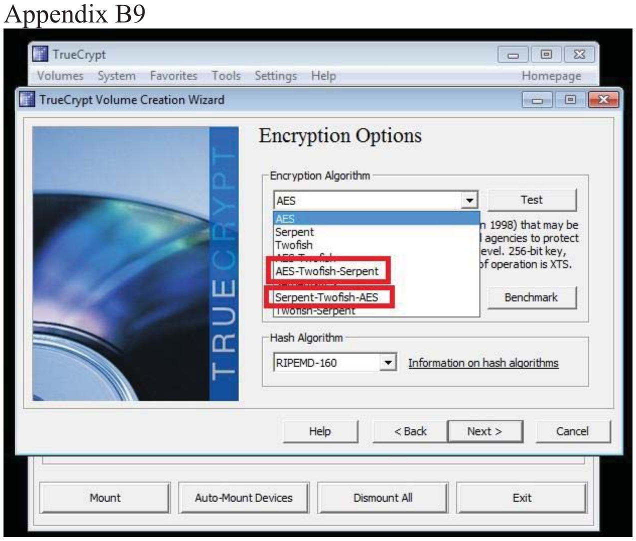
Free space on drive C:\ is 702.39 GB

Please specify the size of the container you want to create.

If you create a dynamic (sparse-file) container, this parameter will specify its maximum possible size.

Note that the minimum possible size of a FAT volume is 292 KB. The minimum possible size of an NTFS volume is 3792 KB.

Appendix B9



TrueCrypt Volume Creation Wizard

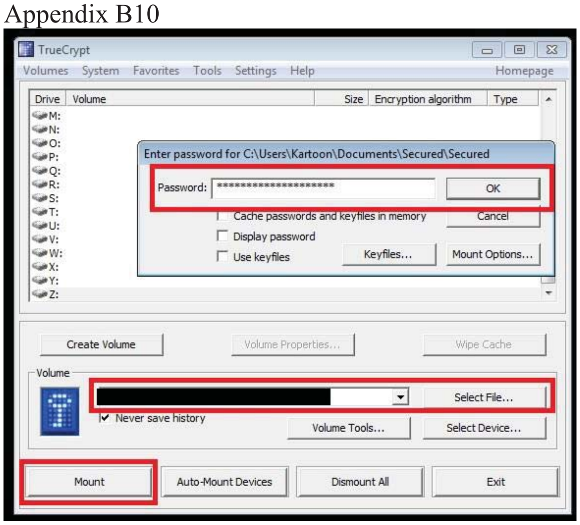
Encryption Options

Encryption Algorithm: AES

n 1998) that may be
l agencies to protect
evel. 256-bit key,
of operation is XTS.

Hash Algorithm: RIPEMD-160

Appendix B10



Enter password for C:\Users\Kartoon\Documents\Secured\Secured

Password: ********************

□ Cache passwords and keyfiles in memory

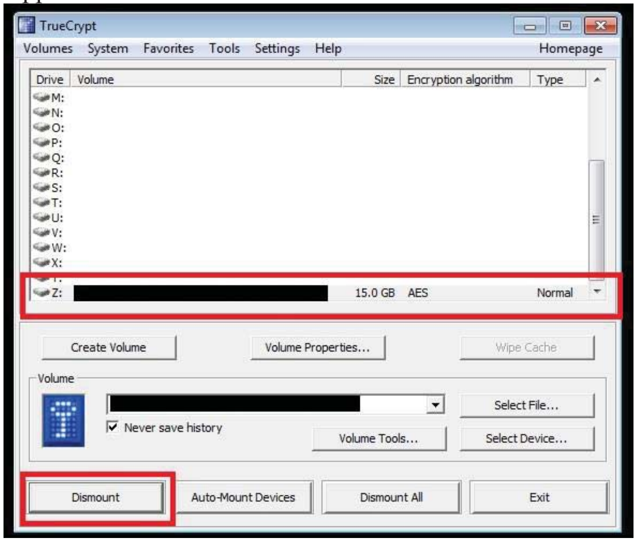
□ Display password

□ Use keyfiles

Volume: [redacted]

☑ Never save history

Appendix B11



Drive	Volume	Size	Encryption algorithm	Type
M:				
N:				
O:				
P:				
Q:				
R:				
S:				
T:				
U:				
V:				
W:				
X:				
Z:	[redacted]	15.0 GB	AES	Normal

Volume: [redacted]

☑ Never save history

Preserving Source-Location Privacy in Wireless Sensor Networks against a Global Eavesdropper

Zhiwen Zeng[1], Xiaoyan Hu[1], Zhigang Chen[1]
[1]School of Information Science and Engineering
Central South University
Changsha 410083, China

Hui Liu[2] *
[2]Department of Computer Science
Missouri State University
Springfield, MO 65897, United States

***Abstract* —While many works to date in wireless sensor networks (WSNs) security have focused on providing confidentiality for message contents, contextual information usually remains exposed. Thus the adversary especially the global eavesdropper can easily obtain the sensitive information such as the location of a target object in a monitoring application, which is critical to the mission of the sensor network. In this paper, we propose an energy-efficient scheme against global eavesdroppers for protecting location privacy and maximizing lifetime of WSNs. Our proposed technique builds fake source areas controlled by the Sink, which disseminates dummy traffic synchronized with real source area. Therefore, we create multiple candidate traces in the network to hide the real traffic generated by the real source. To ensure no impact on lifetime of WSNs, we minimize the energy consumption of hotspots. Analysis and simulation results show that the proposed scheme can significantly improves the security in source-location privacy preservation without reducing the network lifetime.**

Keywords—source-location privacy; wireless sensor network; global eavesdroppers; fake source area;

I. INTRODUCTION

A wireless sensor network is composed of numerous cheap, small, energy-constrained, and spatially distributed autonomous sensors to monitor and study the physical world. It has gained more popularity in recent years and been widely used in lots of important applications such as environment monitoring, military surveillance, and target tracking. Sensors collaborate to gather data and disseminate the data to the sink. Since most of sensors are deployed in unattended or hostile environment, these applications are subject to a variety of security issues.

Among all of these security threats, source location privacy is of special interest to us since it cannot fully addressed by traditional security mechanisms, such as encryption and authentication. Consider a simple example of target tracking in WSNs. A sensor sends a message which includes event-related information to the sink when it detects an event. After this, the location of the event source has been exposed to the adversary who might passively monitor the network traffic, no matter how strong the data encryption key is. Further, the adversary may find out more sensitive information: whether, when and where a particular event occurred, for example, the appearing of a soldier in a target tracking sensor network [1, 2]. This can help the adversary in attacking the soldier, an unfortunate occurrence.

In the past two decades, a number of researches aiming at protecting source location privacy have been proposed. We can generally divide these works into two categories according to the capability of the adversaries: strategies against local eavesdroppers and those against global eavesdroppers. The local eavesdroppers have limited coverage, comparable to that of regular sensors. At any given time only a local area is under the adversary's monitoring and the adversary tires to locate the source node hop-by-hop in a tracing back way. However, global eavesdroppers are much more powerful and formidable. They are able to monitor all the network traffic either by deploying their own cheap sensors that cover the whole area [3] or by employing a powerful site surveillance device with hearing range no less than the network radius.

Several methods have been proposed to preserving the source-location privacy against local eavesdroppers. For example, [4-7]. Kamat et al. [4] propose a classic location protecting protocol based on Phantom Routing. Firstly every message is randomly routed for h hops to find a phantom source, and then the selected phantom source sends the message to the SINK node by flooding. However, both theoretical and practical results demonstrate that if the message is routed randomly for h hops, the message will be largely within $h/5$ hops away from the actual source [5]. Xi et al. [6] propose a two-way greedy random walk named as GROW. In GROW, greedy random walk first creates a static random walk (path of receptors) from the sink node. Subsequently, messages are sent from the source node on a greedy random walk that will eventually arrive at a receptor node, from which the message will be forwarded to the SINK node following the established path. In order to well balance the energy consumption and privacy protection, J. Ren et al. [5] propose a two-phase routing scheme. The source node randomly determines an intermediate node from a pre-determined region around the SINK node called the Sink Toroidal Region (STaR). From the random intermediate node, the message will then be routed to the sink node through the shortest path routing. In STaR, the entire network is divided into grids. One node in every grid is denoted as the head node. The source node randomly selects one grid in the constrained area, of which the head node becomes the random intermediate node. In [7], the authors propose a three-phase routing scheme that addresses the source-location privacy issue by using Radial Routing and Annular Routing based on the phantom source.

* The corresponding author: huiliu@missouristate.edu.

The strategies mentioned above all have outstanding performances in location privacy preservation against local eavesdroppers. However, in order to effectively defend powerful global eavesdroppers, we have to design more suitable schemes to preserve location privacy against them.

Mehta et al. [2, 8] first presented the global eavesdropper model, and proposed two techniques, called periodic collection and source simulation to prevent the leakage of location privacy. The main idea of the periodic collection method is to make the traffic pattern independent of the presence of real objects. However, this approach consumes a substantial amount of energy for latency sensitive applications and largely reduces the lifetime of network since each node periodically sends packets at a reasonable rate regardless of whether it has real data to send or not. The source simulation method creates multiple candidate traces in the network to hide the traffic generated by real objects by simulating the movement patterns of real objects. However, this approach has two problems. First, it is challenging to model the movement patterns of real objects. Second, fake sources also bring much extra energy consumption to the hotspots of the network, which reduces the network lifetime.

The scheme based on proxy filtering is illustrated by Yang et al. [9], in which they select some sensors as proxies that proactively filter dummy packets on their way to the sink. Then, they proposed two mechanisms named as PFS (Proxy-based Filtering Scheme) and TFS (Tree-based Filtering Scheme). Because determining proxies (i.e. selection P elements out of V nodes) is an NP-hard problem as proved in [9], they adopted local search heuristics with no guaranteed maximal network lifetime.

Bicakci et al. [10, 11] investigate the network lifetime in various different proxy assignment strategies and different deployment scenarios. They propose a new filtering idea called OFS (Optimal Filtering Scheme) to maximize the network lifetime of wireless sensor networks while preserving event-unobservability against global eavesdroppers. Through a Linear Programming framework, they claimed that Linear Programming is an effective method to find the optimal locations of proxies under a set of linear constraints.

Recently, Ju Ren [12] adopts cluster structures to construct cyclic interference routing paths, in which the cluster heads will act as proxies to filter fake message generated by the fake source. They allow only real traffic flow to cross the hotspots and build cyclic diversionary routing paths in areas where the sensors have enough abundant energy to support them to maximize the network lifetime.

All of the above-mentioned researches do not solve the problem that the change of the real source location is ahead of that of the fake source location, therefore, the adversary easily detect the location of real source because of the abnormal operations of nodes. In this paper, we use the sink, the center of gathering data in the network, to guarantee the synchronization of data dissemination of real source and fake source, then ensure the preservation of location privacy. Our proposed scheme can achieve a significant improvement in network security without reducing the network lifetime since we take advantage of residual energy in non-hotspots. At the same time, we also consider to minimize the data transmission latency as one of our optimal goals.

The remainder of this paper is organized as follows. In Section II, the system model is described. Details of the proposed location privacy scheme against global eavesdroppers are illustrated in Section III. Section IV compares and analyzes the performances of our scheme based on the simulation studies. Section VI concludes the paper.

II. MODELS

A. Network Model

In this paper, we adopt a homogeneous network model, in which all of the sensors have roughly the same capabilities, power sources, and expected lifetimes. This is a common network architecture for many applications today not only because it is very simple for deployment and maintenance but also it has been well-studied and provides for relatively straightforward analysis. We make the following assumptions about our network model.

- A wireless sensor network is deployed with equal density throughout a circular region with the radius R. The whole network is fully connected through multi-hop communications [7]. The only SINK node is located at the center of the circular network that is the destination location that data packets will be routed to.
- The appearance of the object is randomly distributed in the entire network, so the probability that each sensor detects the information of the object as well as sends data to the SINK is equivalent. We suppose that the time that the message delivered by the farthest sensors in the entire network arrives the SINK is denoted as T, and the sensing radius of the sensor is r.
- We assume that a security infrastructure has already built in; that is, no information carried in the message will be disclosed. The key management, including key generation, key distribution, and key update, is beyond the scope of this paper.

B. Adversary Model

In this paper, we adopt the adversary model given in [12]. Adversaries are assumed to be external, passive and global. More precisely stated, adversaries cannot compromise or control any sensors. They do not conduct any active attacks such as traffic injection, channel jamming, or denial of service attack. However, adversaries can listen to all communication in the network, analyze the collect data and try to determine the location of each sensor node. We assume adversaries deploy their own sensors that cover the target WSNs in order to achieve eavesdropping all the communications of the whole target WSNs.

C. Energy Consumption Model

Energy consumption model [13] is adopted in this paper. We consider only the energy usage of transmitting and receiving messages. Energy consumption for transmitting messages is shown in equation 1, and then equation 2 shows the energy spent for receiving a *l*-bit packet.

$$\begin{cases} E_{member} = lE_{elec} + l\square_{fs}d^2 & d < d_0 \\ E_{member} = lE_{elec} + l\square_{amp}d^4 & ifd > d_0 \end{cases} \qquad (1)$$

$$E_R(l) = lE_{elec} \qquad (2)$$

where E_{elec} is transmitting circuit loss. When the distance *d* between transmitter and receiver is less than the threshold d_0, the free space (d^2 power loss) channel model is considered. Otherwise, the multi-path fading (d^4 power loss) channel model is adopted. $\square_{fs}$ and $\square_{amp}$ are the energy required by power amplification in these two models, respectively. The above parameter settings are given in Table 1 [13].

TABLE I. NETWORK PARAMETERS

Parameter	Value
Threshold distance (d_0) (*m*)	87
Sensing range r_s (m)	15
E_{elec} (nJ/bit)	50
$\square_{fs}$ (pJ/bit/m^2)	10
$\square_{amp}$ (PJ/bit/m^4)	0.0013
Intial Energy (J)	0.5

$b = log_2 \frac{|S_T|}{|S_P|}$ is used to measure the level of privacy preservation in [2]. Depending on the users and applications, this can be easily modified to support different kinds of privacy measurement models. For example, we can define high, medium and low privacy levels using appropriate values of *b*. S_T is defined as a set of sensors that represent the set of possible locations for the objects sensed by the target network. Let S_P be the number of real sources which the adversary finds, then the adversary knows the location of target object. For example, the periodic collection method [2] has every sensor node independently and periodically send packets at a reasonable frequency regardless of whether there is real data to send or not, thus S_T is a set of all the sensors in the network while S_P is the set of potential real sources determined by the adversary. The source simulation method [2] randomly selects a set of sensors and pre-load each of them with a different token. These tokens will be passed around between sensors to simulate the behavior of real objects, thus S_T is a set of sensors which simulate the real objects while S_P is the number of real objects which is 1 in most scenarios.

III. THE ROUTING SCHEME BASED ON FAKE SOURCE AREA (FSA)

In this section, we illustrate our proposed routing scheme based on Fake Source Area (FSA) for location privacy preservation and lifetime maximization in wireless sensor networks. The principles of FSA can be summarized as the following three aspects. Firstly, FSA builds a number of fake source areas to generate interference data gathering areas by using the abundant energy in areas far from the sink, because the sensors in these areas always remain much energy when the network dies. Secondly, since only the center in every fake source area and real source area transmits data to the SINK, the number of transmitted packets as well as the energy consumption in hotspots is reduced, therefore our method increases the network lifetime. Finally, the SINK has knowledge of the movement pattern of real source sensors through data collections, it broadcasts the change to all the sensors and asks to change the locations of centers. Thus, from the view of the adversary, all the centers change at the same time which enhances the difficulty to detect the real source location. The main idea of FSA can be detailed as the following five phases.

A. Deployment and initilization of network

Before deployment, we randomly select a set of sensor nodes and pre-load each of them with a different token. Every token has a unique ID. For convenience, we call the node holding a token the token node. Token nodes are considered as fake sources and centers of fake source areas. The change of token nodes means the change of fake source areas. The number of token nodes indeed represent the number of fake source areas. The random deployment of token nodes can guarantee that when the real object appears anywhere in the network, it will be treated as same as the random fake sources. The forming of real source area and fake source areas are described as the following steps. 1) Each token node and real source act as the centers by broadcasting one data packet, which includes the location of centers and the number of hops, to all the neighboring nodes. For example, if we define the size of area is three hops within the centers, the number of hops will be set as 3 initially. Therefore, the number of hops starts from 3, it decreases by 1 once the data packet is forwarded one more hop. When the number of hops reaches zero, the data packet will be discarded. Only sensors within the area can receive this data packet. 2) When the neighboring nodes receive the broadcasted data packets, they know they are members of this area centered at the corresponding token node. Definitely it is possible that one sensor receives multiple data packets sent by different token nodes, the sensor will select the closest token node among them as its center. So far, fake source areas and real source area are established.

B. Data gathering within areas

Only real source transmits the real packet while all the other nodes send fake packets. As shown in Fig. 1, the data transmission is initiated by the most outside nodes, and continues to the centers within all the areas. Every sensor node checks whether the packet includes real data or dummy traffic when it receives one packet. If the packet is just dummy traffic, the senor node directly discards the dummy traffic and generate a new fake one to send. Otherwise, the sensor node forwards real data to its neighbors until data packet finally arrives the center of its belonged area.

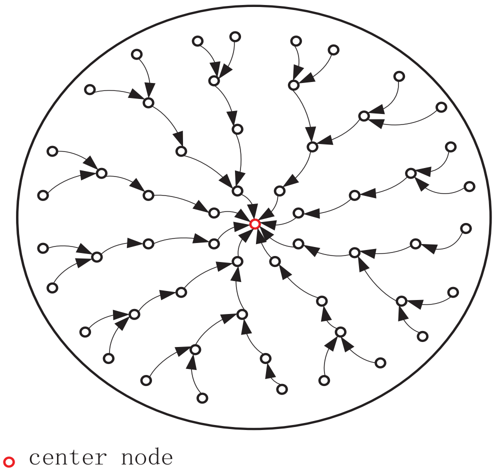


Fig. 1, the data gathering within a single area

C. Data transmission from centers to the SINK

After the centers gather packets from their areas, the centers determine whether they obtain real packet. If the center has the real packet, it will transmit the real packet to the SINK along the shortest path between itself and the SINK. Otherwise, it discards gathered dummy packet, generates one new fake packet, and forwards the fake packet to the SINK also along the shortest path between itself and the SINK. Since the size of every area is fixed and every center sends packets to the SINK almost at the same time after they have gathered data from their area, all the packets sent by the centers arrive the SINK simultaneously. The adversary cannot tell the difference between real source area and random fake source areas because they are synchronized with each other in the aspect of data transmission. Eventually the SINK receives the real data packets which include data related to the tracking target such panda and location information of real source. Then, the SINK will decide whether the fake source areas need to move in next round by analyzing the change information of real source location.

D. The changes of fake source areas

After the first round of data collection of the SINK as shown in Fig. 2, the SINK obtains the real source location. Our method continues the second round of data collection without making any changes of fake source areas. However, after the second round of data collection, the SINK conducts two possible actions according to the changes of the real source location. First, the real source location remains the same as before, which means the target object does not moves at all or maybe only moves within the monitoring area of the previous real source sensor. FSA continues next round of data collection without changes of fake source areas. Second, the real source location changes, which means the target object already moves out of the monitoring area of the previous real source node. So we can predict the real source area definitely changes in next round of data collection. But we can conclude that the next real source should be in one hop area away from the current real source under our assumption that the speed of target object is not very fast and the target object should still be in the range of one hop to two hops distance from the current real source. In a summary, the current fake source randomly select one node within one hop area from itself as the new fake source as the new center of fake source area in order to synchronize the movement patterns of fake source areas and real source area. The token will then be passed to the new selected fake source from the previous fake source. The delivery of such taken between sensor nodes will be always protected by the pairwise key established between them.

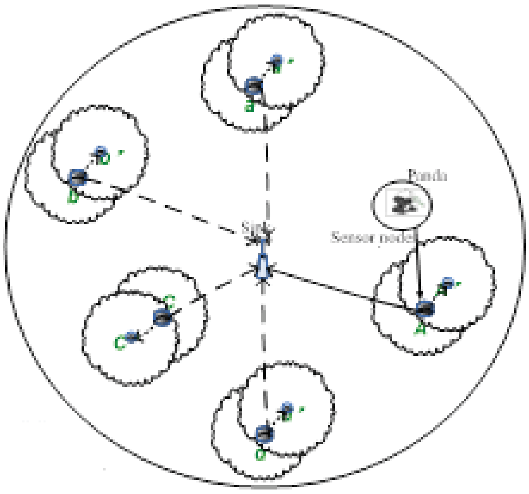


Fig. 2, the data transmission model in the network

E. New fake source areas

After the selection of centers of new fake source areas, FSA continues the next round of data collection according to the above phases *A*, *B*, *C*, *D*, and *E*.

IV. SIMULATION AND PERFORMANCE ANALYSIS

We use the discrete event-based simulator-OMNET++ [14] to simulate the two protocols, FSA and Source Simulation, comparing their energy consumption and network lifetime in order to achieve a certain level of location privacy *b*.

A. Simulation Environment

Assume there are 4000 sensor nodes uniformly distributed in a circular area with the radius of 600 meters, *R*. Every node can communicate with those nodes no more than 50 meters far from it given the assumption that the sensing radius is as the same as the transmission radius $r = 50$ meters. The node density of the network is 0.0035 sensors per square meters. We also suppose the size of source areas is 2 hops and the number of fake source areas is 10. Only one target object appears in the network. The initial location of the target object is random. If the target object moves from coordinates (x,y)

to $(x \pm \Delta x, y \pm \Delta y)$, and $\Delta x^2 + \Delta y^2 \leq r^2$. That means within the interval of two rounds of data collection, the moving distance of the target object is less than that of one hop.

B. *Performance Analysis*

It is known that the better the level of privacy preservation, the more the number of fake source areas or the larger the radius of fake source areas. Our experiments study the communication cost and network lifetime in order to achieve a certain level of location privacy preservation.

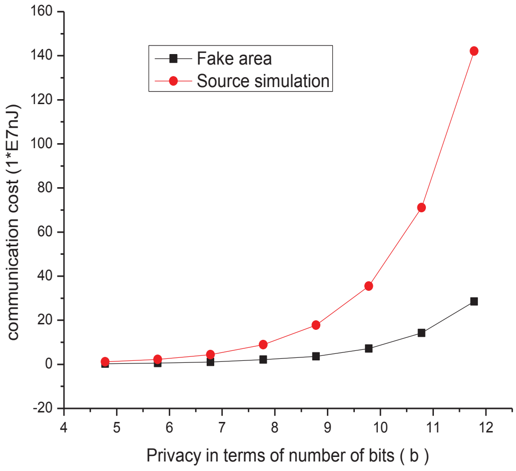


Fig. 3 the comparison of energy consumptions corresponding to different level of location privacy preservation.

From Fig. 3 we observe that the energy consumption of source simulation method [2] is 4 or 5 times as high as our proposed scheme FSA in order to achieve the same level of privacy preservation. Because most nodes only transmit data in short distance in FSA while fake sources in [2] need to send data all the way to the SINK which involve more intermediate nodes and consume more energy compared with FSA.

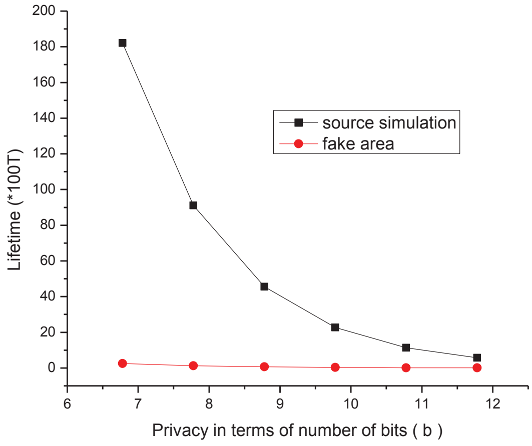


Fig. 4 the comparison of network lifetime corresponding to different level of location privacy preservation.

FSA has two ways to enhance the location privacy preservation, one is to increase the number of fake source areas, and the other is to increase the size of fake source areas. As we know that the network lifetime depends on the energy consumption of the hotspot and has no direct relationship to the whole network energy consumption. Therefore, increasing the size of the fake source areas do not reduce the network lifetime while the increasing the number of the fake source areas directly reduce the network lifetime. In order to prove the advantages of FSA, we adopt to increase the number of fake source areas to enhance the privacy preservation in our simulation experiments. As shown in Fig 4, the network lifetime of FSA is almost 80 to 90 times as long as the network lifetime in source simulation method since FSA fake sources gather packets within areas before sending packets to the SINK, which largely decreases the number of packets sent to the SINK, reduces the number of data transmission in hotspots, and then improves the network lifetime.

V. CONCLUSION

Source-location privacy is significantly important to the successful deployment of wireless sensor networks. In this paper, we propose a novel routing scheme based on the fake source areas for protecting source-location privacy against global eavesdroppers named as FSA. We carry out extensive simulation experiments to compare our proposed method with other existing schemes. Our simulation results demonstrate that the proposed FSA can effectively enhance the privacy preservation as well as maximize the network lifetime.

REFERENCES

[1] P. Kamat, Y. Zhang, W. Trapper, C. Ozturk, "Enhancing source-location privacy in sensor network routing", in ICDCS'05: Proceedings of the 25th International Conference on Distributed Computing Systems, Ohio, USA, June 2005.

[2] K. Mehta, D. Liu, and M. Wright. "Location privacy in sensor networks against a global eavesdroppers", in ICNP, 2007.

[3] Y. Zhu and R. Bettati. "Compromising location privacy in wireless networks using sensors with limited information". In ICDCS 2007.

[4] C. Ozturk, Y. Zhang, and W. Trapper, "Source-location privacy in energy-constrained sensor network routing", in SASN'04: Proceedings of 2004 ACM Workshop on Security of Ad Hoc and Sensor Networks, Washington, DC, USA, October 2004.

[5] Leron Lightfoot, Yun Li and Jian Ren, "Preserving Source-Location Privacy in Wireless Sensor Network using STaR Routing", in Globecom 2010: Proceedings of 2010 IEEE Global Communications Conference, Miami, FL, USA, December 2010.

[6] Y. Xi, L. Schwiebert, and W. Shi, "Preserving Source Location Privacy in Monitoring-Based Wireless Sensor Networks", Parallel and Distributed Processing Symposium, Rhodes Island, Greece, April 2006.

[7] Zhiwen Zeng, Meili Zeng, Hui Liu, "Source-Location Privacy Protection in Wireless Sensor Networks Using AZR Routing", accepted by The 2014 International Conference on Wireless Networks, July 21-24, 2014, Las Vegas, NV.

[8] K. Mehta, D. Liu, and M. Wright. "Protecting location privacy in sensor networks against a global eavesdropper", in IEEE transactions on Mobile Computingm vol. 11, no. 2, pp. 320-336, 2012.

[9] Y. Yang, M. Shao, S. Zhu, and G. Cao, "Towards event source unobservability with minimum network traffic in sensor networks",

Proceedings of the ACM Conference on Wireless Network Security, 2008, pp. 77-88.

[10] K, Bicakci, I. E. Bagci, and B. Tavli, "Lifetime bounds of wireless sensor networks preserving perfect sink unobservability", IEEE Communications Letters, vol. 15, no. 2, pp. 205-507, 2011.

[11] K, Bicakci, H. Gultekin, B. Tavli, and I. E. Bagci, "Maximizing lifetime of event-unobservable wireless sensor networks", Computer Standards and Interfaces, vol. 33, no. 4, pp. 401-410, 2011.

[12] J. Ren, Y. Zhang, and K. Liu, "An energy-efficeint cyclic diversionary routing strategy against global eavesdroppers in wireless sensor networks", Internation Journal of Distributed Sensor Networks, Vol. 2013.

[13] Xianyan Meng, Anfeng Liu, Zhigang Chen, "Strategy of energy balance based on data transfer for wireless sensor networks", Computer Engineering and Applications, vol. 47(1), 2011, pp. 116-119.

[14] OMNet++ Network Simulation Framework, http://www.omnettp.org/

Multimodal Biometrics as Attacks Measure in Biometrics Systems

Ifeoma U.Ohaeri, Michael Esiefarienrhe, Naison Gasela
oh.ifeoma@yahoo.com, Department of Computer Science North-West University,
Mafikeng Campus,
Private Bag X2046, Mmabatho 2735, South Africa.
Michael.Esiefarienrhe@nwu.ac.za, Department of Computer Science North-West University,
Mafikeng Campus,
Private Bag X2046, Mmabatho 2735, South Africa.
Naison.Gasela@nwu.ac.za, Department of Computer Science North-West University,
Mafikeng Campus,
Private Bag X2046, Mmabatho 2735, South Africa.
*Corresponding author: Ifeoma.U.Ohaeri

Abstract: *Multimodal systems are considered to be the most dependable among the other types of biometric systems because it is more difficult to forge multiple traits than a single one. Hence, having a fingerprint multi-algorithmic system would not make authentication more secure considering an attacker trying to gain unauthorized access provided with a fingerprint replica. Therefore, multimodal systems are widely deployed and the most generally accepted.*

Therefore, in this review paper we proposes an integration of fingerprint and face recognition systems to provide anti-spoofing and replay attacks measures making it difficult for an impersonator/imposter to steal multiple biometric traits of a genuine user.

Keywords: Biometric, Multimodal Systems, Verification, and Authentication.

1. Introduction

Biometric system is automated recognition of persons based on their biological and behavioral characteristics. This technique of physical and logical access controls to protect information systems from security threats such as spoofing and replay attacks is becoming increasingly popular compared to traditional token-based or knowledge-based techniques. One of the main reasons for this popularity is the ability of the biometrics technology to differentiate between an authorized person and an intruder who fraudulently acquires the access privilege of an authorized person. Two primary uses of biometrics would be identity verification and user identification. Identity verification is a one-to-one comparison of a person's 'biometric template' with his or her original previously stored in the system. The verification result is a "match"/"no-match", or a similarity measure or class membership degree [1].

Impersonation is a very big security threat to biometric systems. This is performed by the use of artifacts or by finding an existing person with a similar biometric data and then fraudulently assuming that identity to spoof a verification check. points out that a biometric-based verification system works properly only if the verifier system can guarantee that the biometrics data came from the legitimate person at the time of enrollment so that during verification when a user claims an identity it is validated by comparing the stored biometric data against their presented biometric features [2]. For example, wolf attacks use a biometric sample such that the similarities between this sample and a number of templates are resulting in high false matches with these templates; wolf attacks took advantage of vulnerabilities on the specific matching algorithms to achieve their purposes [1]. However, the assassination of Al-Mabhouh, a co-founder of the military wing of Hamas in 2010 highlights the risk that sophisticated attackers can undermine existing identification systems by targeting individuals for impersonation. Interpol and the Dubai police believe the suspects which are up to 29 stole the identities of real people [3]. It is therefore important to examine the accuracy of biometric tools when subjected to such attacks. If biometric systems are to prevent these attacks, the systems need to be made complex for impersonations or impostors by combining more than one biometric data for verification and authentication purposes.

This incessant attack on biometric systems has increased demand of enhanced security systems consequently has led to an unprecedented interest in biometric based person authentication system. Biometric systems based on single source of information are called unimodal systems. Although some unimodal system have got considerable improvement in reliability and accuracy, they often suffer from enrollment problems due to non-universal biometrics traits, susceptibility to biometric spoofing or insufficient accuracy caused by noisy data. Hence, single biometric may not be able to achieve the desired performance requirement in the real world applications. One of the methods to overcome these problems is to make use of multimodal biometric authentication systems, which combine information from multiple modalities to arrive at a decision.

The use of multiple biometric traits (multimodality) is related to capturing more than one biometric characteristic. Some of the commonly used biometric characteristics are: face, fingerprint, hand geometry, palm print, iris, keystroke, voice and gait. However, fingerprint multimodal systems make user authentication more effective because impostors do not gain unauthorized access. Hence, in this review paper we propose Multimodal Biometrics as effective Attack Measures in Biometric Systems in other to guarantee quality of service.

2. Problem Statement

With the widespread deployment of biometric systems in a variety of applications, there are increasing concerns about the security and privacy of biometric technology using single biometric trait [4]. Despite the several advantages offered by these systems, there are challenges to be considered, such as noise in the sensed data, intra-class variations, inter-class similarities, non-universality, interoperability issues and spoof attacks. Therefore, a biometric system that employs one single biometric characteristic is constrained because all of these affect the performance, security and convenience of using such a system. Hence, there is a need to develop a system that are more reliable and provide higher verification rates due to the presence of multiple, independent pieces of evidence [5].

3. Related Work

Several studies concerning integration of two or more biometric traits can be found in the literature [6] [7]. Advantages of using joint face and fingerprint biometrics have appeared in a number of studies. The proposed multimodal biometric method is designed to enhance real time verification and reliability rates by going beyond the technical limitations of single biometric verification methods. Laplacian face is used for face recognition, DFB for fingerprint recognition, and an artificial neural network employed to combine the two [8].

A multimodal authentication model using face and fingerprint on space specified tokens is reported. Space specified tokens require little data and have been used to minimize the data storage. Face mages are encrypted and encoded into fingerprint images. The verification accuracy provides a solution from spoofing and other attacks [9].

Fusion of face and fingerprint considers the fingerprint quality information and compared four different fusion methods. The authors also showed that fusion of multimodal biometrics using data quality information outperforms standard multimodal results and unimodal systems [10].

The matching scores of three traits (face, fingerprint and hand geometry) to enhance the performance of a biometric system are combined. Three different techniques (sum rule, decision tree, and linear discriminant analysis) are used to combine the matching scores. Experiments indicate that the sum rule with normalized scores results in the best performance [11].

Another example of correlated work, which noticeably resembles ours, is the multimodal (face and fingerprint) biometrics verification system proposed and implemented by. Such a system is embedded in an Operating System (OS Linux kernel 2.4.26), in a way that the OS gets sensitive to the presence of the user and, therefore, protect users' login sessions by continuously verifying their presence on this protected computer. The authors point out that their theory can be readily extended to include more modalities in the future, considering that they propose and use a new type of score-level fusion [12].

Fingerprint and face biometric authentication systems using state-of-the-art commercial off-the-shelf products are studied in and it is confirmed that multimodal biometric systems outperform single biometric ones. In addition to examining well-known methods, a new normalization and multimodal fusion method is introduced based on matching score level fusion. The suggested normalization and fusion methods are used for authentication applications that deal with open populations (e.g. airports) [13].

A novel system for identity authentication based on multi-route decision is presented in [14]. This system also uses face and fingerprint information, and includes three modules; an enrollment module, an image preprocessing module and a fusion module. An SVM authentication fusion strategy distinguishes real clients from imposters, depending on self-learning results of the SVM from the enrollment module.

Multimodal biometrics uses a novel face and fingerprint fusion feature modeling approach, and explores using of ridge let and discrete wavelet transforms coupled with different classifiers. The confidence values from the classifiers are fused using various methods to decide the best method for classifier combination [15].

Some state-of-the-art methods are studied. The concept of "uncertainty region" is proposed as a novel serial scheme. This system considered that the case of the first matcher is incapable to obtain enough evidence to classify the subject matching scores. A simple mathematical model is used to simplify the design of the processing chain, and compared serial scheme performance with the best individual matcher [16].

4. Methodology

In this research paper, we shall use the following research methodologies:

4.1 Literature Survey

A thorough literature survey will be conducted on previous research activities within our study with a view to improving our acquaintance with the subject matter and to serve as groundwork to our present investigation.

4.2 Model Formulation

A model will be developed based on the literature survey to serve as anti-spoofing and replay attacks measures.

4.3 Algorithm Development

Also, an anti-spoofing and replay attacks algorithm in a bimodal biometric systems to provide higher verification rates for fraud detection will be developed based on the literature survey.

4.4 Metric Development

A performance metric will be developed to evaluate the performance of the model formulated and algorithm developed.

4.5 Model Implementation

There will be provision of evidence on applicability of the model by showing results in a form of simulation.

4.6 Model Evaluation

The performance of the implemented model will be evaluated. Some of the metrics used to evaluate the model will be False Accepted Rate, False Rejected Rate, Failure to Enroll Rate and its susceptibility to artifacts or mimics.

5 Conclusion

Conclusively, the combination of multiple biometrics tends to improve the performance of the system and reduce the rate of attacks.. We have presented in general the Introduction, Literature Review, and the Methodology of the proposed Multimodal Biometrics as Attack Measure in Biometrics systems. The design and implementation of this system will reduce the high rate of attacks in biometrics systems in other to ensure efficiency and quality of service of biometrics systems.

Future Work

As a work in progress we intend to design and implement a multimodal biometrics system that combines, fingerprint and face characteristics to form a measure of verification and authentication of any system user before an access is either granted or denied. This shall enhance the effectiveness of the biometrics system that is being widely adopted around the world.

References

[1] Tetsushi Ohki, SeiraHidanol& Tatsuya Takehisa. "Evaluation of Wolf Attack for Classified Target on Speaker Verification Systems" Research Institute for Science and Engineering, Waseda University, Japan, vol10, issue 5, page 336.347, 2012.

[2] B. Scheneier.. "The Uses and Abuses of Biometrics Communication." ACM, vol.42, issue No 8, page 615- 625, 1999.

[3] "Bbc news," http://news.bbc.co.uk/l/hi/world/middle_east/8522595.stm.

[4] U. Uludag and A. K. Jain. "Attacks on biometric systems: a case study in fingerprint." In Security, Steganography, and Watermarking of Multimedia Contents'04, vol. 45, issue no 7, pages 622–633, 2004.

[5] Anil K. Jain &Arun Ross. "Learning User-Specific Parameters in a Multibiometric System." Department of Computer Science and Engineering,

Michigan State University, vol 30, issue 5, pages 556-565, 2000.

[6] Abate, A.F, Nappi, M., Riccio. D., and Marsico. M. "Face, Ear and Fingerprint Designing Multibiometric Architectures." 14th IEEE International Conference on Image Analysis and Processing (ICIAP), vol 5, issue 10, pages 778-789, 2010.

[7] Varchol, P., Levickly, D. and Juhar, J. "Multimoda biometric authentication Using Speech and hand geometry fusion." 15th International Computer on System Signals and Image processing (IWSSIP), vol. 15, issue 5, 2008.

[8] Aloysius George. "Bizare Approaches for Multimodal Biometrics."IJCSNS, International Journal of Computer Science and Network Security, vol. 8, issue no7. July 2008.

[9] B. Prasana Lakshmi and AKannammal. "Secured Authentication of Space Specified Token with Biometric Traits Face and Fingerprint." IJCSNS International Journals of Computer Science Network Security, vol., 9, issue 45, July 2009.

[10] Yu-Chiang Wang. and David Casasent. "Multimodal Biometric Fusion Using Data Information." Proceedings of International Security of Photo- Optical Instrument Engineers (SPIE, 2005.

[11] A. Ross, A.K. Jain and J. Qian. "Information Fusion in Biometrics" in Processings AVBPA 01 Halmstad Sweden,. pp 354- 359, June 2001.

[12] Sim T., Zhang S. Janakirama, R and Kumar, S. "Continuous Verification Using Multimodal Biometrics."IEEE Transaction 0n Pattern Analysis and Machine Intelligence, vol, 11, issue no. 5, April 2007.

[13] Robert Snelick, UmutUludag, Alan Mink, Michael Indovina and Anil Jain. "Large Scale Evaluation of Multimodal biometric Authetication Using State-of-the Art System". IEEE Transaction on Pattern, 2005 .

[14] Jun Zhou, Guangda Su, Chunhong Jiang, Yafeng Deng, and Congcong Li. "A Face and Fingerprint Identity authentication system based on Multi-route detection."Neurocomputing Elsevier Science Publishers. Volume 70 Issue 4-6, Pages 922-931 January, 2007.

[15] Djamel Bouchaffra and Abbes Amira, "Structural Hidden Markov Models for Biometrics: Fusion of Face and Fingerprint Pattern Recognition", vol.41, issue 3, pp. 775-777 2008.

[16] G. Zhao and M. Peitikainen."Pattern Analysis and Machine Intelligence", vol. 29, issue no. 6, 915-928, 2007.

SESSION

AD-HOC NETWORKS, MANET

Chair(s)

TBA

An Efficient Crowdsourcing Search Scheme in Vehicular Ad hoc Networks

Chyi-Ren Dow, Duc-Binh Nguyen, Zi-How Lin and Shiow-Fen Hwang
Department of Information Engineering and Computer Science
Feng Chia University, Taichung, Taiwan
{crdow, p0395608, m012313, and sfhwang }@fcu.edu.tw

Abstract *In recent years, Vehicular Ad-hoc Networks (VANETs) have become a popular research field. In addition, conditions happened surrounding our life often have the demand to be supported by other people, and these conditions often occur on the road. This study proposes an efficient crowdsourcing search scheme in VANETs to search objects and collect data. Our proposed mechanism is based on the virtual backbone construction in VANETs. The regional data exchange can be limited, and the packet transmissions can be reduced. We have established coordinator and header mechanisms to manage the information of region nodes, reduce the amount of packet transmission and improve searching efficiency. Experimental results show that our schemes can effectively provide assistance in terms of search efficiency and satisfactory ratio.*

Keywords: VANETs, crowdsourcing, search scheme.

I. INTRODUCTION

Vehicular Ad-hoc Networks are emerging network architecture using wireless network technologies on the intelligent transportation systems (ITS) for mobile information communications. There are many VANETs applications, such as cooperative monitoring of traffic load, prevention and warning of vehicle collision, and location based services (LBS) combined with the nearby area information, etc.

In addition, a model for problem solving is also developed with the Internet. Jeff Howe, a reporter of the Wired magazine has created a whole new jargon “Crowdsourcing” in 2006. Enterprise organizations solicit a large number of volunteers to help them to gather information, provide ideas and solve problems about technical through the Internet. Volunteers usually complete tasks in their spare time and charge a small reward or not, but there may be some opportunities to get more rewards in the future. This solution provides a new way of organizing labors, especially for the software industry and the service industry. For example, Wikipedia [20] is an Internet encyclopedia, and everyone can participate in online editing.

In the past, when people got off a taxi and forgot something in the car, it is not only hard to find the lost property but also wasting time, especially when they do not know the taxi fleet and license plate number. Currently, some taxi fleets and radio stations provide services to help people finding their missing property online. However, it is not immediate when the messages are broadcasted until the driver finds the lost property and returns it. In recent years, with the development of the Internet and the popularity of smart phones, people began to help each other through exchanging messages over the network such as looking for an accident escape, animal abusing, bullying and other cases. These Internet users can assist in finding suspects as long as they have enough information.

In fact, the concept of the crowdsourcing already exists in our daily life. Crowdsourcing mechanisms have been widely applied in many fields since created, and there were many applications developed with sensors on smart phones and tablets. However, in VANETs with crowdsourcing, how to choose the crowd to assist the task and communicate effectively to the demand for them, and how to make proper filtering and screening to ensure the quality and accuracy of the returned data by the crowd, are important issues about crowdsourcing. In addition, the concept of traditional practice of distribution of demand and recovery of data is similar to the broadcast or multicast schemes. It is easy to generate large amounts of streaming data. The influence is perhaps not obvious because there is enough bandwidth in the wired network architecture, but it is a serious problem in VANETs with limited bandwidth.

If we want to provide applications with effective crowdsourcing, we need to reduce the time of request spread and data recovery. Both are closely related to the amount of packet transmission. Anycast [10] is a method of information spreading on the network. Its concept is that if there is anyone who can provide the required services, they will be accepted after the demand spread. This method not only can reduce the amount of packet transmission, but also can choose appropriate service providers. Thus, anycast is more suitable for applications in bandwidth limited wireless network infrastructure compared with the broadcast scheme.

Considering these issues, the application and technology of taxi based VANETs with a group structure and high density to find and track the target in a city are worth a discussion among us. Therefore, we propose an efficient crowdsourcing search scheme in VANETs. It is a distributed searching system based on virtual backbone and geography information in VANETs. It uses vehicles on the backbone to play the role of "crowd", and apportions the searching tasks to the crowd through crowdsourcing in VANETs. The information of vehicles is integrated and exchanged with headers via coordinators in a geographic grid [9]. When there are searching demands generated, our system will ask the coordinators first, and then forward the packet to headers and request vehicles to join the searching task.

II. RELATED WORK

The topic of data searching has been engaged in network research. Lakas et al. [8] proposed a hybrid cooperative through cooperating vehicles using the store-and-forward technique to share collected information. Noguchi et al. [14] proposed a location-aware service discovery scheme. It spreads service discovery messages to nodes inside a geographical area with IPv6 multicast. Some research schemes [6], [13] improved AODV to propose methods to optimize the route of discovery. Lo et al. [12] enabled vehicles to cooperatively aid the source node to discover the location of the destination node without the support of location services. The geographic load balancing routing, namely GLRV is deployed to provide a virtual backbone. GLRV can increase delivery ratio and reduce the transmission latency in hybrid VANETs architecture.

How to process the distribution of a vast number of information from different locations is an important issue. Data aggregation technique aims to solve redundant or distributed data problem to improve the communication efficiency. Tal et al. [16] analyzed various solutions by using Fuzzy Logic in data aggregation schemes is suitable to get benefits in the development process of traffic systems that relies on these schemes. Zhang et al. [19] presented a hierarchical data aggregation scheme to reduce the transmitting of the redundant data which resulted from multi-source data collecting and multi-path data transmitting. Traditional data aggregation methods usually rely on a fixed routing structure to ensure data to be aggregated. However, they cannot be used in highly mobile vehicle environments. Catch-Up [17] dynamically changed the forwarding speed of nearby reports so that they had a better chance to meet each other and be aggregated together. Dietzel et al. [4] proposed a generic model to describe and classify the proposed schemes. DA2RF [18] is an infrastructure-free data aggregation scheme by restricting forwarders to limit the number of forwarders in VANETs. In this way, transmission collisions can be avoided as much as possible.

With the increasing popularity of smart phones in recent years, there are many studies using crowdsourcing techniques, including the sensing data collection, transportation issues, data accuracy, and other interesting applications. CrowdITS system [1] used the smart phones to collect sensing data without additional sensors and communication equipment and make improvements to traffic by collection of traffic information. CrowdOut [2] is based on contributions made by mobile users equipped with smart phones. It allows users to report traffic offense in real time and to map them on a city plan. There are systems sought volunteers to mine the disaster sensing dead space by crowdsourcing [3] and create a noise map in the urban environment through mobile phones with crowdsourcing [7]. Huang et al. [5] presented selection methods of automatic sensor based on crowdsourcing models for unattended acoustic sensor selection. Furthermore, passengers can also be an effective statistical evaluation of the traveling road conditions through the triaxial sensors in the phone [15]. Liu et al. [11] combined people and mobile sensing devices into a live wireless sensor. It uses this combination to remedy the traditional sensor blind, and also conforms the concept of crowdsourcing.

III. THE PROPOSED SEARCHING METHOD

In this section, we introduce our proposed searching mechanisms, including virtual backbone establishment of the tree searching architecture and the design of the crowdsourcing anycast query spreading mechanism. Then, we formulate crowdsourcing data filtering mechanisms.

A. The Virtual Backbone of the Tree Searching Architecture

Traffic could change very often with complex road structure in urban environments as time goes by different distances, directions and speed of the vehicle will cause the network topology changes and affect the stability of the chain. If the city roads are divided by geographic grids to establish a backbone tree, we can limit the area of data exchange of information and reduce unnecessary traffic packets to achieve fast and stable data transmission. Exchanging and maintaining the data table between important nodes in the backbone can simplify data storage and management. The data can be found more quickly, and the searching time can also be reduced with precise management.

The vehicle stays in each geographic grid longest will be elected as a header to manage the information of other vehicles in the grid to reduce the packet number of switched data. Normally, the vehicle closest to the center of the grid will be elected as a coordinator. As shown in Figure 1, $D_i = \sqrt{(x_c - x_i)^2 + (y_c - y_i)^2} < r$, where x_c and y_c are coordinates of the center, x_i and y_i are coordinates of the vehicles in the same grid. r is the range of transmission to consider a vehicle as the header in the grid. As shown in Figure 1, we calculate the distance between each vehicle and the center, the header of the grid is D1 because it has the smallest distance. Headers within grids regularly gather information of vehicles in the same grid.

As shown in Figure 2, the coordinator within the grid gathers information from adjacent grids in order to accelerate the speed of searching, but it will generate a lot of queries and return packets when there are too many coordinators. Therefore, development an appropriate number of coordinators is an important issue. As shown in Figure 3, considering the degree of branching of the tree and beginning with the smallest number of grids. If the degree of the header within the grid is greater than 2, it will be elected as the coordinator. Therefore, this step may lead to a coordinator having too many nearby coordinators. For example, coordinator within grids 22 and 54 are superfluous, because there are too many coordinators.

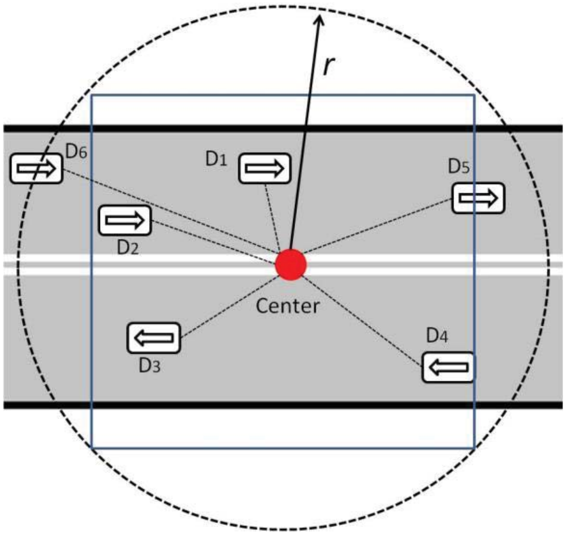


Figure 1 Grid Header Election

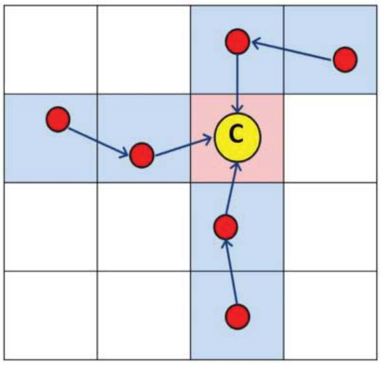


Figure 2 Coordinator Collection

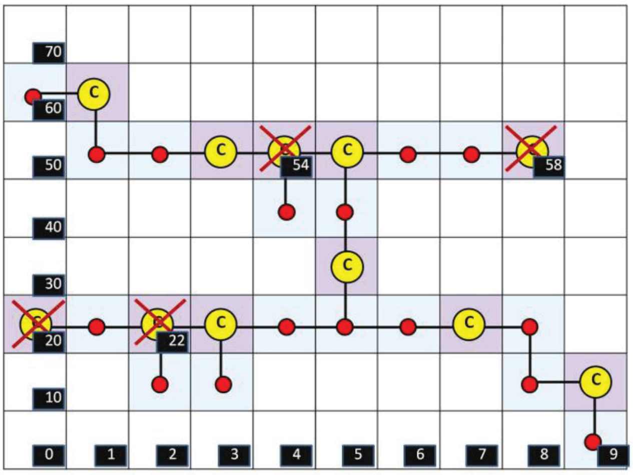
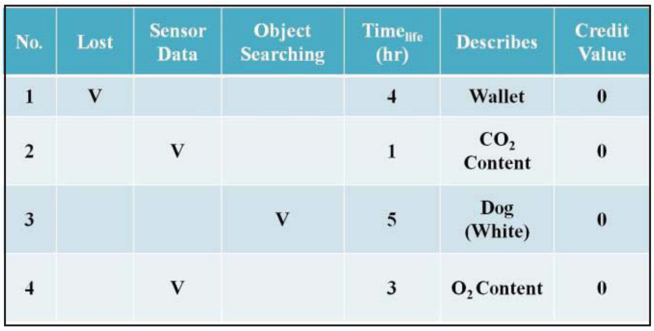


Figure 3 Coordinator Establishment

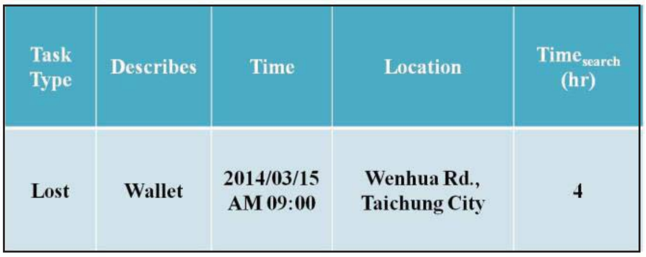
In order to avoid collecting unnecessary information from other paths, a coordinator is created every three hops. However, the coordinator cannot be established if it is located in the leaves of the tree (grids 20 and 58) to avoid excessive control overhead and reduce excess coordinators. Because grids 55 and 58 collect information on the same area, they will result in a large amount of duplicated information and redundant of transmission.

B. The Crowdsourcing Anycast Query Spreading Mechanism



No.	Lost	Sensor Data	Object Searching	$\text{Time}_{\text{life}}$ (hr)	Describes	Credit Value
1	V			4	Wallet	0
2		V		1	CO_2 Content	0
3			V	5	Dog (White)	0
4		V		3	O_2 Content	0

Figure 4 The Reply Table



Task Type	Describes	Time	Location	$\text{Time}_{\text{search}}$ (hr)
Lost	Wallet	2014/03/15 AM 09:00	Wenhua Rd., Taichung City	4

Figure 5 The Query Packet

The header or the coordinator within each grid maintains a reply table as shown in Figure 4 which contains detailed data fields of each informed object. Vehicles will periodically reply the information of board objects (such as lost and sensor data, etc.) to headers within grids, and forward the data to coordinators for management. When there is a task (object searching or data collection, etc.), our system will generate query packets as shown in Figure 5 to send to the nearby coordinator. These query packets contain the purpose of task (lost, target searching, data collection, etc.), content (including target characteristics, time, location, etc.). The coordinator receiving the query packet will be compared with the reply table, and then return to the source node. If the task needs to cooperate for finding the target, the header within the grid near the target area asks the vehicle within the grid for joining the task after receiving the query packet. When any vehicle responses content, the task starts. When the task is completed, the data will be returned by the header and coordinator. If there is no matching information, the coordinator passes the query packets to other coordinators through the adjacent headers within the grids, and they compare with the reply table. In order to maintain the effectiveness of the searching mechanism, we have developed the threshold of searching time to avoid searching too long or endless searching. We will terminate searching and return results if time is over.

Our crowdsourcing search mechanism is listed as follows:

(1) Coordinators regularly update the reply form.

(2) Source node generates a query packet, and delivers to coordinators for comparing data. In case of the data is matched, the coordinator returns to the source node.

(3) If the data is not matching, the query packet will be passed to other coordinators for comparing data by neighbor headers.

(4) After receive the query packet, the header of each neighbor will ask other vehicles if willing to assist or not.

(5) The task starts when any vehicle responses content.

(6) When the task completes, messages will be responded to the source node via headers and coordinators.

(7) If searching time exceeds T_{search}, the task will be terminated.

C. *The Filtering Mechanism*

Although we have designed the query packet format and reply mechanism to provide the format of returned data. However, getting the number of requirements from the return data is also a problem. Hence we need to formulate the filtering mechanism to deal with these situations. If we allow the transmitted of data to be collected together, it would cause a considerable burden for the network. Thus, we must filter redundant information through hierarchical steps to reduce traffic and leave useful information.

We divide data filtering into three types. For the first type, the searching condition given from the source node to the coordinator is precision and the searching condition given from the coordinator to the header is fuzzy. By this way, we can reduce the difficulty of searching by vehicles and increase the number of data. The coordinator can filter those data returned by headers before passing to the source node. The second type) where conditions given from the source node, and the coordinator are both precise. It increases the difficulty of searching by vehicles but reduces the amount of data and reduces the load of networks. The third type represents the conditions given from the source node and the coordinators are both fuzzy, the data returned by vehicles will not be filtered by the coordinator, and the source node can select the information which one does it wants.

When the new beginning data is received, our mechanisms allow the same information received within the time threshold T for maintaining accuracy and reliability of the information. The number of repetitions will be recorded into the reply table and form a credit value. The higher credit value represents the higher reliability of the data. Each cumulative data will get a survival time, T_{life} after time T, and it is not allowed to accept the same data. When the time T_{life} ends, the extra data will be discarded or replaced by new data.

IV. EXPERIMENTAL RESULTS

This section is carried to evaluate our proposed crowdsourcing search scheme. We use the version 2.35 of Network Simulator 2 (NS-2) as our simulator. The simulation vehicular movement mode is generated by Simulation of Urban Mobility (SUMO).

As shown in Figure 6, we compare the control overhead with the time change. At the beginning, since the purpose of our method is to establish the backbone, the amount of control packets required will be greater than AODV and Geogrid. However, once the backbone structure and information forms' contents are completed, the needed amount of control overhead will gradually stabilize.

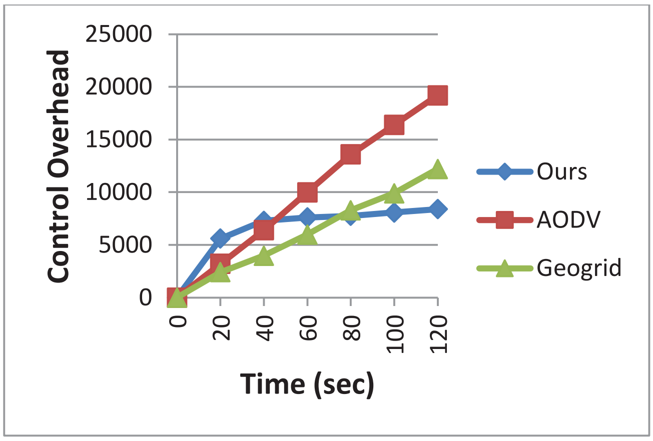


Figure 6 Control Overhead vs. Time

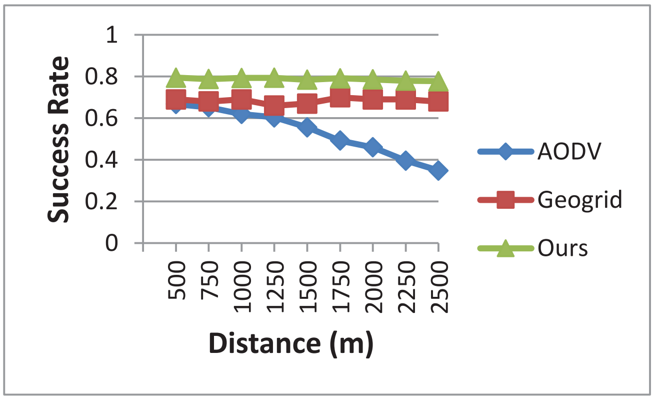


Figure 7 Successful Rate vs. Distance

As shown in Figure 7, the success rate is compared to the change in distance. The range for distance is 500 to 2500. The success rate of AODV decreases as the distance increases, because AODV always starts a new search. Therefore, the success rate will decrease when the distance increases. On the other hand, when Geogrid perform a search, every point is looking for the nearest coordinator for inquiries. Thus, the success rate obtained will display a stable status. In our method, every coordinator works together to maintain and exchange the reply tables. This reduces the search time, which also raises the level of efficiency. In comparison, Geogrid can find the target more effectively.

V. CONCLUSIONS

In this study, we propose an effective crowdsourcing search scheme in VANETs. It can reduce the number of transmission packets through the establishment of coordinator and header, and provide crowdsourcing search of the query instead of blind searching. The proposed scheme not only improves the results of discovery, but also reduces dissemination time and reply time. Design of the table information can help to exchange all reply discoveries.

As the future work, we intend to implement our crowdsourcing search scheme. However, not everyone is willing to participate in crowdsourcing tasks. Maybe we can combine incentives to promote the wishes of people, providing them with virtual or real currency. Furthermore, we will consider more methods for both data aggregation and data filtering in order to achieve more perfect results.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Science and Technology of the Republic of China for financially supporting this research under Contract No. 103-2221-E-035-057- .

REFERENCES

[1] K. Ali, D. A. Yaseen, A. Ejaz, T. Javed and H. S. Hassanein, "CrowdITS: Crowdsourcing in Intelligent Transportation Systems," *Wireless Communications and Networking Conference*, Paris, pp. 3307-3311, Apr. 2012.

[2] E. Aubry, T. Silverston, A. Lahmadi and O. Festor,"CrowdOut: A Mobile Crowdsourcing Service for Road Safety in Digital Cities," *Pervasive Computing and Communications Workshops*, Budapest, pp. 86-91, Mar. 2014.

[3] E. T. Chu, Y. L. Chen, J. Y. Lin and J. W. S. Liu, "Crowdsourcing Support System for Disaster Surveillance and Response," *Symposium on Wireless Personal Multimedia Communications*, Taipei, pp. 21-25, Sep. 2012.

[4] S. Dietzel, J. Petit, F. Kargl and B. Scheuermann, "In-Network Aggregation for Vehicular Ad Hoc Networks," *Communications Surveys & Tutorials*, Vol. PP, No. 99, pp.1, Apr. 2014.

[5] P. S. Huang, M. H. Johnson, Y. Wotao and T. S. Huang, "Opportunistic Sensing: Unattended Acoustic Sensor Selection Using Crowdsourcing Models," *IEEE International Workshop on Machine Learning for Signal Processing*, Santander, pp. 1-6, Sep. 2012.

[6] F. Iqbal, M. I. K. Babar, M. H. Zafar and M. F. Zuhairi, "I-AODV: Infrastructure Based Ad Hoc On-demand Distance Vector Routing Protocol for Vehicular Ad Hoc Networks," *IEEE International Conference on Smart Instrumentation, Measurement and Applications*, TBD Malaysia, pp. 1-5, Nov. 2013.

[7] S. S. Kanhere, "Participatory Sensing: Crowdsourcing Data from Mobile Smartphones in Urban Spaces," *IEEE International Conference on Mobile Data Management*, Lulea, pp. 3-6, Jun. 2011.

[8] A. Lakas, M. A. Serhani and M. Boulmalf, "A Hybrid Cooperative Service Discovery Scheme for Mobile Services in VANET," *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, Wuhan, pp. 25-31, Oct. 2011.

[9] Y. H. Lee, C. R. Dow, L. H. Huang, Y. C. Lin, S. F. Hwang and W. B. Lee, "An Efficient Geo-aware Peer-to-Peer Resource Discovery and Sharing Scheme in Vehicular Ad-hoc Networks," *Ninth International Conference on Information Technology: New Generations*, Nevada, pp. 54-59, Apr. 2012.

[10] P. J. Lin, C. R. Dow, S. C. Chen, C. J. Li and S. F. Hwang, "An Efficient Anycast Scheme for Discovering K Services in Mobile Ad-hoc Networks," *The 5th ACM symposium on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks*, Vancouver, pp. 33-37, Oct. 2008.

[11] J. W. S. Liu, E. T. H. Chu and P. H. Tsai, "Fusing Human Sensor and Physical Sensor Data," *IEEE International Conference on Service-Oriented Computing and Applications*, Taipei, pp. 1-5, Dec. 2012.

[12] C. C. Lo, J. W. Lee, C. H. Lin, M. F. Horng and Y. H. Kuo, "A Cooperative Destination Discovery Scheme to Support Adaptive Routing in VANETs," *IEEE Vehicular Networking Conference*, New Jersey, pp.202-208, Dec. 2010.

[13] H. Maowad and E. Shaaban, "Efficient Routing Protocol for Vehicular Ad Hoc Networks," *IEEE International Conference on Networking, Sensing and Control*, Beijing, pp. 209-215, Apr. 2012.

[14] S. Noguchi, M. Tsukada, T. Ernst, A. Inomata and K. Fujikawa, "Location-aware Service Discovery on IPv6 GeoNetworking for VANET," *International Conference on ITS Telecommunications*, Petersburg, pp.224-229, Aug. 2011.

[15] C. Song, J. Wu, M. Liu, H. Gong and B. Gou, "RESen: Sensing and Evaluating the Riding Experience Based on Crowdsourcing by Smart Phones," *8th International Conference on Mobile Ad-hoc and Sensor Networks*, Chengdu, pp. 147-152, Dec. 2012.

[16] I. Tal and G. Muntean, "Using Fuzzy Logic for Data Aggregation in Vehicular Networks," *Distributed Simulation and Real Time Applications*, Dublin, pp. 151-154, Oct. 2012.

[17] B. Yu, C. Z. Xu and M. Guo, "Adaptive Forwarding Delay Control for VANET Data Aggregation," *Parallel and Distributed Systems*, Vol. 23, No. 1, pp. 11-18, Jan. 2012.

[18] Y. Yuan, J. Luo, W. Yan, T. Zhao and S. Lu, "DA2RF: A Data Aggregation Algorithm by Restricting Forwarders for VANETs," *Computing, Networking and Communications*, Anaheim, pp. 393-397, Feb. 2014.

[19] L. Zhang and B. Jin, "QoS-Oriented Data Dissemination in VANETs," *Parallel and Distributed Processing Symposium Workshops & PhD Forum*, Shanghai, pp.21-25, May 2012.

[20] Wikipedia.
http://en.wikipedia.org/wiki/Main_Page

Wireless Power Control for Tactical MANET: Power-Rate Bounds

Sarah Lauff, Jeff Allen, Dave Schwartz
SSC PACIFIC, 53560 Hull Street, San Diego, California, United States

Abstract - *Two conflicting objectives of mobile ad-hoc networks (MANET) are maximizing throughput while simultaneously minimizing power. This paper uses the geometry of the Power-Rate (PR) image to compute the best possible tradeoffs between throughput and power that any wireless power control (WPC) algorithm delivers in a given tactical RF scenario. These optimal power-rate tradeoffs are computed for a two-link tactical scenario to benchmark distributed WPC algorithms operating with delayed and noisy measurements.*

Keywords: Wireless Power Control, Tactical MANET, Network Performance, Pareto Front.

1. Performance Bounds

Tactical mobile wireless networks serving small unit ground forces operate with different requirements than commercial systems: The radios are single transceivers; hardware upgrades are not likely; frequencies differ from commercial systems; LPD may be critical so that the noise level may be extreme; battery power is always a problem; jamming will be intelligent and hostile; nodes routinely drop out; messaging priorities vary; and throughput requirements are increasing. These limitations point to software only solutions operating on the existing radio hardware. The Office of Naval Research (ONR) is developing dynamic spectrum access (DSA) for tactical wireless networks [1]. Assessing performance of these algorithms in tactical RF scenarios is a basic task to quantify their performance.

Power and throughput are basic tactical network performance metrics. Well-designed algorithms lead to optimal power-rate tradeoffs [2]. Best possible bounds on the power-rate tradeoffs in tactical RF scenarios provide absolute benchmarks to rank system performance and guide development in these challenging RF scenarios.

This paper develops power-rate bounds for a scenario: a two-link network where one node moves along a straight-line track. These bounds assume centralized control and computation, instantaneous messaging, and error-free measurements. The bounds quantify the loss of performance that distributed power control must incur.

Section 2 introduces the RF urban scenario and two-link system. Section 3 develops the power-rate bounds for a single point on the straight-line track. Section 4 applies these bounds over the entire track. Section 5 makes explicit challenges that distributed adaptive algorithms must overcome in this tactical RF scenario and the narrowband modeling limitations. In particular, the single most important observation is that the substantial gain variations along the track point to robust WPC algorithms [3].

Symbol	Description
G	Channel gain matrix
$\mathbf{p}_C$	Power control vector (W)
$\mathbf{p}_N$	Additive noise power vector (W)
γ_M	Measured SINR vector
γ_X	External or QoS SINR vector

2. Wireless RF Link Model

Figure 1 is a side view of the 3-D city that supports the RF propagation modeling [4]. The buildings are modeled with a flat plate dielectric corresponding to concrete. The ground plane is also a flat plate dielectric corresponding to wet earth.

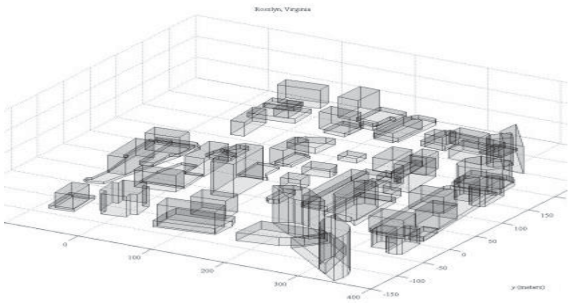


Figure 1: 3-D Urban Model.

Figure 2 is the top view of the city showing the transmitters (TX-1, TX-2) and the receivers (RX-1, RX-2) of the 2-link network. All nodes operate at the street level (2 meters above the ground plane) excepting RX-2. This receiver operates at the edge of the rooftop of the large building overlooking the downtown area. The network is mobile because TX-2 moves along the *x*-axis track.

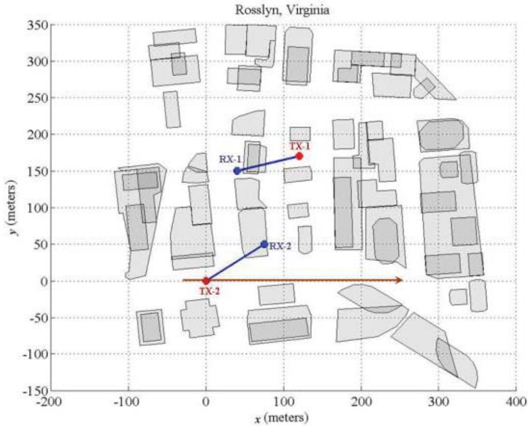


Figure 2: Top view of the links; RX-2 is mounted on the rooftop (80m); the others are street level (2m).

Figure 3 is a top view of the city showing the channel gain generated by TX-1 to any street-level receiver at 300 MHz. All nodes use a single ½ wavelength vertical whip antenna. The gain is computed from the input to the TX-1 antenna to the output of the receiver antenna—no amplifiers are modeling in the RF link. The 3-D propagation model is a collection of ray paths that approximate the 3-D wave equation [5]. These paths include not only multiple reflections off of the building plates and the ground plane but also encompass multiple edge and vertex diffractions. These diffraction terms allow the RF energy to propagate around corners and over building tops. The 3-D model produces magnitude and phase to coherently sum the complex-valued rays. The magnitude of the coherently summed rays is the channel gain between TX-1 and a receiver.

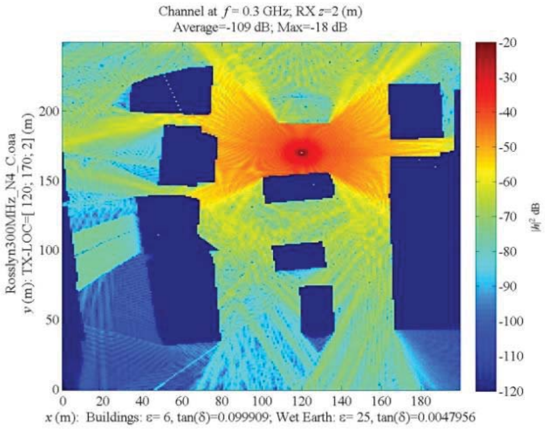


Figure 3: Channel from TX-1 to street-level locations (2m); 3-D propagation.

The channel gain matrix for this narrowband wireless network when TX-2 is located at the origin is

$$G = \begin{array}{c c} & \begin{array}{cc} TX_1 & TX_2 \end{array} \\ \begin{array}{c} RX_1 \\ RX_2 \end{array} & \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \end{array} = \begin{bmatrix} -72 & -75 \\ -71 & -80 \end{bmatrix} \text{[dB]} \,. \qquad (1)$$

More generally, when TX-2 drives along its track, the gain matrix has the form

$$G = \begin{bmatrix} -72 & g_{12}(x) \\ -71 & g_{22}(x) \end{bmatrix}. \qquad (2)$$

Any wireless power control algorithm must adapt to delayed, noisy, distributed and indirect measurements of this narrowband and varying gain matrix. The next section uses this gain matrix to compute bounds on the power and rate of this network.

3. Power-Rate Bounds

Figure 4 is a schematic of a two-link network. In this two-link system, the 2×2 narrowband channel gain matrix $G(t)$ maps the transmitted power to the received power

$$\mathbf{p}_R(t) = G(t)\, \mathbf{p}_T(\, t - \Delta t_D(t)\,) \qquad (3)$$

subject to the downlink delay Δt_D. This delay is on the order of a few micro-seconds for the small unit tactical scenarios confined to a few urban blocks. Each receiver is equipped with power controller. The controllers control power command to the transmitter on its link. The transmitter actually broadcasts a delayed and noisy replica of the control power vector

$$\mathbf{p}_T(t) = \mathbf{p}_C(\, t - \Delta t_U(t)\,) + \mathbf{p}_E(t) \qquad (4)$$

subject to the uplink delay Δt_U. This delay accounts for all the RF measurements and signal processing to estimate the SINR. Depending on the measurement scheme (e.g., each packet may have blank slots), this delay can range up to 10's of micro-seconds. Figure 4 shows the controllers setting the power control command by comparing the external SINR $\gamma_X(t)$ to the measured SINR $\gamma_M(t)$. The SINR on the l-th link is

$$\gamma_{M,l}(t) = \frac{g_{ll}(t)\mathbf{p}_{T,l}(t-\Delta t_D(t))}{\mathbf{p}_{N,l}(t)+\sum_{l\neq k} g_{lk}\,(t)\,\mathbf{p}_{T,k}(t-\Delta t_D(t))}, \qquad (5)$$

where $\mathbf{p}_N$ denotes the additive noise vector. Not shown is the messaging and schemes to measure the SINR.

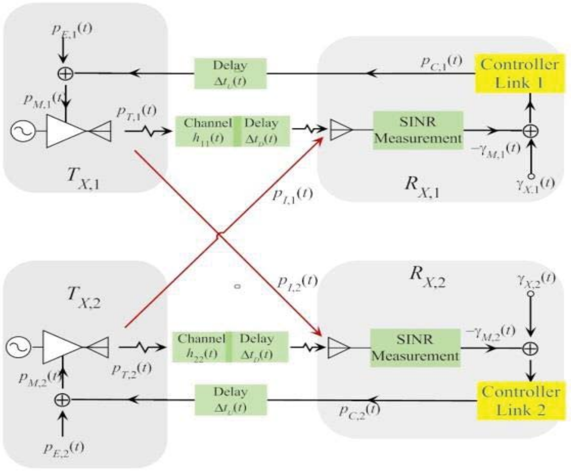


Figure 4: Schematic of a two-link wireless network.

The throughput is bounded by the capacity equation

$$R_b(t) \;=\; f_B \times \log_2(1 + \gamma_M(t)\,), \qquad (6)$$

where f_B denotes the bandwidth. This throughput does not account for any messaging overhead or specific coding. Consequently, $R_b(t)$ is the upper bound on the throughput. The total power of this model network is the sum of the transmitted power:

$$\mathbf{p}_\Sigma(t) \;=\; \textstyle\sum_{l=1}^{L} \mathbf{p}_{T,l}(t). \qquad (7)$$

This power does not account for any overhead power. Consequently, $\mathbf{p}_\Sigma(t)$ is a lower bound on the total power consumed by the network.

Maximizing throughput $R_b(t)$ and minimizing network power $\mathbf{p}_\Sigma(t)$ under a quality of service (QoS) constraint is a canonical WPC problem. The next section bounds the best possible power and information rates that wireless network could deliver. These bounds assume that the network has the channel matrix, that both messaging and optimization are instantaneous and without error. The instantaneous messaging removes the delays and therefore allows for optimization using a constant gain matrix. Equivalently, an all-knowing network "genie" inhabits the network to deliver instantaneous optimization. The subsequent feasible power vectors determine the power-rate image.

3.1. The Power-Rate Image

[6, Section 16.3.2] Let $G \in \mathbb{R}^{L\times L}$ be a given channel gain matrix for L wireless links. Define the signal and interference matrices as

$$G_S = \begin{bmatrix} g_{11} & 0 & \cdots & 0 \\ 0 & g_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{LL} \end{bmatrix};\; G_I = G - G_S. \quad (8)$$

Let $\mathbf{p}_N \in \mathbb{R}^L$ denote the constant noise vector. Let $\boldsymbol{\gamma}_M\colon \mathbb{R}^L \to \mathbb{R}^L$ denote the measured SINR vector

$$\boldsymbol{\gamma}_M(\mathbf{p}_C) \,=\, \frac{G_S\mathbf{p}_C}{\mathbf{p}_N + G_I\mathbf{p}_C}. \qquad (9)$$

Let $\boldsymbol{\gamma}_X$ denote a constant, user-specified SINR constraint vector. Let P_C denote the set of *feasible control vectors*:

$$P_C \,=\{\mathbf{p}_C \in\, \mathbb{R}^L\!: \boldsymbol{\gamma}_X \le \boldsymbol{\gamma}_M\,(\mathbf{p}_C)\}. \qquad (10)$$

Let Γ_X:=diag($\boldsymbol{\gamma}_X$) and denote the *scaled interference matrix* as

$$F \,=\, G_S^{-1}\Gamma_X G_I. \qquad (11)$$

If the spectral radius of F is strictly less than 1, the vector

$$\mathbf{p}_{\min} = (I - F)^{-1} G_S^{-1}\Gamma_X \mathbf{p}_N \qquad (12)$$

is well-defined, non-negative, and is the vertex of the convex cone of all feasible control vectors:

$$P_C \,=\, \mathbf{p}_{\min} +\, (I - F)^{-1}\mathbb{R}_+^L. \qquad (13)$$

The network performance function is the mapping from on the set P_C of feasible control vectors into the Power-Rate plane:

$$\mathfrak{N}(\mathbf{p}_C) = \begin{bmatrix} \mathbf{p}_\Sigma(\mathbf{p}_C) \\ R_b(\mathbf{p}_C) \end{bmatrix} \begin{matrix}\text{Watts} \\ \text{Mbps}\end{matrix}. \qquad (14)$$

All possible network performances are determined by the Power-Rate image:

$$\mathfrak{N}(P_C) = \{\, \mathfrak{N}(\mathbf{p}_C) : \mathbf{p}_C \in P_C\} \qquad (15)$$

3.2. Example: Urban Links at 300 MHz

The matrix equation $\mathbf{p}_R$=$G\mathbf{p}_T$ models the RF channel as narrowband system. For TX-2 parked at x=0, the gain matrix is

$$G = \begin{array}{ccc} & TX_1 & TX_2 \\ RX_1 & g_{11} & g_{12} \\ RX_2 & g_{21} & g_{22} \end{array} = \begin{bmatrix} -72 & -75 \\ -71 & -80 \end{bmatrix} [\text{dB}]. \qquad (16)$$

A bandwidth of f_B=1 MHz bandwidth sets the standard noise power at $\mathbf{p}_N$=−131 dBW for suburban noise. Figure 5 shows the power-rate (PR) image for this gain matrix and noise level. The (blue) dots are the mapping of random feasible power vectors $\mathfrak{N}(\mathbf{p}_C)$. The dense sampling of the feasible set PC near $\mathbf{p}_{\min}$ "fills" in the PR image. The (green) line is the Pareto front computed by maximizing the single objective function constrained to each feasible power level:

$$\mathbf{p}_\Sigma \rightarrow \max\{R_b(\mathbf{p}) : \mathbf{p} \in P_c ; \mathbf{p}_\Sigma = \mathbf{1}^T\mathbf{p}\}. \qquad (17)$$

By construction, the Pareto front is an upper bound for network performance [7].

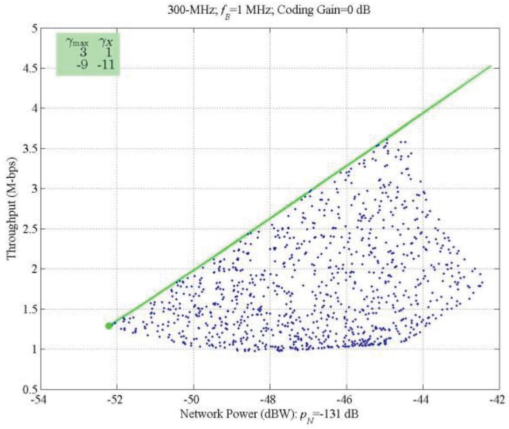


Figure 5: Power-Rate Image for 300-MHz urban links (x=0) in standard urban noise.

The plot also reports the external SINR vector $\boldsymbol{\gamma}_X$ and a maximum SINR vector $\boldsymbol{\gamma}_{\max}$. The maximum SINR vector is an upper bound on the external SINR vector. Selecting $\boldsymbol{\gamma}_X \leq \boldsymbol{\gamma}_{\max}$ guarantees the scaled interference matrix F has spectral radius less than one so that the set of feasible power vectors P_C is not empty. The external SINR vector $\boldsymbol{\gamma}_X$ is also called the QoS vector because requested throughput $R_{b,X}$ on each link is set as

$$R_{b,X} = f_B \times \log_2(1 + \gamma_X). \qquad (18)$$

Figure 5 reports that the external SINR vector was set by reducing the maximum SINR vector by 2 dB: Equivalently, the QoS in terms of throughput is 1.2 Mbps on Link 1 and 0.1 Mbps on Link 2.

The small unit tactical scenarios typically use low transmit power and uniform throughput on each link. For ease of discussion, the RF noise will be increased to −70 dBW. The external SINR vectors are set to −10 dB or 0.1375 Mbps per link. Figure 6 shows the effect of this high noise: for total network power of $\mathbf{p}_\Sigma$=2 Watts, network throughput of 0.5 Mbps may be realized—for the gain matrix $G(x)$ at x=0. There is the substantial question of how this network performs when TX-2 moves along its track.

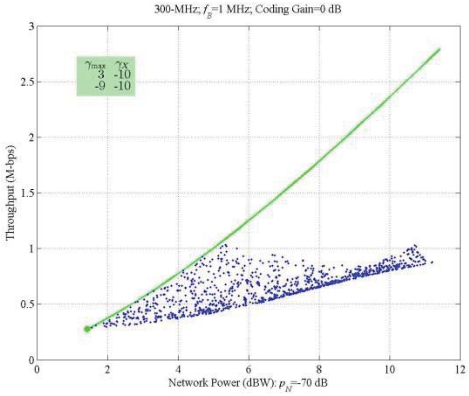


Figure 6: Power-Rate Image for 300-MHz urban links (x=0) in high noise.

4. Tactical High-Noise MANET

Referring to Figure 2, TX-1, TX-2, and RX-1 are all at street level; RX-2 is on the roof of an 80-meter building and slightly back from the rooftop edge. TX-2 travels along the x-axis track. This example shows the variation of the gain matrix along the track and the subsequent variations in the SINR and throughput while operating in high RF noise.

4.1. The Gain Matrix

The gain matrix is determined by the position of TX-2 along the x-axis track and is plotted in Figure 7:

$$G(x) = \begin{bmatrix} -72 & g_{12}(x) \\ -71 & g_{22}(x) \end{bmatrix} \quad [\text{dB}]. \qquad (19)$$

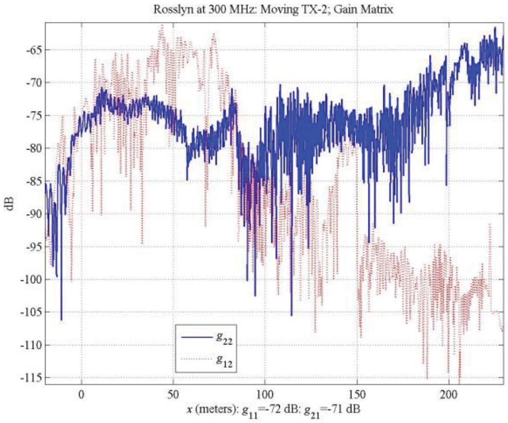


Figure 7: Gain matrix along the x-axis track.

The figure shows TX-2's interference on RX-1 maximizes around x=50 meters because the transmitter and receiver are within line of sight of and then sharply decreases when TX-2 is blocked by the buildings. For communication between TX-2 and RX-2, the TX-2 signal is weak at the start of the track because TX-2 is behind a building. The signal gets stronger as TX-2 comes into view of RX-2, but then decreases around 50 meters due RX-2 sitting on top of the building that TX-2 is passing. The sharp jumps in this graph are due to the blocking and unblocking with TX-2.

4.2. SINR

A tactical scenario typically uses low-power matched to the local RF noise. For convenience, this baseline case uses one Watt on each transmitter power but sets the RF power equal to the noise ($\mathbf{p}_N$=–70 dBW). Figure 8 shows the SINR's for both links as TX-2 travels along the x-axis track.

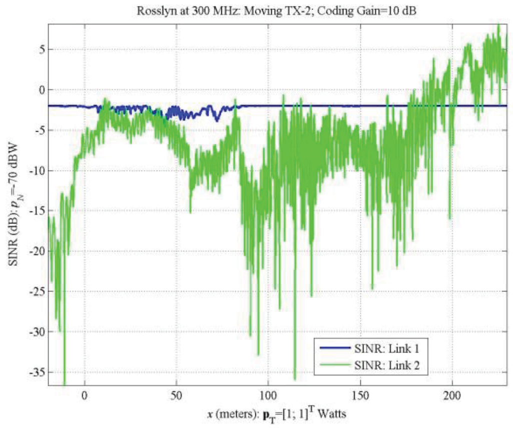


Figure 8: SINR along the x-axis track; unit power on each TX; high noise.

For Link 1, the power from TX-1 to RX-1 is not changing whereas the interference from TX-2 is rolling off as TX-2 travels along the x-axis ($g_{12}(x)\downarrow$). In fact, the interference power rolls off well below the noise power. In these noise-limited regions, the SINR on Link 1 will be constant:

$$\gamma_{M,1}(x) = \frac{g_{11}\,\mathbf{p}_{T,1}}{\mathbf{p}_N + g_{12}(x)\,\mathbf{p}_{T,2}} \approx \frac{g_{11}}{\mathbf{p}_N}. \qquad (20)$$

For Link 2, the power from TX-1 to RX-2 is not changing whereas the power from TX-2 is increasing as TX-2 travels along the x-axis ($g_{22}(x)\uparrow$). Consequently, RX-2 receives more signal power from TX-2 so its SINR will increase.

4.3. Throughput

The SINR's on each link determine the throughput on each link. Figure 9 plots these link throughputs (Link 1: blue; Link 2: green) and their sum (black). The plot shows that Link 1 carries almost all the information whereas Link 2 is shut down in more than one location. Indeed, when one link is the strongest, maximizing throughput is accomplished by loading all the power on that link. Some fairness is achieved by forcing a quality-of-service constraint.

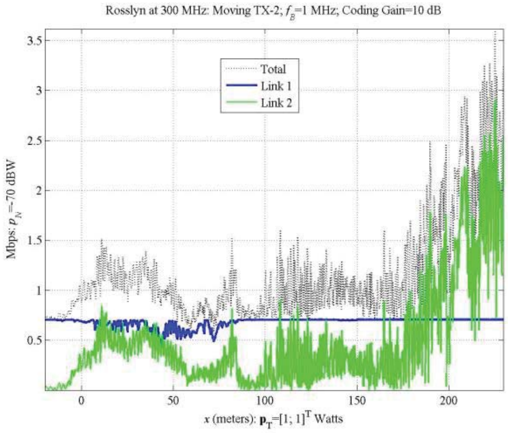


Figure 9: Throughput along the x-axis track; unit power on each TX; high noise; No QoS.

4.4. Quality of Service

Figure 10 reports the throughput that the Pareto Font computation (Section 3) can deliver when applied to every gain matrix along the track under the constraints:

- The total network power is fixed to 2 Watts for comparison with the fixed unit power of Figure 9.
- There is QoS constraint that each link carry a minimum of 0.25 Mbps.

The figure plots realized throughputs on each link (Link 1: blue; Link 2: green) and their sum (black). The plot shows that Link 1 typically reduces power and Link 2 increases power in comparison to the fixed unit power simulation. However, not every position on the track can support this QoS: only 61% of the track can deliver this QoS. In comparison to the fixed power of Figure 9, this QoS is realized "by accident" over 31% of the track.

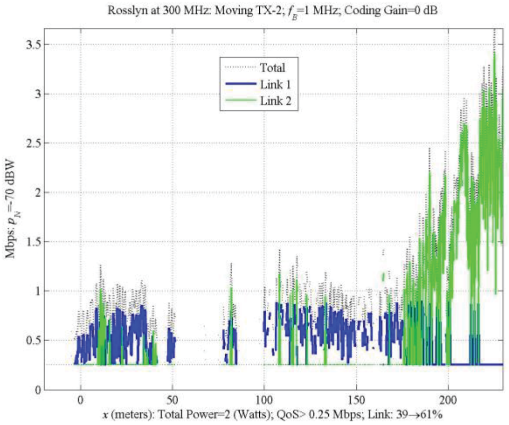


Figure 10: Throughput along the x-axis track; high RF noise; QoS constraint setting the TX power.

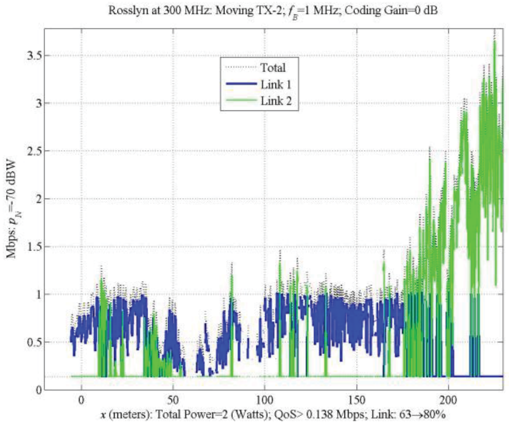


Figure 11: Throughput along the x-axis track; high RF noise; QoS constraint reduced to 138 kpbs.

5. Applications

Figure 10 is the payoff for this paper. Best possible Wireless Power Control is computed with a QoS constraint assuming instantaneous messaging, perfect SINR measurements, and instantaneous optimization. Figure 10 sets an upper bound on network throughput and the subsequent benchmarking for DPC algorithms that are driven along the track. A lower bound is supplied by the fixed unit power of Figure 9. A credible DPC algorithm must deliver performance between these two extremes.

This quasi-realistic simulation also shows that DPC algorithms that rely on forward predictions or are sensitive to stale measurements may be confounded by rapid and wide dynamic range of the gain matrix along the track. The rapid variations also highlight that the DPC algorithms must also have channel-adaptive update schedules.

Figure 10 also highlights the non-trivial problem of setting the QoS constraint. The QoS must be dynamic [8], [9]. Indeed, Figure 11 shows that relaxing the QoS constraint to 0.1375 Mbps (See Figure 6) allows the network to deliver QoS performance over 80% of the track while the fixed power increases to only 63% of the track.

Figure 6 also offers another feature that is related to Pareto optimization. A close inspection reveals that the PR mapping is folding the 2-D feasible power vectors. This folding will be problematic for Pareto-front algorithms: termination may happen at a local front rather than the global front. This folding may also explain problematic convergence for DPC algorithms.

Figure 12 challenges the narrowband modeling by plotting the gain matrix over the 10-MHz band centered at 300 MHz. The plot shows that the 1-MHz band at 300 MHz has 5 dB variations across the band. Opening the bandwidth to 2 MHz incurs 20-dB variations across the band. Therefore, wideband wireless power control modeling must account for these tactical RF channel variations.

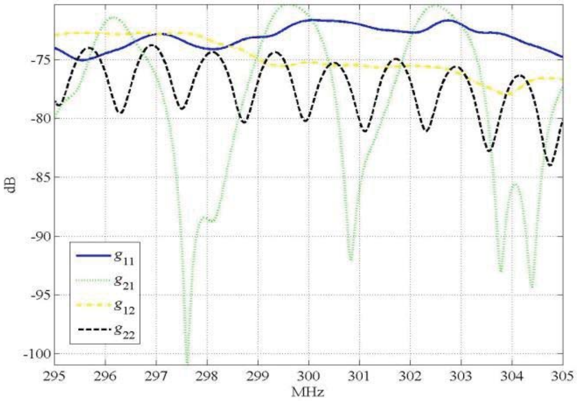


Figure 12: Wideband Gain matrix.

6. Conclusions

The bounds developed in this paper show best possible power-rate performance any WPC algorithm could deliver in this tactical scenario. The power-rate image provides the wireless engineer with a practical graphic to quantify best possible network performance. The engineer must then decide if these best bounds are sufficiently strong to justify proceeding with algorithm development and the expectation that the realized network performance will still be "good enough" or that another approach is needed. For example, Figure 10 shows that QoS can be improved to cover 20% more of the track—at best. If this payoff is too small, no algorithm or associated tweaks will boost this payoff. Instead, the wireless network must employ a new gain matrix—obtained by CDMA coding, or DSA migration to another frequency band (See Figure 12)—or equip the radios with new antennas to deliver adaptive nulling, beamforming, MIMO and interference alignment, or polarization diversity—or the network must employ MAC-layer schemes to mitigate local interference.

[1] Single-Transceiver Dynamic Spectrum Access (ST-DSA) Navy SBIR 2014.2 - Topic N142-112 http://www.navysbir.com/n14_2/N142-112.htm

[2] Jiang Canming; Yi Shi; Sastry Kompella; Y. Thomas Hou; Scott F. Midkiff [2013] Criteria Optimization in Multihop Wireless Networks: Characterizing Throughput-Energy Envelope, *IEEE Transactions on Mobile Computing*, 12(9).

[3] Naveed Ul Hassan, Mohamad Assad, Hamidou Tembine [2013] Distributed H-Infinity-Based Power Control in a Dynamic Wireless Environment, *IEEE Communications Letters*, 17(6).

[4] Ontiveros, M.; Arceo, D.; Allen, J.; James, J.; Daly, M., [2014] Beam-Space MIMO simulations in a 3-D suburban environment, *Antennas and Propagation Society International Symposium (APSURSI), 2014 IEEE*, pp.472–473.

[5] Marhefka, Ronald J. [2002] *Numerical Electromagnetics Code-Basic Scattering Code, (NEC-BSC 4.2) User's Manuel*, Department of Electrical Engineering, Ohio State University, Columbus Ohio.

[6] Goldsmith, Andera [2005] *Wireless Communication*, Cambridge University Press, Cambridge, UK.

[7] Boyd, Stephen; Lieven Vandenberghe [2005] *Convex Optimzation*, Cambridge University Press, Cambridge, UK.

[8] Fabiano de Sousa Chaves; Mohamed Abbas-Turki; Hisham Abou-Kandil; João Marcos Travassos Romano [2013] Transmission Power Control for Opportunistic QoS Provision in Wireless Networks, *IEEE Transactions on Control Systems Technology,* 21(2).

[9] Viswanath, Pramod; David N. C. Tse; Rajiv Laroia [2002] Opportunistic Beamforming Using Dumb Antennas, *IEEE Transactions on Information Theory*, 48(6).

Fault-Tolerant Wireless Multihop Transmissions with Byzantine Failure Detection

Norihiro Sota[1] **and Hiroaki Higaki**[1]
[1] Tokyo Denki University, Japan

Abstract— *Wireless multihop networks consist of numbers of wireless nodes. Hence, introduction of failure detection and recovery is mandatory. Until now, various failure detection and recovery methods such as route switch and multiple routes detection have been proposed based on an assumption with stop failure model. However, the assumption that failed wireless nodes never transmit any messages is too restrict the area where the proposed methods can be applied. In order to solve this problem, we propose a novel failure detection and notification method that supports not only stop failure but also Byzantine failure. That is, it is possible for failed wireless nodes to transmit malicious messages not according to the data message transmission and the failure detection and notification protocols unconsciously due to failure or even intentionally. Here, the design of failure detection and notification protocols is critical. In this paper, Byzantine failures in an intermediate node are detected by its multiple neighbor wireless nodes cooperatively since the neighbor wireless nodes are also vulnerable and might transmit erroneous failure notifications. From the performance viewpoint, no additional control messages are required to be transmitted while no failure wireless node is detected, i.e., in usual data message transmissions.*

Keywords: Ad-Hoc Networks, Fault-Tolerant Wireless Networks, Byzantine Failure, Cooperative Watchdog, Protocol, Ad-Hoc Routing.

1. Introduction

In mobile wireless ad-hoc networks (MANETs) and wireless sensor networks, data messages are transmitted according to wireless multihop transmissions where each intermediate wireless nodes along the wireless multihop transmission route forwards them from the source wireless node to the destination one. Usually, the wireless transmission range of each wireless node is limited and the wireless nodes are assumed to be distributed densely enough for all the wireless nodes to be possible to communicate with some neighbor wireless nodes directly and to communicate with almost all the other wireless nodes by the wireless multihop communication. This is because, all the observation area is required to be covered by at least one sensor node and the sensor data messages are required to be transmitted to one of the sink nodes in sensor networks and enough high connectivity by wireless multihop transmissions is required in usual mobile wireless ad-hoc networks.

Such wireless multihop networks consist of numbers of wireless nodes. Hence, it is impossible to operate such wireless multihop networks continuously without failure detection, notification and recovery mechanisms. That is, higher resilient wireless multihop networks are required. Until now, various techniques for fault-tolerant distributed systems such as distributed failure detection, notification and recovery algorithms and systems have been proposed [3], [10]. For wireless multihop networks, only a naive watchdog method and its slight extensions have been proposed. Here, almost only the stop failure model in which failed wireless nodes become silent and never transmit any data and control messages is supported. Even though some methods support the Byzantine failure model, desirable behavior such as only erroneous data messages are transmitted is assumed. As discussed in this paper, erroneous and/or malicious data message transmissions deviated from the application protocols and erroneous and/or malicious failure detection and notification transmissions are required to be supported. This paper proposes a novel cooperative watchdog method and designs a data message transmission protocol with an extension of the Byzantine failure detection and notification and a routing protocol for detection of watchdoggable wireless multihop transmission routes based on flooding based ad-hoc routing protocols such as AODV [7].

2. Related Works

Suppose a wireless multihop transmission route $\mathcal{R} := ||N_0(= N^s) \ldots N_n(= N^d)\rangle\rangle$ from a source wireless node N^s to a destination one N^d in a wireless multihop network such as a mobile wireless ad-hoc network and a wireless sensor network. If one of the intermediate wireless nodes N_f $(0 < f < n)$ is detected to be failed by one of its neighbor wireless nodes, a failure notification message is transmitted to the source node N^s and another wireless transmission route $\mathcal{R}'$ without N_f is searched and detected. Then, data messages are transmitted through not $\mathcal{R}$ but $\mathcal{R}'$. Until now, some failure detection, notification and recovery by re-routing have been proposed [4], [9]. In addition, for avoidance of high communication and time overhead for search of a detour wireless multihop transmission route, various multiple route detection protocol have also been proposed where multiple wireless multihop transmission routes are detected in a routing protocol and the routes are switched each time a failed intermediate wireless node is detected along an available wireless multihop transmission route [1], [5], [8]. These papers only discuss the methods to switch wireless multihop transmission routes after detection of failure of one of the intermediate wireless nodes. The discussion of failure detection and notification is almost out of range.

There are some various failure model for wireless nodes [10]. Almost all of them assume that wireless nodes fail according to the following stop failure model where the failed wireless nodes become silent and the stop failure is detected by at least one of the other wireless nodes by using periodically transmitted "Hello" or "I'm alive" messages.

[Stop Failure Model]

A failed wireless node stops. It becomes silent, i.e., it never transmits and receives any data and control messages. □

A stop failure usually detected by using a timer [2]. A neighbor wireless node Q of another wireless node P sets its timer. If Q does not receive a message from P before the expiration of the timer, Q detects failure of P. In wireless multihop data message transmissions along $\mathcal{R}$, in cases that there are no failed wireless nodes in $\mathcal{R}$, within a certain interval after the time when an intermediate wireless node N_{i-1} forwards a data message m to its next-hop wireless node N_i, N_i forwards m to its next-hop wireless node N_{i+1}. As shown in Figure 1, under an assumption of the disk model wireless signal transmissions, N_{i-1} is surely within the wireless transmission range of N_i and m transmitted from N_i to N_{i+1} is surely overheard by N_{i-1}. Hence, if N_{i-1} does not overhear m forwarded by N_i to N_{i+1} during a certain interval after N_{i-1} forwards m to N_i, N_{i-1} detects that N_i is failed.

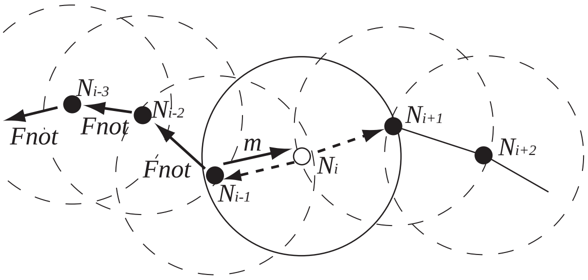


Fig. 1: Stop Failure Detection in Wireless Multihop Networks.

3. Proposal

3.1 Problems

As discussed in the previous section, the stop failure model is supported in various fault-tolerant methods for wireless multihop networks. The Byzantine failure model is more general than the stop failure model and it is much difficult to support [10].

[Byzantine Failure Model]

Failed wireless nodes do not always become silent. They might transmit and receive data and control messages. In addition, the transmission of the messages are not always according to application protocols. The failed wireless nodes might transmit erroneous and/or malicious data and control messages. □

Different from the stop failed wireless nodes, the Byzantine failed intermediate wireless nodes in a wireless multihop transmission route might transmits different data messages from those they have received to their next-hop wireless nodes and might transmits data messages to their next-hop wireless nodes even though they have not yet receive any messages from their previous-hop wireless nodes. For such problems, some watchdog methods by the previous-hop nodes have been proposed [6]. If the wireless transmissions are based on the disk model, the transmitted data message from an intermediate node N_i to its next-hop wireless node N_{i+1} is overheard by its previous-hop node N_{i-1}. As shown in Figure 2, if N_i transmits a different data message m' to N_{i+1} from m that N_i has received from N_{i-1}, N_{i-1} detects the failure of N_i by receipt of m' different from m. That is, the Byzantine failure in N_i is detected by N_{i-1} by the comparison of data messages received and transmitted by N_i.

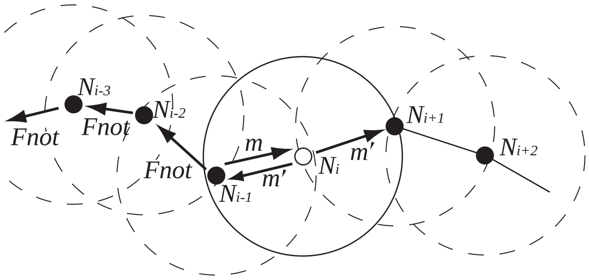


Fig. 2: Byzantine Failure Detection in Wireless Multihop Networks.

However, if two successive intermediate wireless nodes N_i and N_{i+1} simultaneously fail, it is impossible for N_{i-1} to detect the failure especially in N_{i+1} in Figure 3. Though N_i correctly forwards a data message m received from N_{i-1} to N_{i+1}, a failed intermediate wireless node N_{i+1} transmits a different data message m' from m to its next-hop wireless node N_{i+2}. Since N_i overhears m' from N_{i+1}, it can detect the failure in N_{i+1} due to the comparison of m and m'. However, if N_i also fails, N_i does not transmits any failure notification control messages to its neighbor wireless nodes and no failure recovery such as rerouting without failed wireless nodes is initiated. Generally, $n-$simultaneous failure is defined as follows [3]:

[$n-$simultaneous Failure]

The number of failed wireless nodes are at most n at any instance. Failed wireless nodes are never recovered by themselves and removed from the wireless network system by a certain maintenance procedure. □

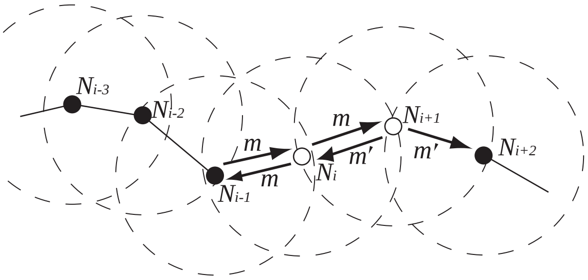


Fig. 3: Simultaneous Byzantine Failures in Wireless Multihop Networks.

On detect the failure of an intermediate wireless node N_i, a wireless multihop transmission of a failure notification message $Fnot$ to the source node N^s is initiated by N_{i-1}. On receipt of the $Fnot$, N^s searches a wireless multihop transmission route $\mathcal{R}'$ to the destination wireless node N^d without the failed intermediate wireless node N_i. Until now, the failure detection is assumed to be correctly done in any intermediate wireless node. However, the failed intermediate wireless node N_{i-1} might erroneously detect a failure of its neighbor wireless node especially its next-hop intermediate wireless node N_i and initiate the transmission of the failure notification control message by transmission of a failure notification message $Fnot$ of N_{i+1} to its previous-hop wireless node N_{i-2} even though N_i does not fail as shown in Figure 4. Since it is impossible for N_{i-2} to find the $Fnot$ is transmitted by N_{i-1} erroneously, N_{i-2} and the other intermediate wireless nodes forwards the message to their previous-hop wireless nodes along $\mathcal{R}$. Here, the source node is notified for requirement of re-routing due to failure not in N_{i-1} but in N_i. Hence, newly detected wireless multihop transmission route surely excludes not N_{i-1} but N_i, which is a serious problem to be solved.

The failure notification control message $Fnot$ of N_i transmitted by N_{i-1} is also received by N_i. Hence, it can detect the erroneous or malicious transmission of $Fnot$. In order to notify the failure of N_{i-1} to N^s, an additional wireless transmission route from N_i to N^s without N_{i-1} is required. In addition, since N^s receives two different failure notification messages from N_i and N_{i-1}, N^s is required to select one of them for recovery.

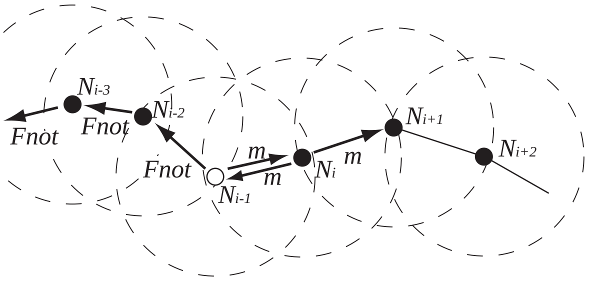


Fig. 4: Erroneous Failure Detection of Byzantine Failure.

3.2 Neighbor Watchdog Wireless Nodes

In order to solve the problem discussed in the previous subsection, that is, under the 1-simultaneous Byzantine failure assumption, one of the intermediate wireless nodes along a wireless multihop transmission route might erroneously or maliciously transmit a failure notification control message, this paper proposes a cooperative watchdog method with the help of a neighbor wireless node O_i of N_{i-1} and N_i as shown in Figure 5. Here, a neighbor watchdog wireless node O_i is within the wireless transmission ranges of both N_{i-1} and N_i. Hence, O_i overhears the data messages transmitted both from N_{i-1} to N_i and from N_i to N_{i+1}. Hence, same as N_{i-1}, O_i also detects the failure of N_i by comparison of data messages transmitted from N_{i-1} to N_i and from N_i to N_{i+1}. Therefore, even if N_{i-1} erroneously or maliciously transmits a failure notification message $Fnot$ of N_i to N_{i-2}, O_i detects that the $Fnot$ message while N_i correctly works.

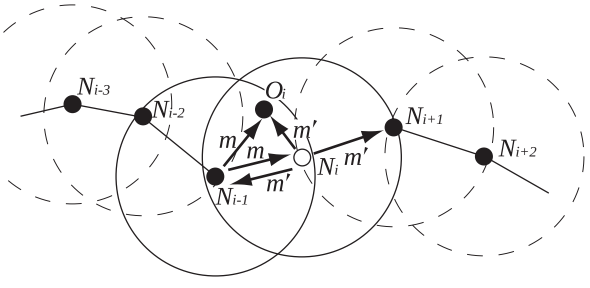


Fig. 5: Cooperative Watchdog Neighbor Wireless Nodes.

In cases that O_i detects the erroneous transmission of the $\mathcal{F}not$ message, O_i should prevent the wireless multihop transmission of $Fnot$ of N_i to N^s and initiate the wireless multihop transmission of $Fnot$ of N_{i-1} since O_i has detected the failure of N_{i-1}. Hence, a control message $Fnot$ for notification of failure of N_{i-1} is transmitted from O_i to N^s through N_{i-2}. However, N_{i-2} is not always a neighbor wireless node of O_i and the $Fnot$ message is required to be transmitted not through the failed intermediate wireless node N_{i-1}. In order to realize the later discussed lower overhead route detection based only on the neighbor node information in each wireless node, O_i and N_{i-2} are required to be 1-hop neighbor or 2-hop neighbor through an intermediator wireless node I_i as shown in Figure 6. The role of I_i is only forwarding the $Fnot$ message from O_i to N_{i-2}.

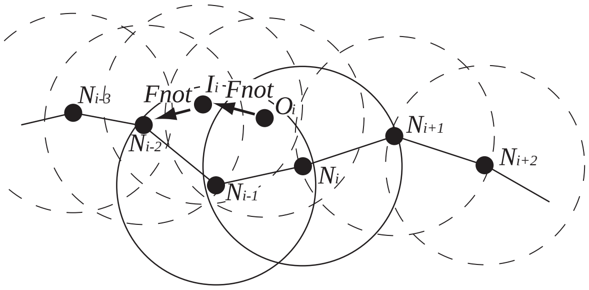


Fig. 6: Intermediator Wireless Nodes for Notification.

Now, we discuss the procedure in wireless nodes N_{i-1}, N_i, O_i and I_i for detection and notification of the 1-simultaneous Byzantine failure of one of these nodes to N_{i-2}. In the following discussion, the $Fnot$ message from O_i is transmitted to N_{i-2} through I_i; however, almost the same procedure is possible to be applied without the intermediator node I_i.

First, in the cases free from the Byzantine failures of all the intermediate, the neighbor watchdog and the intermediator wireless nodes, a data message m is transmitted through the wireless transmission route $\mathcal{R}$ according to the forward of m by the intermediate wireless nodes N_i as shown in Figure 7. There are no additional control message is required to be transmitted.

In cases that the intermediate wireless node N_i fails according to the Byzantine failure model, the data message m forwarded from N_{i-1} to N_i is not transmitted from N_i to N_{i+1}, a different data message m' from m is transmitted from N_i to N_{i+1} or a data message m'' is transmitted from N_i to N_{i+1} even though no data message is transmitted from N_{i-1} to N_i. Anyway, as shown in Figure 8, both N_{i-1} and the neighbor watchdog wireless node O_i detect the difference of data messages transmitted through the wireless links from N_{i-1} to N_i and from N_i to N_{i+1}. At this time, the same failure notification control messages $Fnot$ for the failure of N_i are transmitted from N_{i-1} to N_{i-2} and from O_i to N_{i-2} through I_i. Thus, N_{i-2} receives these two $Fnot$ messages.

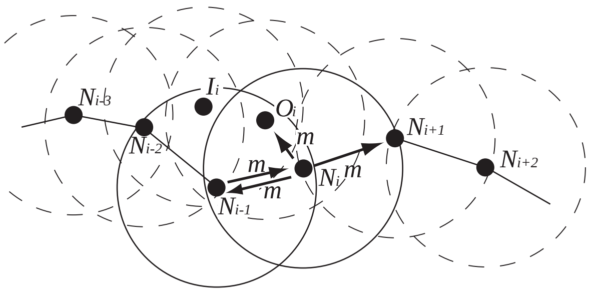


Fig. 7: Data Message Transmissions with No Node Failure.

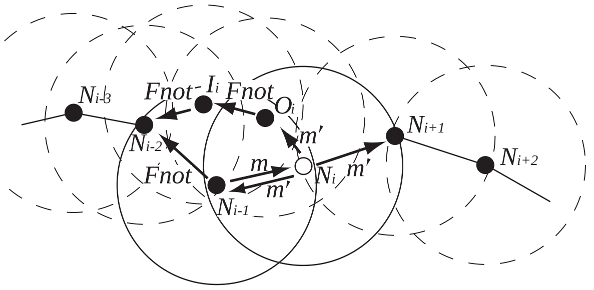


Fig. 8: Detection of Failure in N_i.

In cases that N_{i-1} transmits a failure notification message $Fnot$ for N_i to N_{i-2} though N_i works correctly, N_{i-1} fails according to the Byzantine failure model as shown in Figure 9. Due to the 1-simultaneous Byzantine failure assumption, N_i does not fail. O_i detects that N_{i-1} transmits the $Fnot$ message for N_i to N_{i-2} though N_i does not fail by overhearing the transmitted data and control messages. Thus, O_i transmits a failure notification message $Fnot$ for N_{i-1} to N_{i-2} through I_i.

Same as the previous cases, even though N_i does not fail and works correctly, O_i erroneously detects the failure of N_i and notifies it to N_{i-2} through I_i as shown in Figure 10. Due to the 1-simultaneous Byzantine failure assumption, N_{i-1} does not fail. N_{i-1} detects that O_i transmits a failure notification control message $Fnot$ for N_i though N_i does not fail by overhearing the transmitted data and control messages. Then, N_{i-1} transmits a failure notification message $Fnot$ for O_i to N_{i-2}. Thus, N_{i-2} receives two different failure notification messages $Fnot$ for N_i from O_i and for O_i from N_{i-1}.

Finally, in cases that N_i does not fail and one of O_i and N_{i-1} fails according to the Byzantine failure model and

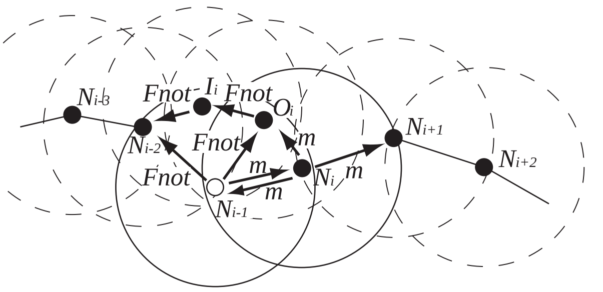


Fig. 9: Detection of Failure in N_{i-1}.

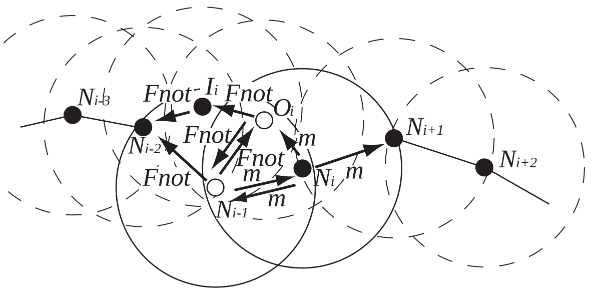


Fig. 11: Failure Detection and Notification in O_i or N_{i-1}.

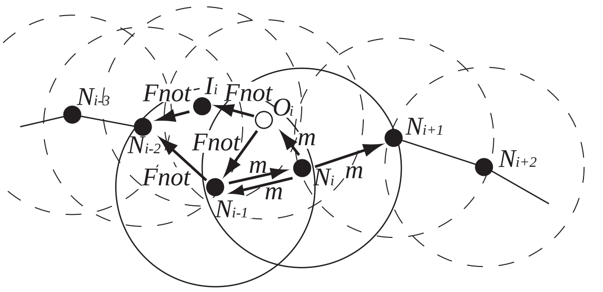


Fig. 10: Detection of Failure in Watchdog Neighbor Wireless Nodes.

Table 1: Failure Node Determination in N_{i-2}.

Failure Node in $Fnot$ from N_{i-1}	Failure Node in $Fnot$ from O_i	Failure Node
N_i	N_i	N_i
N_i	N_{i-1}	N_{i-1}
O_i	N_i	O_i
O_i	N_{i-1}	N_{i-1} or O_i

transmits a failure notification control message $Fnot$ for the other to N_{i-2} as shown in Figure 11. Here, the correct wireless node detects the erroneous or malicious transmission of the failure notification control message $Fnot$ from the failed one. Thus, it transmits another failure notification control message $Fnot$ to N_{i-2}. Hence, N_{i-2} receives two different $Fnot$ messages for N_{i-1} and O_i.

The following Table 1 summarizes the above discussion. If one of the wireless nodes N_{i-1}, N_i and O_i fails, two failure notification control message $Fnot$ from O_i and N_{i-1} are transmitted to N_{i-2}. Thus, when N_{i-2} receives one $Fnot$ message for one of the wireless nodes N_{i-1}, N_i and O_i from I_i or N_{i-1}, it waits for receiving another $Fnot$ message. Then, N_{i-2} determines the really failed wireless node in accordance with Table 1 and transmits a composite failure notification control message to N_{i-3}, which is transmitted to N^s along $\mathcal{R}$ for re-routing for the removal of the failed wireless node.

Usually, a failure of an intermediate wireless node N_j is detected by its neighbor watchdog wireless node O_j and/or its previous-hop wireless node N_{j-1} and a transmission of a failure notification control message $Fnot$ is initiated. Based on the 1-simultaneous Byzantine failure assumption, all the intermediate wireless node between N_{j-2} and N^s are surely correct. So that, these intermediate wireless nodes safely forward the failure notification control message to their previous-hop nodes. However, since the Byzantine failure model is assumed, a transmission of a failure notification message for an intermediate node N_j might be initiated by another intermediate wireless node N_i ($i < j - 1$) erroneously or maliciously. As a result, an intermediate wireless nodes in a wireless multihop transmission route $\mathcal{R}$ might forward an erroneous or malicious failure notification control message which increases the communication overhead in the wireless multihop network.

The unique chance to detect the erroneous or malicious failure notification control message is when the message is initiated. If the $Fnot$ message for N_j is initiated by N_i, N_i transits a $Fnot$ message for N_j though it has not received the message from N_{i+1}, all of which is observed by the neighbor watchdog wireless node O_{i+1}. Hence, it is possible for O_{i+1} to transmit the $Fnot$ message for N_i to N_{i-1} and to induce the confirmation procedure in N_{i-1}. However, if this confirmation procedure is introduced in each intermediate wireless node for transmission of $Fnot$ message hop-by-hop, longer transmission delay is required for $Fnot$ transmission since transmitted $Fnot$ message and the additional $Fnot$ message from O_{i+1} are required to be synchronized at N_{i-1} for confirmation. The transmission delay overhead for the failure notification control message is too high for realization of fault-tolerant wireless multihop networks. Thus, in our protocol, for confirmation of the failure notification message, digital signature of the initial wireless node of the failure notification control message is attached to the $Fnot$ control message.

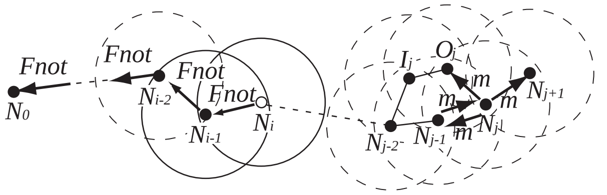


Fig. 12: Erroneous or Malicious Failure Notification to Source Node.

3.3 Watchdoggable Wireless Multihop Transmission Route

As discussed in the previous subsection, for realizing 1-simultaneous Byzantine failure detection in wireless multihop transmissions, all the wireless communication links $|N_iN_{i+1}\rangle$ in a wireless multihop transmission route $\mathcal{R} = ||N_0 \dots N_n\rangle\rangle$ should be $watchdoggable$. The condition for a watchdoggable wireless communication link is as follows:

[Watchdoggable Wireless Communication Links]

A wireless communication link $|N_iN_{i+1}\rangle$ is watchdoggable if and only if there is a neighbor watchdog wireless node O_{i+1} satisfying the following conditions (Figure 13):

- O_{i+1} is a neighbor wireless node of N_{i+1}.
- O_{i+1} is a neighbor wireless node of N_{i-1} or there is a intermediator wireless node I_i neighboring to N_{i-1}, N_i and O_{i+1}. □

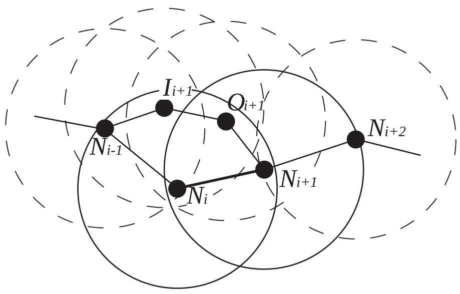


Fig. 13: Watchdoggable Wireless Links.

For determination whether a wireless communication link $|N_iN_{i+1}\rangle$ is a watchdoggable one or not, the neighbor relation with N_{i-1} is required. Hence, in order to determine the possible next-hop wireless nodes satisfying the watchdoggable wireless communication links, each node requires the neighbor relation of two hop neighbor wireless nodes. Thus, each wireless node achieves its location information by using GPS and advertise the location information to its 2-hop neighbor nodes.

The detailed proposed protocol would be discussed in our future research papers.

4. Evaluation

By using the data message transmission protocol with the Byzantine failure detection and notification, fault-tolerant wireless multihop transmissions of data messages are provided. In order to apply the proposed failure detection and notification, the wireless multihop transmission route is required to consist of only watchdoggable wireless communication links. Such a route is able to be detected by a flooding-based routing protocol such as AODV. Here, the protocol has two phases; a flooding phase for a route request control message $Rreq$ transmissions and a unicast phase for a route reply control message $Rrep$ along a detected wireless multihop transmission route $\mathcal{R}$. There are no additional control message transmissions and no additional synchronization overhead for data message transmissions without failure of intermediate wireless nodes.

However, in order to detect the watchdoggable wireless multihop transmission route based on the flooding of an $Rreq$ control message as discussed in the previous section, each candidate of an intermediate node is required to keep the two-hop neighbor relation as discussed in subsection 3.3. That is, each wireless node broadcasts its location information to all its 2-hop neighbor nodes by using TTL centric broadcasts independently of the transmission requests. For data message transmissions, no additional data and control messages are required to be transmitted. Additional control message transmissions are only required to detect and notify the failure of N_{i-1}, N_i and O_i to N^s. These $Fnot$ control messages are transmitted to N_{i-2} and synchronized there which requires communication and synchronization overhead.

In the proposed method, a wireless multihop transmission route is required to consist of only watchdoggable wireless communication links. Hence, a part of wireless communication links are not included in the wireless multihop transmission routes and the available wireless communication links ratio is expected to depend on the density of wireless nodes. Thus, we evaluate the effect on the route detection ratio by the restriction on the wireless communication links in the proposed method in simulation experiments. Figure 14 shows the simulation settings. N^d is a destination wireless node and N^s_is are a source wireless node or intermediate ones. Additionally 1,000–20,000 wireless nodes are randomly distributed in the 600m×600m simulation area whose wireless transmission ranges are 10m.

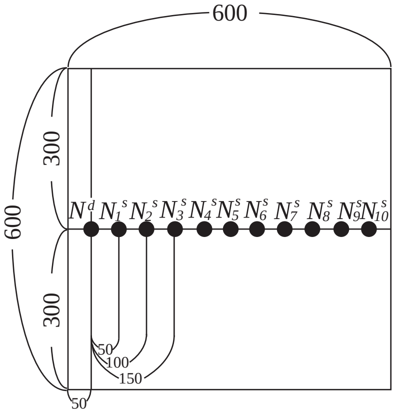

600
600
300
300
N_1^d N_1^s N_2^s N_3^s N_4^s N_5^s N_6^s N_7^s N_8^s N_9^s N_{10}^s
50
100
150
50

Fig. 14: Simulation Setting.

Figure 15 shows the simulation results. The x-axis represents the numbers of wireless nodes, y-axis represents the distance from the source wireless node to the destination one, and z-axis represents the successful route detection ratio. For comparison, the route detection ratio in AODV is also evaluated. In both method, the route detection ratio monotonically increases according to the number of wireless nodes and is almost independent of the distance from the source wireless node to the destination one. In highly dense and sparse distribution of wireless nodes environment, the route detection ratio is almost constant. In the middle range, the route detection ratio steeply changed. In AODV, the threshold of high route detection ratio is 8,000 and the threshold of low route detection ratio is 6,000. On the other hand, in the proposed method, the threshold of high route detection ration is 11,000 and the threshold of low route detection is 6,000. Thus, in the range 8,000-11,000, the proposed method reduces the route detection ratio, which is almost only the disadvantage of the proposed method. The detection, notification and recovery of the Byzantine failed wireless nodes are critical technique for achieving the fault-tolerant wireless multihop networks and the merits of the proposed method surpass the disadvantage for reliable wireless multihop transmission requirements.

5. Concluding Remarks

This paper has proposed a novel communication protocols, i.e., for wireless transmission route detection and for data message transmissions in wireless multihop networks with failure detection, notification and recovery. Though almost all the conventional methods only support the stop failure, the proposed method supports the Byzantine failure where failed wireless nodes does not become silent and continues to communicate with the others out of their application protocols, i.e., erroneous and malicious data messages are transmitted independently of the application protocols. In addition, various erroneous and malicious control messages are also transmitted. This makes difficult to realize the failure detection and notification. The proposed method introduces the cooperative watchdog method where two successive intermediate wireless nodes and an additional neighbor watchdog wireless node cooperate. In the proposed protocol, no additional control message transmissions are needed and the failed wireless node is correctly removed. In addition, the simulation experiments show that the proposed method has a little disadvantage on the successful route detection ratio. However, in the usual density of wireless node to assure the wireless multihop connectivity, almost no reduction in route detection ratio is expected.

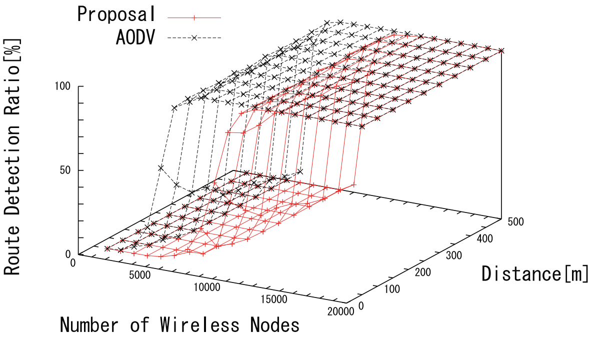


Fig. 15: Route Detection Ratio (Simulation Results).

References

[1] Adibi, S. and Erfani, S., “A Multipath Routing Survey for Mobile Ad-Hoc Networks,” Proceedings of the 2nd Annual IEEE Consumer Communications and Networking Conference, pp. 984–988 (2005).

[2] Cho, Y., Qu, G. and Wu, Y., “Insider Threats against Trust Mechanism with Watchdog and Defending Approaches in Wireless Sensor Networks,” Proceesings of the IEEE Symposium on Security and Privacy Workshop, pp. 134–141 (2012).

[3] Fokkink, W., “Distributed Algorithms: An Intuitive Approach,” *The MIT Press* (2013).

[4] Harada, Y., Wang Hui, Fukushima, Y., Yokohira, T., "A reroute method to recover fast from network failure", Information and Communication Technology Convergence , pp. 903 – 908 (2014).

[5] Kaur, R., Mahajan, R. and Singh, A., “A Survey on Multipath Routing Protocols for MANETs,” International Journal of Emerging Treands and Technology in Computer Science, vol. 2, no. 2, pp. 42–45 (2013).

[6] Pandit, V., Jung, H. and Agrawal,D.P ., “Inherent Security Benefits of Analog Network Coding for the Detection of Byzantine Attacks in Multi-Hop Wireless Networks,” ,IEEE 8th International Conference on Mobile Adhoc and Sensor Systems , pp. 697–702 (2011).

[7] Perkins, C., Belding-Royer, E. and Das, S., “Ad Hoc On-Demand Distance Vector (AODV) Routing,” RFC 3561 (2003).

[8] Perlyasamy, P. and Kathikeyan, E., “Survey of Current Multipath Routing Protocols for Mobile Ad Hoc Networks,” International Journal of Computer Network and Information Security, vol. 5, no. 12, pp. 68–79 (2013).

[9] Po-Kai Tseng and We-Ho Chung ., “Local Rerouting and Channel Recovery for Robust Multi-Hop Congnitive Radio Networks ” IEEE International Conference on Comunications, pp. 2895–2899 (2013).

[10] Raynal, M., “Distributed Algorithms for Message-Passing Systems,” *Springer* (2013).

Self Healing MANET Using Weight Factor

C.Jinshong Hwang[1], Ashwani Kush[2], Rozy Pawar[3]
[1]Dept of Comp Science, Texas State University, San Marcos, Texas, USA, cjhwang@txstate.edu
[2]Dept of Comp Science, UC KUK INDIA 136119, akush20@gmail.com
[3]Dept of Comp Science, UC KUK INDIA 136119, rosy_pawar@yahoo.com

Abstract: *The challenge of wireless communication is that, the environment that wireless communications travel through is unpredictable. Wireless networks that fix their own broken communication links may speed up their widespread acceptance. An effort has been made to propose an on-demand distributed algorithm for self-organizing, multihop, mobile packet radio large network. These nodes are independently controlled and are dynamically reconfigured as nodes may move from one range to another. It is seen that when the network size increases, per node throughput of an ad hoc network rapidly decreases. This is due to the fact that in large scale networks, flat structure of networks results in long hop paths which are prone to breaks. A weight factor has been added that will be broadcast to each node in the network. The proposed algorithm is robust due to the motion, failure, insertion or deletion of nodes. This non periodic algorithm reduces the cost due to computation and communication. Simulation experiments evaluate the performance of the proposed scheme.*

Keywords: Mobile Ad Hoc networks, Self healing, Load balancing, Mobile computing, Routing

1. Introduction

In developing broadband digital networks, a short service-outage such as a link failure or a node failure can cause a serious impairment of network services. It is due to the volume of network traffic carried by a single link or node. Moreover, the outage can stimulate end users to try to re-establish their connections within a short time. The retrials, however, make the problem worse because the connection establishment increases the traffic volume further. Fast restoration from a network failure becomes a critical issue in deploying high-speed networks. Self-healing algorithms have been recognized as a major mechanism for providing the fast restoration. A self-healing system [1,2,8] should recover from the abnormal state and return to the normal state, and should start functioning as it was prior to failure. One of the key issues associated with self-healing networks is to optimize the networks while expecting reasonable network failures [3,4,5,8]. Self-healing network (SHN) [6,8] is designed to support transmission of messages across multiple nodes while also protecting against recursive node and process failures. It will automatically recover itself after a failure occurs. The problem of self-healing is in networks that are reconfigurable in the sense that they can change their topology during an attack. One goal is to maintain connectivity in these networks [9], even in the presence of repeated adversarial node deletion. Modern computer systems are approaching scales of billions of components. Such systems are less akin to a traditional engineering enterprise such as a bridge, and more akin to a living organism in terms of complexity. A railway overbridge must be designed in such a way that, key components never fail, since there is no way for the bridge to automatically recover from system failure. In contrast, a living organism can not be designed so that no component ever fails: there are simply too many components. For example, skin can be cut and still heal. Unfortunately, current algorithms ensure robustness in computer networks through hardening individual components or, at best, adding lots of redundant components [7].

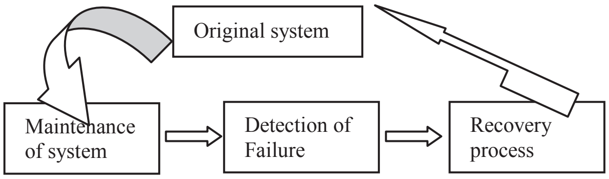


Figure 1: Self healing cycle

Self healing cycle has been shown in Figure 1. Critical issues [10] in self-healing systems typically include ; Maintenance of system health, recovery processes to return the state from an unhealthy state to a health one. Self-healing components or systems typically have the following characteristics [10] : (a) perform the productive operations of the system, (b) coordinate the activities of the different agents, (c) control and audit performance, (d) adapt to external and internal changes and (e) have policies to determine the overall purpose of the system. Most of the self-healing concepts are still in very early stages; still some possible areas explored are Grid computing, software agents, middleware computing, ad hoc networks. Emphasis here is on ad hoc network self healing characteristic.

A Mobile Ad Hoc Network, properly known as MANET [20] is a collection of mobile devices equipped with interfaces and networking capability. Hosts [19] can be mobile, standalone or networked. Such devices can communicate with another node within their radio range or one that is outside their range by multi hop techniques. An Ad Hoc Network is adaptive in nature and is self organizing. It is an autonomous system of mobile hosts which are free to move around randomly and organize themselves arbitrarily. In this environment network topology may change rapidly and unpredictably. The main characteristic of MANET strictly depends upon both wireless link nature and node mobility features. Basically this includes dynamic topology, bandwidth, energy constraints, security limitations and lack of infrastructure. MANET is viewed as suitable systems which can support some specific applications as virtual classrooms, military communications, emergency search and rescue operations, data acquisition in hostile environments, communications set up in exhibitions, conferences and meetings, in battle field among soldiers to coordinate defense or attack, at airport terminals for workers to share files etc.

In this paper an efficient weighted probabilistic algorithm for Mobile Ad Hoc network has been proposed to self heal, which considers the number of nodes in routing which can handle ideally, transmission power, mobility and battery power. The proposed algorithm selects a node with probabilistic weight considering effect of the three factors as Power factor, stable routing and backbone nodes on Ad Hoc networks. The rest of the paper is organized as: Section 2 presents a review of significant contribution in the area of routing for Ad Hoc networks and their limitations. In Section 3, the design philosophy and the Methodology of the routing scheme has been presented. Section 4 discusses the Simulation results and performance evaluation. Conclusions are given in the last section.

2. Routing

Routing protocol is needed whenever a packet needs to be handed over via several nodes to arrive at its destination. A routing protocol finds a route for packet delivery and delivers the packet to the correct destination. Routing Protocols have been an active area of research for many years; many protocols have been suggested keeping applications and type of network in view. Routing protocols can broadly classify into two types as (a) **Table Driven or Proactive Protocols:** here each node maintains one or more tables containing routing information to every other node in the network. All nodes keep on updating these tables to maintain latest view of the network. Some of the existing table driven or proactive protocols are: DSDV, GSR, WRP, ZRP and STAR. and (b) **On Demand or Reactive Protocols:** here routes are created as and when required. When a transmission occurs from source to destination, it invokes the route discovery procedure. The route remains valid till destination is achieved or until the route is no longer needed. Some of the existing on demand routing protocols are: DSR, DDR , TORA, AODV and RDMAR.

Study has been concentrated for reactive protocols because they work well in dynamic topology. Surveys of routing protocols for ad hoc networks have been discussed in [18,19,20].

2.1 Self Healing Issues

This section provides an analysis of various schemes that can be used as self healing schemes.

a) Self Healing in Routing : The most promising developments in the area of self-healing wireless networks are ad hoc networks. Automated network analysis through link and route discovery and evaluation are the distinguishing features of self-healing network algorithms. Through discovery, networks establish one or more routes between the originator and the recipient of a message. Through evaluation, networks detect route failures, trigger renewed discovery, and in some cases, select the best route available for a message.

b) Self healing in RF: Environmental radio-frequency (RF)[10,11] "noise" produced by powerful motors, other wireless devices, microwaves—and even the moisture content in the air can make wireless communication unreliable. Despite early problems in overcoming this pitfall, the newest developments in self-healing wireless networks are solving the problem by capitalizing on the inherent broadcast properties of RF transmission. The changes made to the network architectures are resulting in new methods of application design for this medium.

c) Self healing in Power efficiency: As the network is always on, conserving power is more difficult. One solution is On-demand discovery [11]. On-demand discovery networks are only "on" when called for. This allows nodes to conserve power and bandwidth and keeps the network fairly free of traffic. Once routes have been established, they must generally be maintained in the presence of failing equipment, changing environmental conditions, interference, etc. This maintenance may also be proactive or on-demand. Another solution can be Single-path routing[11]. As for routing, network algorithms that choose single-path routing, as the name suggests, single out a specific route for a given source-destination pair. Sometimes, the entire end-to-end route is predetermined. Sometimes, only the next "hop" is known. The advantage of this type of routing is that it cuts down on traffic, bandwidth use, and power use. If only one node at a time needs to receive the packet, others can stop listening after they hear that they're not the recipient.

3. Design Philosophy and Algorithm of a WNA

The objective of the proposed algorithm is to get a stable, power efficient protocol. To perform this role, however, a node does not require additional resources (e.g. buffers, processing power etc) since protocol support functions are well distributed among all nodes.

The network formed by the nodes and the links can be represented by an undirected graph $G = (V, E)$, where V represents the set of nodes v_i and E represents the set of links e_i. Note that the cardinality of V remains the same but the cardinality of *E* may change with the creation or deletion of links [5].

To decide the probability of a node to participate in routing, four main factors as its degree, transmission power, mobility and battery power are taken into account. Unlike other existing algorithms which are invoked periodically resulting in high communication overhead, proposed algorithm is adaptively invoked based on the mobility and battery power consumption of the nodes. Thus, concern is more on power awareness of each node so that node remains alive for the longer period of time. The Proposed algorithm is divided into two parts. In first part we consider the weights of nodes using parameters like mobility, degree, and its distance from its neighbors. In second part the available battery power of each node has been considered.

The *battery power* can efficiently be used within certain transmission range, i.e., it takes less power for a node to communicate with other nodes if they are within close distance to each other. However if nodes have maximum battery power to start with, then it would be more accurate metric to measure the power currently available at the node. This in turn depends on the node's initial power and the power expanded based on the actual network traffic and length of the links used to support it. Battery power of a node depends on two factors i.e. transmission range and type of applications. *Mobility* is considered as an important factor. It uses Random waypoint model. It is desirable to elect a node that does not move very quickly. Radio signals transmission is affected by interference, diffraction and shadowing. There should be some limit of speed so that transmission occurs effectively. Less mobility may cause more stable of topology also.

Whenever need for a new route arises in case of route break, check for network nodes are made, and a new route is established. Route tables are updated at each hello interval as in AODV with added entries for network nodes. Whenever a break in the route phase occurs due to movement of participant node, node damage or for other reasons; these idle nodes which have been termed as network nodes take care of the process and start routing. The whole process becomes fast and more packet delivery is assured. Each route table has an entry for number of network nodes surrounding it and their hop distance from the node.

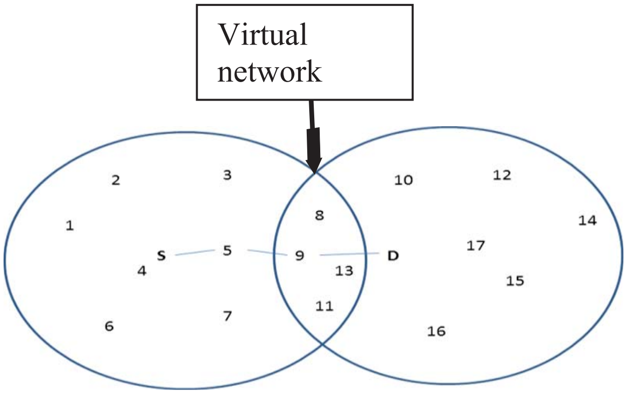


Figure 2: creation of virtual network, It is two hop situation.

Figure 2 represents creation of virtual network . Route established will be **S-5-9-D**. In case of failures of network, self healing nodes will be from virtual network established i.e. nodes 8,11,13. This is case of 2 hop situation. In case of multi hop refer to figure 3.

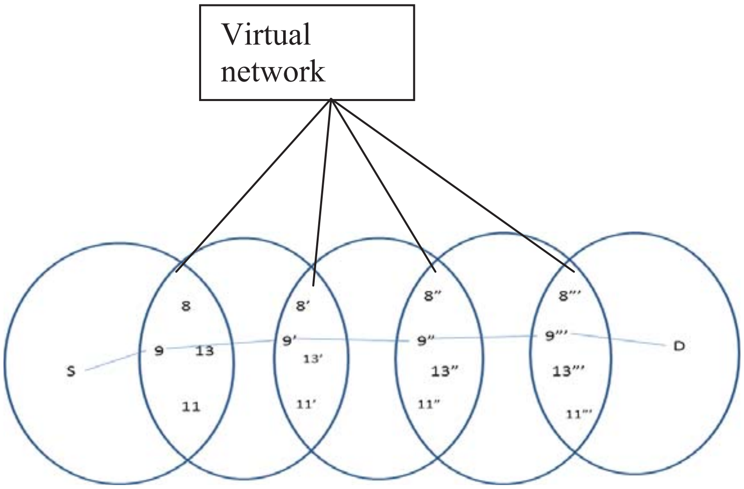


Figure 3: creation of virtual network for multi hop.

Now weight factor is calculated as per algorithm, this factor is updated in each route table at each beacon.

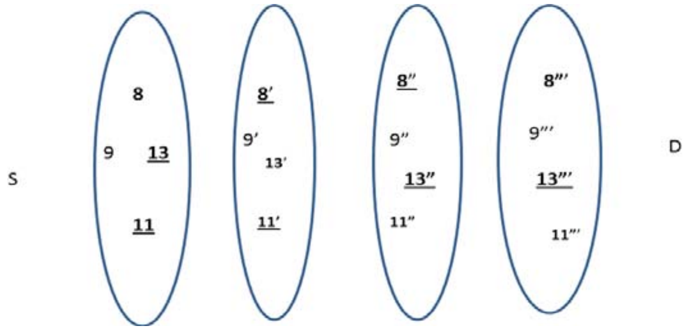


Figure 4: underlines nodes are weighted nodes to participate in case of failures.

In case of failures nodes that are ready to participate are 11,13,8',11',8",13",8"' and 13"'.

Based on these features, an algorithm called *weighted Network algorithm* (*WNA*) is designed. It effectively combines each of the above system parameters.

To select a node in a route, it is to be seen that it can remain for some specified period of time. It is assumed same battery power of all nodes at the initial stage. This assumption helps to consider the battery drainage which gives a direct measure of the available battery power. However, if the nodes have different battery powers to start with, then it would be a more accurate metric to measure the power currently available at the node. This in turn depends on the node's initial power and the power expended based on actual network traffic and length of the links used to support it.

The minimum probabilistic weight W_m of a node m is determined by Δm, the probabilistic degree, B_m the probabilistic mobility and S_m the probabilistic distance of the node m. The probabilistic degree Δm of a node can be defined by the ratio of the capacity of the node to the degree of the node d_r where d_r is given by [5]

$$d_r = \sum_{r'\in V, r'\neq r}\left\{dist(r,r') < t_{xrange}\right\} \qquad -- \quad 1$$

A node having degree less than the threshold value $\boldsymbol{\delta}$ always has Δm=0. As number of degree increases, it increases load on channel access. The probabilistic mobility B_m of the node is incorporated by the ratio of the actual mobility of the nodes B_r to the maximum allowed mobility of the nodes B_e. The value of B_r can be computed as follows [5]

$$B_r = (1/T)\sum_{t=1}^{T}\sqrt{(u_t - u_{t-1})^2 + (v_t - v_{t-1})^2} \qquad -- \quad 2$$

Where (u_{t-1}, v_{t-1}) and (u_t, v_t) are the coordinates of the node r at time $(t\text{-}1)$ and t, respectively and T is the time period. Also, it is assumed that

$$\left\{B_r \leq B_e \| B_r = 0 then B_m = 0\right\}$$

The third component, S_m plays the role of power consumption. The probabilistic distance S_m is given by the ratio of the sum of distances S_r with its neighbors and the maximum distance sum S_n. The sum of the distances S_r with its neighbors is defined by

$$S_r = \sum_{r'\in N(r)}\left\{dist(r,r')\right\} \qquad --- \quad 3$$

The motivation of S_r is mainly related to energy consumption. It is known that more power is required to communicate to a larger distance. And, the maximum distance sum S_n is given by

$$S_n = t_{xrange} \times d_r \qquad --- \quad 4$$

As the nodes moves, communication may become difficult due to mainly signal attenuation with increasing distance. Like in [13], $\boldsymbol{E_t}$ is assumed as energy cost to communicate information through free space directly between two nodes. $\boldsymbol{E_t}$ is a strong function of distance d between the nodes. More precisely, $\boldsymbol{E_t}$ is defined by

$$E_t = \beta \times d^{\gamma}, \qquad --- \quad 5$$

Where γ >1 as path lose exponent and β is a proportionality constant describing the overhead per bit.

Based on the above discussion the algorithm has been designed which consists of following steps.

1. Compute the *probabilistic degree* Δm by (δ / dr) where

$$d_r = |N(r)| = \sum_{r'\in V, r'\neq r}\left\{dist(r,r') < t_{xrange}\right\}$$

and δ is a threshold value. If $d_r \leq \delta, \Delta m$ is assumed to be 0.

2. Compute the *probabilistic mobility* $B_m = B_r / B_e$, where B_r is the running average of the speed for every node till current time T.

3. Compute the *probabilistic distance sum* $S_m = (S_r / S_n)$

4. Compute the combined *probabilistic weight* W_m for each node m as the sum of three factors Δm, B_m and S_m.

5. Compute the energy $\boldsymbol{P_{req}}$ of the node m by $P_{req} = (d_a \times T_m) \times C \times E_t$ where d_a is the delay between the signal propagation, T_m is t_{xrange} / B_e, c is the link capacity for a given physical range and $\boldsymbol{E_t}$ is the energy per bit.

6. Broadcasts its id, W_m and $\boldsymbol{P_{req}}$ by each node at each beacon **.**

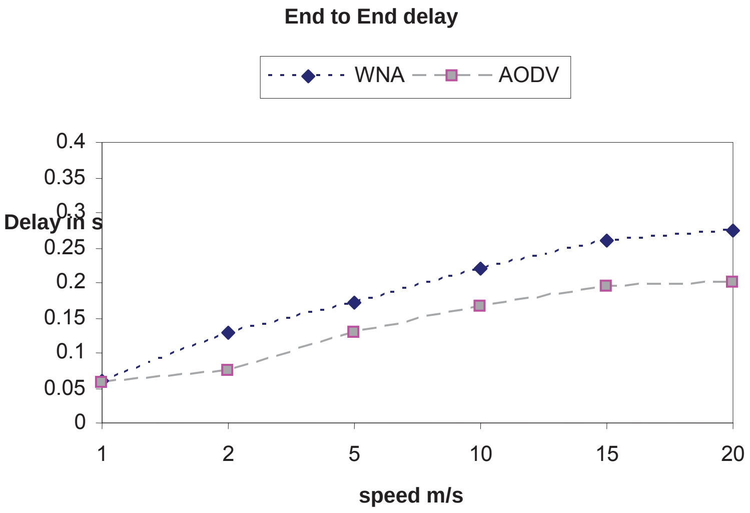
The weight factor is added to routing table of each node. This broadcast occurs at each beacon of AODV. The algorithm has been incorporated on AODV. Changes have been made and protocol is executed on simulator to check its efficiency.

4. Simulation

This section deals with the simulation of an environment having nodes on a 1Km ×1 Km grid. It is assumed that the nodes can move in all possible directions with displacement varying uniformly between 0 to a maximum value in one unit of time.

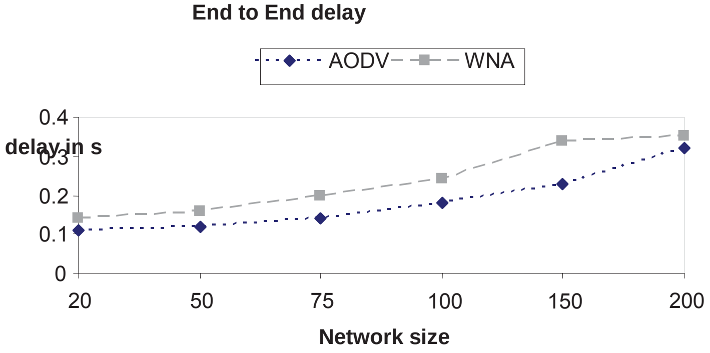
Simulation study has been carried out to study the Performance study of proposed protocol. Simulation Environment used is NS-2 (network simulator) version NS2.34 to carry out the process. Simulation results have been compared with AODV. Simulation study has been performed for packet delivery ratio, Throughput and End to End delay evaluations.

In simulation study 100 nodes were taken in a random scenario of size 1 km × 1 km. Two parameters have been taken as Pause time and speed. The study has been conducted at different pause times. Pause time of 0 means maximum mobility and 500 is minimum mobility. The sources connected are 20-24 using TCP connection.

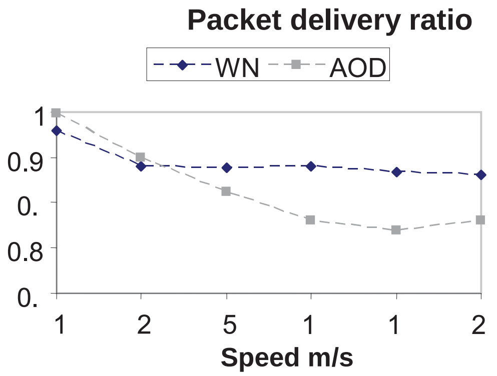


Graph 1: delay for 100 nodes using speed as function

This graph represents delay calculated at various speeds. It is clear that WNA actually has more delay. But the reason is obvious that it requires more calculations. But in the end it provides better routes. Graph 2 has been used to show the effect of delay using various size of network nodes. Number of nodes have been varied and effect has been depicted in Graph-2. Delay is more and it is as per the theory expected.

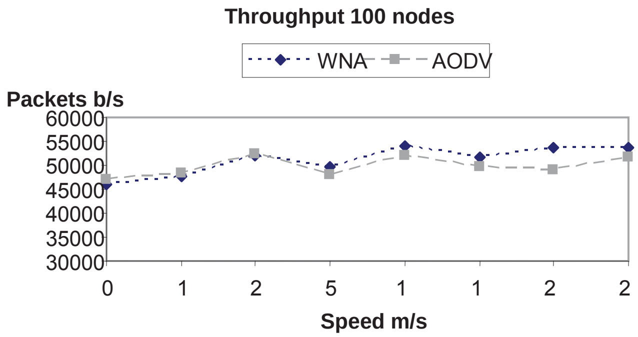


Graph 2: Delay calculated using various network size.

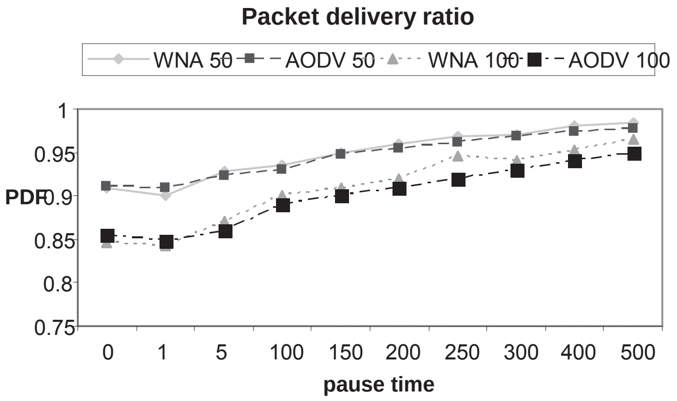


Graph 3: PDR using speed as function

Packet delivery ration denotes that WNA is much better in terms of packet delivery. Moe packets have been delivered. Same is case with Throughput. Graph-4 indicates that WNA has better throughput than AODV at all speeds.



Graph 4: throughput using speed as function



Graph 5: comparison using pause time as function

Using pause time as a function a scenario has been generated using various network sizes. Use of AODV and WNA has been shown in graph-5. It is evident from the graph that WNA outperforms AODV in all cases and at all intervals of pause time.

5. Conclusions

A new scheme has been presented that utilizes weight as a factor for better routing. The scheme can be incorporated into any ad hoc on-demand unicast routing protocol to improve reliable packet delivery in the face of node movements and route breaks. Alternate routes are utilized only when data packets cannot be delivered through the primary route. As a case study, the proposed scheme has been applied to AODV and it was observed that the performance improved. Simulation results indicated that the technique provides robustness to mobility and enhances protocol performance. It was found that overhead in this protocol was slightly higher than others, which is due to the reason that it requires more calculation initially. This also caused a bit more end to end delay. The process of checking the protocol scheme is on for more sparse mediums and real life scenarios and also for other metrics like Path optimality, Link layer overhead.

6. References:

[1] A. Amis and R. Prakash, *Load balancing clusters in wireless ad hoc networks*, Proceedings of ASSET 2000, Richardson, TX, March 2000, pp. 25–32.

[2] D. J. Baker and A. Ephremides, *A distributed algorithm for organizing mobile radio telecommunication networks*, Proceedings of 2nd International Conference on Distributed Computer Systems, April 1981, pp. 476–483.

[3] S. Basagni, I. Chlamtac and A. Farago, *A generalized clustering algorithm for peer-to-peer networks*, Proceedings of Workshop on Algorithmic Aspects of Communication, July 1997.

[4] A.Kush,Divya,Vishal ," Energy efficient Routing for MANET" , Intl conf on applied and communication tech, **Elsevier Publ**, pp 189-194., 2014

[5] M. Chatterjee, S. K. Das and D. Turgut, *WCA: A Weighted Clustering Algorithm for Mobile Ad Hoc Networks*, Cluster Computing 5, Jun 2002, pp. 193-204.

[6] I. Chlamtac and A. Farago, *A new approach to the design and analysis of peer-to-peer mobile networks*, Wireless Networks 5(3), August 1999, pp. 149–156.

[7] M. Gerla and J. T. C. Tsai, *Multicluster, mobile, multimedia radio network*, Wireless Networks 1(3), 1995, pp. 255–265.

[9] C.-H.R. Lin and M. Gerla, *A distributed architecture for multimedia in dynamic wireless networks*, Proc. IEEE GLOBECOM*,* 1995, pp. 1468–1472.

[10] A.K.Parekh, *Selecting routers in ad-hoc wireless networks*, Proceedings of SBT/IEEE International Telecommunications Symposium, August 1994.

[11] A. B. McDonald, T. F. Znati, *Design and performance of a distributed dynamic clustering algorithm for ad hoc networks*, Proceedings of 34th Annual Simulation Symposium, 2001, pp. 27-35.

[12] C. E. Perkins, *AD HOC NETWORKING*, Addison-Wesley, 2001, pp. 76.

[13] J. M. Rabaey, M. J. Ammer, Julio L. da Silva Jr., Danny Patel, Shad Roundy, *Pico Radio Supports Ad Hoc Ultra-Low Power Wireless Networking*, IEEE Computer Magazine on Wireless Computing.

[14] M. S. Corson, S.Papademetriou, P.Papadopoulos, V.Park, and A.Qayyum, *An internet MANET encapsulation protocol (IMEP) specification.* Internet Draft, August 1998.

[15] C. E. Perkins and P. Bhagwat, *DSDV Routing over a Multihop Wireless Network of Mobile Computers*, *AD HOC NETWORKING*, Addison-Wesley.

[16] Z. J. Hass and M. R. Pearlman, *The Zone Routing Protocol(ZRP) for Ad hoc Networks*, Internet draft - Mobile Ad hoc Networking (MANET) Working Group of the Internet Engineering Task Force (IETF), November 1997.

[17] Kaixin Xu and Mario Gerla, *A Heterogeneous Routing Protocol based on a new Stable Clustering Scheme*, IEEE MILCOM 2002, Anaheim, CA, Oct. 2002

[18] A.kush, Sunil Taneja, "A Survey of Routing Protocols in Mobile Ad Hoc Networks" in International Journal of Innovation, Management and Technology, Vol. 1, No. 3, August 2010, pp 279-285

[19] A.Kush, Divya, "Performance comparison of energy efficient AODV protocols" in International Journal of Computing and Business Research IJCBR, Vol 2, Issue 1 January 2011.

[20] A.Kush, Seema "Evaluation of Routing Schemes for MANET" in A. Mantri et al. (Eds.): HPAGC 2011, CCIS 169, pp. 575–580, 2011. © Springer-Verlag Berlin Heidelberg 2011

[21] A.Kush, D. Sharma, "Maximizing Lifetime of a Mobile Ad hoc network using energy Efficient Routing:", ICIRITO 2014**, IEEE Explore** , 2014. Pp1-5

Baseband Modem Design for Multiparty Voice Communications

Dong Jegal and Byoungchul Ahn
Department of Computer Engineering, Yeungnam University, Gyungsan, Korea

Abstract - *This paper presents a design of the TDMA baseband modem to support 4-party voice communications. The baseband modem supports an ad-hoc function to join the voice networks. The designed baseband modem is expandable by adding slots up to 8 nodes since the bandwidth is limited TDMA method and voice quality. To implement the 4-party communications, the baseband modem has a cycle network controller, a sync time controller, a buffer controller and an asynchronous serial device. The sync time controller is implemented to synchronize precisely to the master's control signal for slaves. The protocol is programmed on Xilinx Zynq 7000 with Verilog HDL. The baseband modem is tested and verified its functions for voice communications, and its measured maximum delay time is less than 30msec for voice transmission.*

Keywords: Ad-hoc Network, Multi-party, TDMA, Baseband Modem, WPAN

1 Introduction

WPANs(Wireless Personal Area Networks) are applied to many places such as small group meetings, short distance multi-way communications, remote speakers and so on. Typically WPANs do not have a multiparty communication function. Also it does not have functions to join networks or voice data using an ad-hoc function. Also the communication distance of WPANs is very limited because of sharing ISM bandwidth. For small group communications, it is much efficient to use TDMA technology than CDMA technology since TDMA increases the efficiency of the channel by removing collisions.

To implement the ad-hoc function for voice communications, there are several technical challenges in ad-hoc networks. First, ad hoc networks are characterized by high bit error rates and path breaks due to changing network topology. High bit error rates reduce the quality of the network service. Second, the transmission frame of wireless networks are included not only preamble for synchronization but also a payload of limited length. Therefore the length of the data packets that are available in the ad-hoc network is short in the wireless network. There is a disadvantage in multi-hop wireless networks. This is why the throughput of data is reduced as the number of nodes increases. In particular, the transmission delay of wireless network communication is increased as the number of hops is increased. But there is no product to support the relay protocol and the ad-hoc function for WPANs.

As processors speeds are increased, it is possible to overcome the above technical challenges. This paper presents multiparty voice communications with a relay protocol based on the TDMA technology and the ad-hoc function[1].

This paper presents a design of the baseband modem for multiparty voice communications. The basic protocol is described in Section 3 and the design of baseband modem is described in Section 4 and 5.

2 Related work

There are many researches for ad-hoc networks. For real-time speech on wireless ad-hoc is studied by Kargl, Kwong and Venkat[4][5][6]. Frank Kargl *et al*. have discussed voice transmission over Bluetooth and presented a new routing protocol called Bluetooth Scatternet Routing(BSR)[4]. But they have discussed its possibility, and the chip of BSR has not been implemented. Kwong *et al*. have used multi-path routing protocol called MSDR to improve speech quality[5]. But processing overheads have not been solved.

G. Venkat Raju *et al*. have proposed a Localized Distributed heuristic for Minimum number of Transmissions(LDMT)[6]. In order to reduce transmission delay, this algorithm minimizes voice retransmission only. Several researches have studied to solve the problem of capacity reduction in multi-hop wireless networks[11][12]. They have observed that the performance degrades quickly as the number of hops increases due to using a single radio for transmitting and receiving packets. A good way to improve the capacity of wireless is to use more network interfaces or to use speech compression in the case of voice applications. Another way to improve the capacity of wireless is to use schedule transmission slots in time and to use multiple non-interrupting frequency channels[13]. Chen *et al*. have observed transmission traffic is decreased as the number of hops increases when single frequency is used in wireless networks[12]. There is no related paper which has been implemented the baseband modem with multi-party communications and ad-hoc function.

3 Network protocol

The proposed network cycle time is 92.5*msec* and expandable up to sixteen slots to communicate 4 persons at the same time as shown in Figure 1. The time length of each slot is 18.5*msec* and each slot has 5frames and each frame is 3.7*msec*. Each node uses one of its own control section assigned in its frame. Slot 0 is used for the start of cycle(SoC) and allocated for contention slot. If a new node requests to join the network, it should use this contention slot. If there are collisions, random back-off time is used to join the network. The first frame of each slot is used for control frame, which is used by the master node and all slave nodes to receive network information.

In Figure 1, 5 slots are designed to implement the TDMA network cycle for four-party voice communication. First slot is used SoC and four slots are used for voice communications. The second slot is the master slot, and other slots are slave slots under the control of the master.

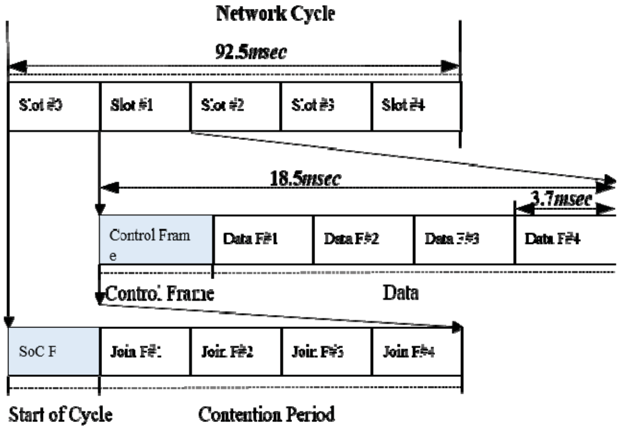


Figure 1. Network Cycle of Relay Protocol

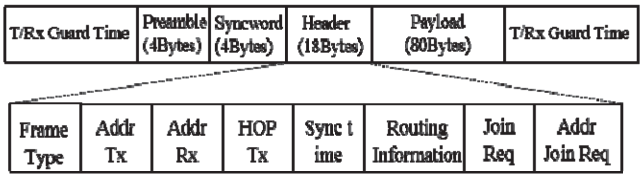


Figure 2. Basic Structure of Frame

The basic structure of each frame as shown in Figure 2 consists of six sections, which are two Tx/Rx guard time, preamble, sync word, header, and payload. Two Tx/Rx guard time is 500 μsec of Tx/Rx turnaround time of RF modules and 500 μsec of guard time for adaptive synchronization of slave nodes. 2*msec* guard time can adjust synchronization timing among nodes like the IEEE802.15.4e-2012 TSCH. The preamble is used for stabilizing time of the frequency generator of RF modules and users can specify the length of preamble for specific RF chips. Sync word is 4-byte and has a special pattern in order to synchronize the payload data.

$$\text{Frame time} = 2\,(T_{rf} + T_g) + T_b \qquad (1)$$

where, T_g is guard time for adaptive synchronization of slave nodes,
T_{rf} is guard time for transmission and reception time,
T_b is frame time except guard time.

The header has frame type, the addresses of the sender and receiver, sync time, network routing information and join request information. Sync time measures the transmission time of the master and make slaves adaptive synchronization. Payload is for 64Kbps voice data. The voice is sampled 8-bit 8KHz sampling frequency and 80 Bytes can be sent in 10 *msec*. When communication bandwidth is 500Kbps and the size of payload is 80Bytes, T_b is 1.7 *msec* and frame time is 3.7 *msec* from Equation (1). The time length of each slot is 18.5 *msec* and one network cycle time is 92.5 *msec* for four-party communication. For *N*-party voice communications, it is easy to expand easily by adding *N* frames. All joined nodes can transmit voice data every 18.5 *msec*. The number of party is limited up to 8 by the TDMA method and voice quality.

3.1 Start of cycle

The SoC(start of cycle) is generated and transmitted by the master. It indicates the start of the network cycle and all slave nodes should be synchronized to this signal to work on the network. This includes sync time code for synchronization, which is the actual transmission time calculated by the internal timer. To synchronize this cycle, the default common frequency is used to all nodes.

The contention frames followed by the SoC are used to request to join the network and total contention frames are 4 frames. The slave which joins the network sends the message "*connection request* (*JoinREQ*)" to the master and the master replies back the message "*connection permission* (*JoinACK*)" with its node number and network address for routing. After the number of joint slaves is equals to the number of 4 frames, the contention slot can be used for voice communications.

3.2 Control frame and data frame

The control frame is generated and sent by the master. The control frame is used to manage the network using the common frequency. This includes frequency and hopping data. The master node uses the header of the control frame to exchange routing information. All Slave nodes can join the network after they look at routing information of the header.

Data frames are divided into 4-party. Nodes at the network have their assigned frame number from routing information and transmit their voice data at the assigned frame number. Slaves are at listening mode and receives voice data from other nodes. Nodes not joined the network can receive only voice data even though they cannot transmit voice data.

4 Modem design

The baseband modem for 4-party voice communication protocol is implemented on Xilinx Zynq 7000 using Verilog HDL. The baseband modem block diagram is shown in Figure 3. The modem has a network cycle controller, buffers and a buffer controller, a sync time controller, an asynchronous serial transceivers.

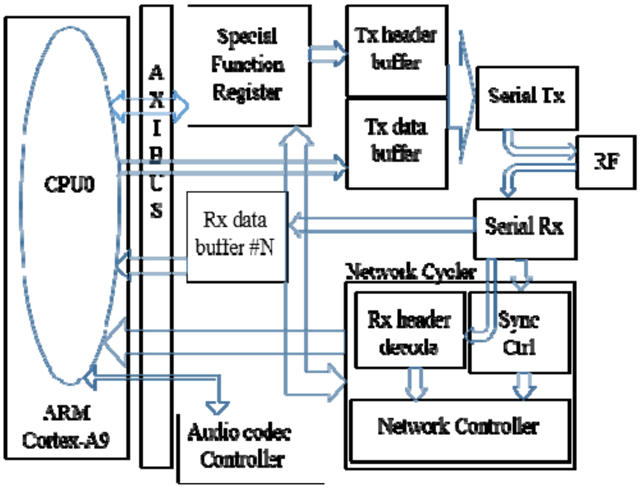


Figure 3. Block Diagram of the Baseband Modem

4.1 Asynchronous serial transceivers

The baseband modem is designed to communicate an asynchronous RF module chip. The RF frequency is ISM(Industrial Scientific and Medical) band. The asynchronous serial transceiver uses a message-bit counter and a comparator to detect a Sync word and a shift register is used to convert serial data to parallel data or vice versa.

4.2 Buffer memory and controller

To interface the baseband modem with the processor and a RF module, buffer memories are implemented. The buffer memories store voice data from the processor and send them to the RF module. They store data from the RF module and pass them to the processor to play voice data. Therefore, one Tx buffer and four Rx buffers are implemented to handle data from four slaves simultaneously. Each buffer is assigned to one of four data sections and is used to a dedicated voice data. Nodes have one transmission buffer because they transfer one voice data at a given time but they have *N*-1 receive buffers to receive data for 4-party communications. The header data is updated and transmitted according to the network cycles and the status of nodes. All data is designed to access AXI bus of the Zynq 7000 chip. To meet the processing time of the baseband modem, a special function register is designed for fast memory access.

4.3 Network cycle controller

The network cycle controller controls the basic cycle function of the baseband modem. It enters a state of frequency search after it is initialized. In this state, the modem determines whether it is the master node or not. If the master mode is set, immediately it transmits the SoC to provide a service to slave nodes to operate TDMA network cycles.

If the slave mode is set, it searches the SoC with the common frequency. After it receives the SoC, it synchronizes the network cycle times by the sync time controller. When slave nodes are scanning mode, the sync time controller stops an internal timer for synchronization and set a timer to the sync time data of the header. The master and slaves maintain the network cycle time. In Figure 4, if slaves do not receive signals of SoC and control frame from the master during one network cycle time, slave nodes stop the network cycle and start scanning mode again to synchronize again. Each slave is synchronized for one network cycle time.

And then it sends out the message *"connection request (JoinREQ)"* and searches the control frame of the master and retrieves its routing address assigned by the master. If a slave node receives routing address from the master, it works for transmitting and receiving voice data by joined node.

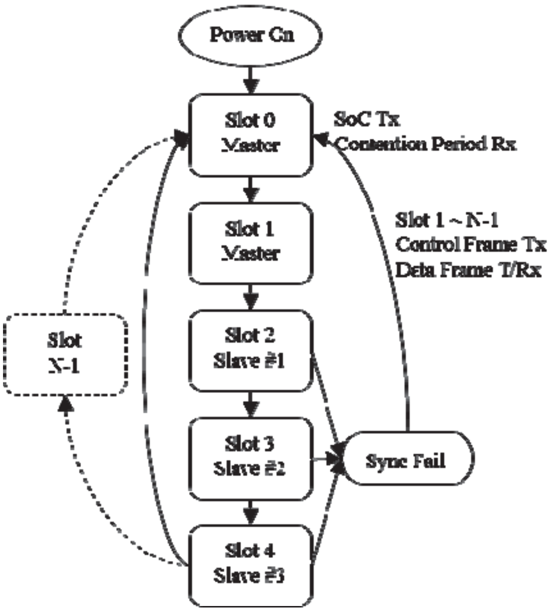


Figure 4. State Diagram of Network Cycle

5 Implementation

In order to evaluate relay protocol functions of the baseband modem, it has been assembled on the system board shown in Figure 5. The system board has a Zynq 7000 SoC chip from Xilinx, a CC2500 chip and a CC1200 chip from TI and an ADAU1761 audio codec chip from Analog Devices. The Zynq 7000 SoC integrates an ARM dual Cortex-A9 based processor system with Xilinx 7-series FPGA and Xilinx Vivado ver14.2 is used as a design tool.

The baseband modem is programmed on Xilinx Zynq 7000 SoC using Verilog HDL. The CC2500 chip is a 2.4GHz RF module and the CC1200 chip is a Sub-1GHz RF module. The ADAU1761 chip is used for encoding and decoding voice data. Voice data and routing and the baseband modem controlled by ARM dual Cortex-A9 based processor system.

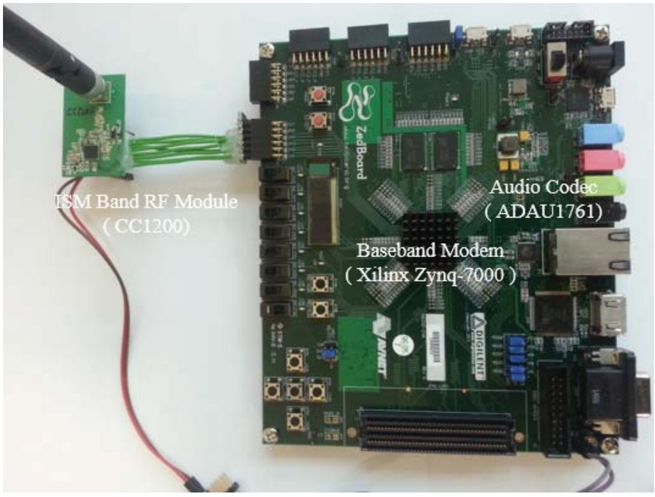


Figure 5. Design Evaluation Board

In order to evaluate the performance of the modem, experiments have been carried out as shown in Figure 5. First, the master node sends the first frame with the SoC and the control frame according to the network cycles. Typically, slave nodes received voice data from the master and other nodes. At transmission mode, slave nodes transmit voice data at the designated slot from 4 data frames.

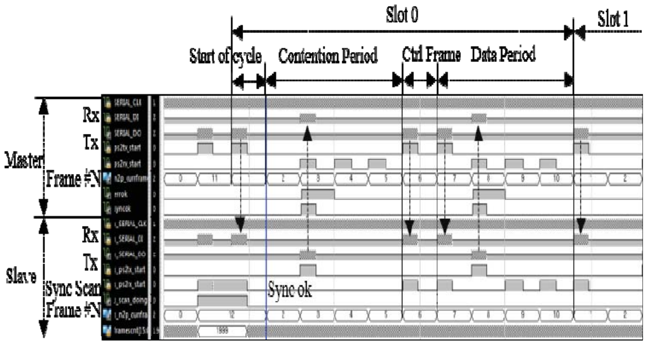


Figure 6. Output Signals of the Network Cycle Controller

Figure 6 shows simulation output signals measured on FPGA pins when the master node transmits data to Slave node #1. The master sends out the start of cycle signal after it starts power-on or the network cycles. Slave nodes starts scanning mode to find the start of cycle. The sync time controller waits the SoC and stops the internal timer. It synchronizes by receiving the start of cycle and keeps synchronizing precisely as shown in Figure 6. And it synchronizes to the control signal of the master. Slaves send "*Join request*" signal during the contention period. After slaves join the networks, they send out their voice data to the networks at the designated frame and receive voice data.

6 Conclusion

In this paper, a TDMA baseband modem has been designed on Zynq 7000. Its performance has been verified by experiments and tested its performance to four-party voice communication. The voice data is sampled at 8-bit 8KHz sampling rate and the tested RF frequency is 900MHz. Each node can send and receive signals by its assigned time slot. To implement the 4-party communications, the baseband modem has a cycle network controller, a sync time controller, a buffer controller and an asynchronous serial device. The sync time controller is implemented to synchronize precisely to the control signal of the master for slaves. The protocol is programmed on Xilinx Zynq 7000 with Verilog HDL. Experiment results show the stable communication channel and the synchronization of the network cycle between the master and slaves.

The WPAN network can be configured as various topology such as line, star or tree and so on. Also the modem can be used with various RF modules for specific applications with different data rates.

Please address any questions related to this paper to Byoungchul Ahn by Email (b.ahn@yu.ac.kr).

7 References

[1] B. Ahn, S.-H. Hwang, C.-H. Park, S.-H. Moon, "Small Group Relay Protocol using TDMA Contention", Proceedings of ICWN, pp. 182-188, CSREA, Jul. 2012.

[2] N. Jain, S. R. Das, and A. Nasipuri, "A multichannel CSMA MAC protocol with receiver-based channel selection for multihop wireless networks," in Proceedings of the 9th International Conference on Computer Communications and Networks, pp.432-439, 2001.

[3] C. H. Lin, H. Dong, U. Madhow, A. Gersho, "Supporting Real-Time Speech on Wireless Ad-hoc Networks: Inter-packet Redundancy, Path Diversity, and Multiple Description Coding", in Proceedings of ACM workshop on WMASH, pp.11-20, Oct. 2004.

[4] F. Kargl, S.Ribhegge, S. Schlott, M. Weber, "Blue tooth-based Ad-hoc Networks for Voice transmission", in

Proceedings of 36th Annual Hawaii International Conference on System Sciences, Jan. 2003.

[5] M. Kwong, S. Cherkaoui, R. Lefebvre, “Multiple description and multi-path routing for robust voice transmission over ad-hoc networks”, in IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob, pp.262-267, 2006.

[6] G. Venkat Raju, T. Bheemarjuna Reddy Shyamnath Gollakota, and C. Siva Ram Murthy, “A near optimal localized heuristic for voice multicasting over ad-hoc wireless networks”, in Communications, 2007 ICC’07. IEEE International Conference on, pp. 1648–1653, June 2007.

[7] D. B. Johnson, D. A. Maltz, “Dynamic Source Routing in Ad-hoc Wireless Networks”, in Computer Communications Review – Proceedings fo SIGCOMM’ 96, Aug. 1996.

[8] S. Corson, J. Macker, “Mobile ad-hoc networking (MANET): Routing protocol performance issuse and evaluation considerations”, IETF 1999.

[9] T. Camp, J.Boleng, V.Davies, “A survey of mobility models for ad-hoc network research”, Wireless Communications & Mobile Computing(WCMC): Special Issue on Mobile Ad-hoc Networking: Research, Trends, and Applications, Vol. 2, no. 5, pp. 483-502, 2002.

[10] P. Gupta and P. R. Kumar, "The capacity of wireless networks," IEEE Trans on Info Theory, Mar 2000.

[11] J. Li, C. Blake, D. S. J. De Couto, H. I. Lee, and R. Morris, "Capacity of ad-hoc wireless networks," in MOBICOM, 2001.

[12] T. W. Chen, J. T. Tsai, and M. Gerla, "QoS routing performance in multihop, multimedia, wireless networks," in Proceedings of IEEE ICUPC'97, 1997.

[13] Yu-Ching Hus, Tzu-Chieh Tsai, Ying-Dar Lin, and Mario Gerla, "Bandwidth routing in multi-hop packet radio environment," in Proceedings of the 3rd International Mobile Computing Workshop, 1997.

[14] Xilinx device manual, Zynq-7000 all programmable SoC overview. December 2013. http://www.xilinx.com/support/ documentation/data_sheets/ds190-Zynq-700-Overview.pdf.

[15] Xilinx device manual, Zynq-7000 All Programmable SoC Technical Reference Manual, UG585 (v1.10) February 23, 2015.

[16] Xilinx device manual, AXI Reference Guide, UG761(v14.3) November 15, 2012

[17] Cha, Bong-Sang, Jeong, Eui-Hoon, Jeon, Gwangil, Seo, Dae-Young, " A study on improvement of ISO/IEC 29157 MAC protocol," The Journal of the Institute of Webcasting, Internet and Telecommunication, Volume 13, Issue ,5, 2013, pp.17-26.

[18] Tengfei chang, Thomas Watteyne, Kris Pister, Qin Wang, "Adaptive synchronizatioin in multi-hop TSCH networks," Computer Networks, Volume 76, 15 January 2015, Pages 165-176.

SESSION

GREEN NETWORKS AND ENERGY EFFICIENT SYSTEMS: METHODS AND POWER AWARE SYSTEMS

Chair(s)

TBA

A Survey of Base Station Sleeping Technologies for Green Cellular Networks

Mwashita Weston [1], **Odhiambo Marcel Ohanga**[2]
[1]College of Science, Engineering and Technology, University of South Africa, PO Box 5172 Windhoek, Namibia
[2]Department of Process Control and Computer Systems, Faculty of Engineering and Technology
Vaal University of Technology, Private Bag X021 Vanderbijlpark,1911, Andries, Portgieter, South Africa

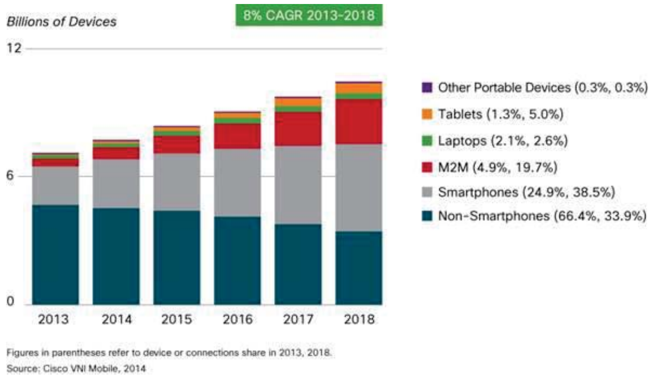
Abstract - *This paper presents a survey of the latest technologies that have been advanced by both academia and industry in an attempt to reduce the energy consumed by Base stations (BS) in cellular networks. Since BSs are the primary energy consumers in cellular networks, BS sleeping technologies are promising proposals in reducing BS energy consumption. The main goal of the survey is to gain an in-depth understanding of the benefits and shortcomings of these proposed technologies. The survey presented the authors with an opportunity to offer clear insights to researchers working on Green Cellular Networks for them to choose and adapt the most efficient ways of reducing BS energy consumption without compromising Quality of Service (QoS).*

Keywords: Green Cellular Networks, BS sleeping, energy efficient networks, green base stations, cell zooming.

1 Introduction

Lately, there has been an exponential growth [1] in mobile cellular systems as mobile data services are well on their way to becoming necessities for many network users. Many people nowadays are ever demanding ubiquitous wireless and Internet services. Cisco, in [2] predicts that globally, mobile devices and connections will grow to 10.2 billion by year 2018. Authors in [3] forecast that by 2020, over thirty billion things with over two hundred billion intermittent connections will be in place. Demand for multimedia-rich mobile communication devices like smart phones has been on the upward trend. Figure 1 shows this enormous growth as predicted by CISCO.

One network provider, China Mobile, has been doubling its number of BSs [4] in order to provide better network coverage and capacity. The operators are forced to deploy more and more BSs per unit area to meet the ever increasing traffic demand. The tremendous increase in BSs has resulted in an exponential increase in energy consumption and carbon footprint especially in remote areas that rely on diesel generators for their power requirements**.** [5]



Figures in parentheses refer to device or connections share in 2013, 2018.
Source: Cisco VNI Mobile, 2014

Figure1: Global mobile devices, connections growth [2]

Renewable power sources like solar and wind generators are being used though to a limited extent in cellular networks but their only problem is that of uncertainty. They are weather driven. Weather is highly unpredictable and this makes the amount of generated energy also to be unpredictable. For solar energy, the amount of energy produced depends on the availability of the sunlight.

The main power supply for BSs comes from the electrical grid. Most of the electricity generation methods produce CO_2 emissions. Nuclear generation for example, produces CO_2 emissions in the uranium enrichment process. Authors in [5] give a figure of 650Kg CO_2 /MWh. Japan's Central Research Institute of the Electric Power Industry published a life cycle CO_2 emission figures for the various electricity generation technologies and the results are as shown in table 1.

Table 1: CO_2 emissions of various electricity generation technologies in Japan, Sweden and Finland [5]

g/kWh CO_2	Japan	Sweden	Finland
coal	975	980	894
gas thermal	608	1170 (peak-load, reserve)	-
gas combined cycle	519	450	472
solar photovoltaic	53	50	95
wind	29	5.5	14
nuclear	22	6	10 - 26
hydro	11	3	-

The figure for British Energy's Torness Nuclear in 2002 was 5.05g/KWh [5]. All this proves that even electricity production technologies that do not seem to result in CO_2emissions actually do emit the gas.

From table 1, it can be seen that coal, gas thermal and gas combined cycle, produce the highest percentage of CO_2 emissions. As of 2010, according to statistics that was availed by the International Energy Agency (IEA), slightly over 81% [6] of the world-wide energy consumption comes from oil, gas and coal which happen to be the sources of energy that have highest CO_2 emissions. Scientists have estimated that ICT (with cellular communication systems included) will be responsible for 3% [7] of all global emissions by year 2020. This figure is the same as the percentage that is contributed by all the airlines combined. It is not only the environment that is affected by the high energy consumption, operational expenditure, commonly known as OPEX is substantially increased. Authors in [4] reported that a collective cellular network OPEX of USD22 billion was incurred in 2013 alone.

This has led researchers from industry to team up with academia to carry out research projects on what has been coined "Green Cellular Communication" [8]. Green Cellular Communication refers to the practice of using energy efficient cellular communication technologies, minimising resources whenever possible thereby limiting the amount of CO_2 greenhouse gas emissions. The EARTH (Energy Aware Radio network technology) [9] under the European Framework Program 7, was a major European research project that had 15 partners that included 10 countries which focussed in energy efficiency in the next generation access networks.

Reducing the energy that is consumed by BSs has recently become a very important research topic. This is because a BS is the greatest energy consumer in a cellular mobile network and as such produces the highest CO_2 emissions as shown in figure 2.

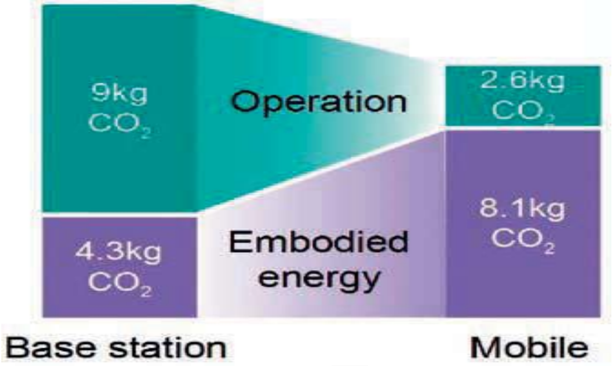


Figure 2 Network components, CO_2 emissions [5]

According to [5], a BS can consume 60-80% of the total energy that is consumed in a cellular mobile network. Several methods have been identified that are all aimed at reducing BS energy consumption. Scientists estimate that 65% of energy consumed by a BS is consumed by a power amplifier (PA). For this reason, several methods have been suggested to improve the efficiency of the power amplifier. Special materials that include Si and GaAs can be used for the construction of the amplifier. The Doherty amplifier [10] is one amplifier that has been designed with efficiency as a prerequisite and a drain efficiency of 50% can be achieved. Some methods involve the reduction of energy lost in the AC/DC conversion in the power supply section. Scientists have also designed high efficient air conditioning systems for cooling BS components. Beam forming and OFDMA technologies are other methods that are used to reduce energy consumption at the BS site. There have been quite a number of surveys on technologies that can be used to reduce BS energy consumption. Most of these surveys[7,11,12,13,14,15,16,16,17,18,19] have tended to include a cocktail of schemes used to reduce energy consumption. This paper will only focus on the latest BS sleeping technologies. Section 2 will give the various recent BS switching technologies that have been proposed. The benefits and shortcomings of these technologies are discussed in this section. Section 3 gives the concluding remark, implications and applications of the research survey.

2 BS SWITCH ON/OFF TECHNOLOGIES

Authors in [20] proposed a method of switching underutilised BSs in a smart way in a bid to save energy. Their first method makes use of the distance between the BS and the UE. The algorithm proposed switches off BSs that have the maximum average distances after estimating distances of associated UEs. They reasoned that a higher power is naturally required by both BSs and UEs to connect distant users. The remaining BSs are then made to cover those areas where BSs would have been switched off. The scheme involves BSs estimating the distances of UEs they are serving and then taking an average which they share amongst themselves. The BS with the highest average is the one to be switched off first provided that action does not result in the degradation of QoS. According to the authors, energy savings of 40% were achieved using this method.

The authors in [20] went on to propose another BS sleeping algorithm that switches off BSs according to traffic load variations. Certain BSs are switched off from 7pm to 7am when traffic is very low. In the morning, BSs are switched back on, gradually to follow the rate at which traffic will be picking up. With this strategy energy saving of 70% was achieved.

Their last proposal was an algorithm that they used to find the maximum number of BSs to be switched off. The algorithm involves the selection of the most energy efficient combination of active BSs to be switched off.

This combination is then switched off. This last algorithm that the authors used is very effective as the maximum number of BSs at any given point in time is switched off. This results in considerable amount of energy savings. The authors managed a 71% energy saving using the scheme. Their solution is also not so complex hence implementation is very easy. However, the solution is silent on the actual process involved in re-associating UEs to the remaining active neighbouring BSs. This is a very delicate process which, if not handled properly will result in a compromised QoS.

The scheme that was presented by authors in [1] involves the monitoring of individual BS activity and the moment inactivity or underutilisation is detected, the BS is switched off. The underutilised BS goes into wilting or a progressive switch off to prevent call interruptions and subsequently call drops. During the wilting phase, the few UEs within the cell in which a BS is switching off , are handed over to still active neighbouring BSs that have to increase their transmit powers slightly to cover those areas. Figure 3 shows an example of cell wilting.

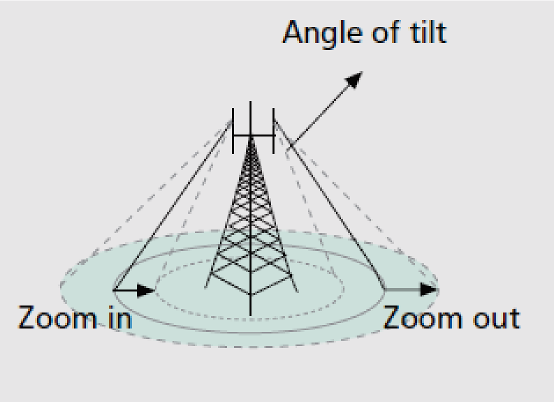


Figure 3: Cell wilting [1]

With this proposed solution, if a UE encounters some unacceptable degradation with no possibility of being handed over to a willing neighbour, it alerts the about to switch off BS and the switching off process is halted. On switching on a sleeping BS, the process is such that the BS power is raised slowly through a progressive switch on process the authors called BS blossoming. This process prevents denial of service and call drops.

This ensures that an acceptable QoS is maintained at all times. From the case study that the authors carried out on a portion of Munich in Germany, they came to the conclusion that wake up transients associated with the system are very short and thus are acceptable in a network where wilting and sleep modes technologies are used.

The biggest benefit scheme has is its mechanism to handle BS sleep and switch on transients. A strong signal from a BS that is switched on too fast can cause a UE to lose connection that might be receiving a weak signal from a distant BS because of the generated strong interference. The scheme handles this process very well. On the negative side, the proposed solution does not adequately address the problem of inter-cell interference and coverage holes. As BSs gradually increase their spheres of influences, their areas of overlap increase and hence mutual interference also increase. It is not mentioned in the report how this problem is dealt with.

In [21], authors proposed a distributed BS switching algorithm that they called SWES where BSs and UEs periodically share amongst themselves information such as signal strength and system load. There is no central controller required. The BS that has the least network impact is switched off and users served by that BS are handed over to the second best BS. Network impact helps to quantify how the switching off process affects the system load. Mathematically, it is given by:

$$F_b = \max_{n \in N_b}(\rho_n + \rho_{b \to n}), \forall b \in B^{on}. \qquad (1)$$

Where F_b is network impact,
ρ_n is the internal system load of BS n,
N_b are neighbouring BSs,
$\rho_{b \to n}$ is the external load from b to n.

Using the network impact, each BS decides whether it has to switch off or to remain on. Before a BS switches off, it sends a request to switch off to neighbouring BSs and then switches off after being cleared to do so by the neighbours. This is done to ensure that the neighbouring BSs do not end up being overloaded by the extra traffic load coming from the switching off BS. For the BS to switch back on, neighbouring BSs play a leading role in the process by ensuring that the sleeping BS is switched on when the traffic load reaches the same value that it was originally switched off on. This algorithm the authors are proposing here is the same one that they proposed in [22]. With their algorithm, the authors claimed that 80% energy savings can be achieved which is quite substantial. The fact that the algorithm does not require a central controller means it does not suffer from problems associated with failure of a central controller that tends to bring down the whole network in the event that the controller fails.

However, since the switching ON and OFF is not centrally controlled, there is a strong possibility of having coverage holes in the cellular network using this scheme. Also, in the event that the Request to Switch off (RTSO) and the Clear to Switch off (CTSO) signals are exchanged simultaneously, the proposed algorithm might operate ineffectively leading to a compromised QoS.

Researchers in [23] developed an algorithm that can be used to switch ON/OFF BSs in densely deployed networks where each BS would be serving a reasonably small number of UEs. A decision to switch off a BS is reached after considering the traffic load of neighbouring BSs. The proposed heuristic algorithm involves ranking the BSs according to their traffic load and BSs whose traffic load is below a certain prescribed threshold; taking into consideration a certain call blocking probability constraint, are switched off. The proposed algorithm also considers a minimum holding time during which time a BS cannot change its state. This prevents frequent mode switching so as to save equipment. A BS can only be switched off if all its current UEs can be handed over to neighbouring active BSs otherwise the switch off process is halted. The proposed scheme, when simulated, was able to maintain a target of 1% blocking probability almost all the time. The authors however could not guarantee the target blocking probability if system is monitored over an extended period of time. Their results show that the proposed algorithm leaves the network with a number of active BSs that matches the varying network traffic. This ensures that the energy the network consumes at any given point in time is the energy that is required at that time. There is no wastage and that is Green Cellular Communications at its best. The scheme has an element of holding time during which time the BS cannot change state. This is good for the equipment since there is no frequent mode switching. The proposed scheme is not so complex hence implementation is easy.

The holding time can also work against the system in that BSs might fail to switch ON or OFF at the time they are supposed to do so in order to observe the holding time. This can result in coverage holes which are one of the biggest problems in schemes that in involve temporarily switching OFF BSs.

Another BS sleeping algorithm was advanced by researchers in [24]. The algorithm makes use of two thresholds, the lower and upper thresholds. The difference between these two limits should be large enough to prevent the so called ping pong effect. Under this scheme, BSs in a cluster of seven BSs share information pertaining to their traffic loads. BSs with an utilisation that is lower than the prescribed lower threshold broadcast their utilisation values. From these broadcast messages, a BS that has the lowest utilisation then broadcasts intention to switch off. If neighbouring BSs after checking their own utilisation, agree to take over the extra load from this BS, they give thumbs up signal for the handover and the switching off procedure to commence. This long process ensures that no two BSs can switch OFF at the same time within the same cluster otherwise some handovers might be affected in the process. When the BS finally switches OFF, then the remaining BSs memorise the status of this sleeping BS. The authors simulated the proposed solution on a homogeneous network of fifteen macro BSs and the results obtained showed a marked reduction in energy consumption in a cellular network. Again, due to the presence of a long holding time, the equipment is saved from frequent mode changes and this increases the equipment's reliability. The system is decentralised as switching ON/OFF decision can be taken for a cluster of seven cells. This makes manageability easier because of the modularity that the system uses. The system also posted substantial energy savings from the simulations made.

On the negative side, the system was simulated on a homogenous network and considering that future networks are most of them going to be heterogeneous networks, its performance on such networks remains unknown.

Authors in [25] proposed a centralised algorithm where BSs share their ON/OFF status with a central controller unit. This information sent to the controller is sorted out using the following sorting rules:

- Least-Load (LL) - In this sorting, BSs are arranged starting with a BS with the smallest number of associated UEs to the one that has the highest number of UEs.
- Most-Overlapped (OV)-This strategy takes into consideration overlapping coverage areas among neighbouring BSs.

The algorithm then checks whether there is a candidate BS that can be put out to sleep. A candidate BS, if found, is removed from the current BS topology. If this residual topology can still fulfil coverage and capacity requirements, then the candidate BS is switched OFF after its few users would have been handed over to neighbouring BSs. From simulations conducted by the researchers, it was discovered that there were significant energy savings during the time when the network is not so busy. They discovered that the energy that is consumed by a network using their proposed solution is 50% of the energy that is consumed by a network where an Always-ON scheme is used during the time that traffic is very low. With LL being used, energy savings of between 17% and 28% can be achieved.

Since the system is centralised, it can effectively deal with the issue of coverage holes. By making use of the OV strategy, the proposed technology also caters for interference that is likely to creep into the system when BSs increase their transmit powers to cover areas where BSs would have switched OFF.

The solution is however silent on issues of BS-UE association which is very critical in BS sleeping technologies.

The researchers in [26] came up with a BS sleeping algorithm that makes use of an adaptive threshold. The authors combined BSs and Relay Stations (RS) in their proposed network. RSs are located at the edges of cells. The BSs and the RSs on the cell edges cannot be ON at the same time, which means that if the central BS is on, the surrounding RSs will be in dormant mode. The RSs may at certain instances amplify and forward, or decode and forward signals between BSs and UEs. The algorithm uses the prevailing network traffic to fix the switching ON/OFF pattern of the central BS dynamically. The decision to switch off is made at fixed intervals. The authors managed to evaluate the performance of their proposal and they compared it to other modes of BS operation by using extensive simulations. The adaptive threshold scheme managed a 53% energy saving. Figure 4 shows a comparison of the adaptive threshold method compared to the fixed threshold of 0.5.

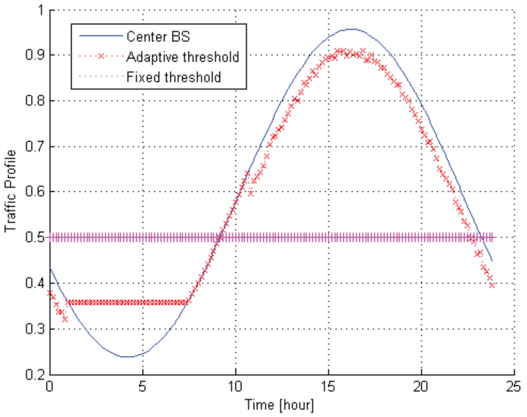


Figure 4: Normalised traffic profile of the centre BS with fixed and adaptive switching thresholds levels [26]

As can be seen from figure 4, high thresholds can be achieved with the adaptive threshold scheme. A higher switching threshold maintains the desired QoS and allows a longer interval for the BS to stay in either the RS or sleep mode thereby ensuring a higher energy saving. The energy savings are summarised in Table 2.

Table 2: Energy saving percentages in different operation modes with fixed and adaptive thresholds [26]

Switching algorithm	Always on (%)	BS sleeping (%)	BS-RS switching (%)
Fixed-threshold	0	20.4	39.4
Adaptive - threshold	0	32.8	53.4

The adaptive threshold method results in substantial energy savings as can be noticed on table 2.

The adaptive threshold method, though being effective for energy savings, is a rather complex mechanism which makes implementation on an extensive scale very difficult.

In [27] researchers came up with a Dynamic Traffic Aware (DTA) and Dynamic traffic and Interference Aware (DTIA) algorithms that can be used together with BS sleeping schemes to decide on the best BS to switch OFF for maximum energy saving. DTA switches off BSs if traffic has gone below a prescribed threshold. Before a BS is switched OFF, it has to be ascertained that the neighbouring BSs can handle the extra traffic coming from the BS that has to be switched OFF. The set of BSs that are put to sleep however might not be the best since the DTA only uses the fixed threshold to switch OFF the BSs. To always have a situation where the best set of less active BSs being switched OFF, the authors introduced DTIA which incorporates interference into the switching off process. With this scheme, each active BS collects information pertaining to the prevailing traffic and also interference levels from UEs in their neighbouring cells. This information is then used to prepare a utility function for each BS that takes into consideration the impact of both interference and traffic. The authors then went on to simulate 5 by 5 hexagon cells having an inter-side of 1.732km and they chose 0.6 as the traffic threshold and produced the results in figure 5.

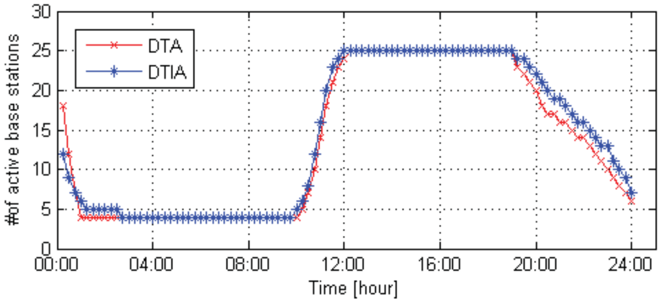


Figure 5: Comparing DTA and DTIA over a 24 hour period [27]

It can be seen from figure 5 that DTIA keeps more BSs in the sleep mode in the morning while it switches off less in the evening.

The energy savings are quite significant. The system also uses interference in deciding when to switch off BSs and this assists in maintaining the QoS at the desired levels.

The process of BS-UE association is also thoroughly dealt with and this is good for QoS. The strategy introduced in [28] makes use of Holt-Winter forecast method to decide at what point to switch off BSs. The technique is a recursive scheme that updates at each observation of the phenomena in question. This technique is an extension of exponential smoothing that is used to forecast future values using past values. This prediction is then used to decide the points at which certain BSs have to be switched OFF.

If it so happens that the number of channels needed at a certain time is far less than what the network can offer, then the BS with the lowest load is switched OFF first. The BS is switched OFF after verifying that all communications currently using the BS are successfully handed over to neighbouring BSs. The authors evaluated their proposed scheme by use of software simulations and concluded that it produced greater energy savings at night.

Where the prevailing traffic conditions differ slightly with the norm, the proposed solution is likely to introduce some degradation in the QoS since the system makes use of the Holt-Winter forecast method to decide on the threshold.

3. Contributions of this paper

In this paper several methods that have been specifically been advanced by researchers that deal with the reduction of energy consumption by the BS have been investigated and analysed. This makes it easier for researchers to find information of BS sleeping technologies in one place and they can then use the information to come up with the best scheme that can be used.

4. Conclusion

In this paper, the latest technologies that researchers have proposed with regards to BS sleeping technologies have been investigated. Most of these technologies result in a significant amount of energy saving. However, there have not been many proposals specifically for heterogeneous networks, which are networks of the future. More still needs to be done in this regard.

5 References

[1] A. Conte, A. Feki, L. Chiaraviglo, D. Ciullo, M. Meo & M.A. Marsan, "Cell wilting and blossoming for energy efficiency," Wireless Communications, IEEE, vol. 18, no. 5, pp. 50-57, October 2011.

[2] CISCO (2014, Feb.) Cisco visual networking index: Global mobile data traffic forecast Update, 2013-2018. [Online]. Available: www.cisco.com/c/en/us//solutions/collateral/service-provider/visual-networking-index-vni/whit_paper_c11-520862.pdf

[3] J. Wu, S. Rangan & H. Zhang, "Internet of Things – Converging Technologies for smart environments," New York: CRC Press, 2013.

[4] J. Wu, S. Rangan & H. Zhang, "Green communications: Theoretical Fundamentals, algorithms and applications," New York: CRC Press, 2013.

[5] Energy Balances and CO_2 Implications. [Online]. Available: http://www.world-nuclear.org/info/Energy-and-Environment/Energy-Balances-CO2-Implications/

[6] Y. Fan, X. Wang & C.H.J Peter, "Green Cellular….Towards Sustainable Networks," [Online].Available: www.mobile.ecei.tohoku.ac.jp/COE/seminar_2010_06_2/Green.pdf

[7] M.H Asharif, R. Nordin & M. Ishmail, "A survey of green radio communication Networks: Techniques and recent advances," Journal of computer networks and communications, vol. 2013, 2013.

[8] E. Hossain, V.K Bhargava, G.P Fettweis (Ed), "Green Radio Communication Networks," Cambridge: Cambridge University Press, 2012.

[9] EARTH. [Online]. Available: https://www.ict-earth.eu/

[10] D. W. Runton, M. D. LeFevre & M. K. Mellor, "Doherty power amplifier design" [Online], Available: http://www.rfmd.com/sites/default/files/resources/migration/presentations/commDRuntonPASymposium11.pdf

[11] S. A. Rahane, "A survey of Green Wireless Communications," International Journal of Electronics, Communications and Instrumentation Engineering Research and Development, vol. 3, issue 2, pp. 25-35, June 2013.

[12] S. Taruna, B. Pahwa & I. Kaur, "Energy efficient cellular networks: A survey," Advance in Electronic and Electric Engineering, ISSN 2231-1297, vol. 3, number 1, pp. 127-136, 2013.

[13] C. R. Muthy & C. Kavitha, "A survey of green base stations in cellular networks," International Journal of Computer Networks and Wireless Communications, ISSN 2250-3501, vol. 2 no. 2, pp. 232-236, April 2012.

[14] Z. Hasan, H. Bootstanimehr & V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," Communications Surveys & Tutorials, IEEE, vol. 13, no. 4 pp. 524-540, Fourth Quarter 2011.

[15] M. H. Alsharif, R. Nordin & M. Ismail, "A review on intelligent base stations cooperation management techniques for greener LTE cellular networks," Journal of Communications, vol. 9, no. 9, pp. 937-945, December 2014.

[16] A. D. Domenico, E. C. Strinati & A. Capone, "Enabling green cellular networks: A survey and outlook," Computer Communications, vol. 37, pp. 5-24, January 2014.

[17] D. Feng, C. Jiang, G. Lin, L. J. Cimini, G. Feng & G. Y. Li, " A survey of energy efficient wireless communications," Coms Surveys & Tutorials, IEEE, vol. 15, no. 1, pp. 167-178, First Quarter 2013.

[18] M. Ismail, W. Zhuang, K. Qaraqe & E. Serpedin, "A survey on green mobile networking: From the perspectives of network operators and mobile users," Communication surveys & Tutorials, IEEE, vol. 1, no. 99, pp. 1-23, 2014, doi: 10.1109/COMST.2014.2367592.

[19] R. Vijayasrathi, A. S. Monika, N. R. Himaja & S. Saraswathy, "A survey report on cell zooming for an energy efficient cellular network," International Journal of Engineering Research & Technology, vol. 3, issue 12, pp. 655-660, December 2014.

[20] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso & C. Verikoukis, "Energy efficient schemes for base station management in4G broadband systems,"[Online], Available: https://www.researchgate.net/publication/257921530_Energy_Efficient_Schemes_for_Base_Station_Management-_in_4G_Systems

[21] E. Oh, K. Son & B. Krishnamachari, "Dynamic base station switching –on/off strategies for green cellular networks," IEEE Transactions on Wireless Communications, vol. 12, no. 5, pp. 2126-2136, May 2013.

[22] E. Oh & B. Krishnamachari, "Energy savings through dynamic base station switching in cellular wires access networks," Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE, vol. 2 no. 1, pp. 1-5, December 2010.

[23] J. Gong, S. Zhou, Z. Niu & P. Yang, "Trffic-aware base station sleeping in dense cellular networks," Quality of Service (IWQoS), 2010 18th International Workshop on, vol. 3, no. 3 pp. 1-2, June 2010.

[24] G. Lee, H. Kim, Y. T. Kim & B. H. Kim, "Delauny triangulation based green base station for self-organising networks," 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things, Cyber Physical and Social Computing.

[25] L. Chiaraviglio, D. Ciullo, G. Koutitas, M. Meo & L. Tassiulas, "Energy–Efficient Planning and Management of Cellular Networks," Wireless ON-demand Network Systems and Services (WONS), 2012 9th Annual Conference on, vol. 1, no. 1, pp. 159-166, 9-11 January 2012.

[26] A. S. Alam, L. S. Dooley, A. S Poulton & Y. Ji, "Energy savings using an adaptive base station relay station switching paradigm," International Conference on Wireless Communications and Signal Processing (WCSP'12), 25-27 October 2012, Huangshan, China.

[27] A. S. Alam, L. S. Dooley & A. S Poulton, "Traffic-and-aware base station switching for green cellular networks," 18th IEEE International Workshop on Computer-Aided Modeling Analysis and Design of Communication links and Networks (CAMAD 2013), 25-27 September 2013.

[28] S. Morosi, P. Piunti & E. D. Re, "Sleep mode management in cellular networks: A traffic based technique enabling energy saving," Transactions on Emerging Telecommunications Technologies, 2013, doi: 10.1002/ett.2621.

Improving Energy Efficiency for Carrier Aggregation in Cellular Networks via Desirability-Based Resource Allocation

Pei-Rong Li, Szu-Chen Yeh[†], and Kai-Ten Feng
Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan
[†] Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan
shockbowwow.cm01g@nctu.edu.tw, [†]girl8222@gmail.com, and ktfeng@mail.nctu.edu.tw

Abstract— *Hyper-dense deployment of base stations (BSs) with carrier aggregation (CA) is being considered as a promising technique to improve the capacity and coverage for Long Term Evolution-Advanced (LTE-A) user equipments (UEs). However, the increasing energy consumption used to support this type of structure might lead to a significant impact on operating expenditure of their service providers. To efficiently leverage energy for network throughput enhancement, a weighted energy efficiency (EE)-based resource allocation (WEERA) algorithm is proposed to improve network EE under the quality-of-service (QoS) requirement. The proposed algorithm is a desirability-based heuristic method together with convex optimization over the subset of decision variables, which can efficiently perform BS association, component carrier (CC) assignment, resource block (RB) scheduling, and power allocation. Simulation results show that our method enjoys higher EE in CA-based cellular networks.*

1. Introduction

Carrier aggregation (CA) is an important feature of cellular networks, which is supported in 3rd Generation Partnership Project (3GPP) Release 10 to allow cellular networks to make the best use of spectrum. The component carrier (CC) can have a bandwidth of 1.4, 3, 5, 10, 15 or 20 MHz and a maximum of five CCs can be aggregated, thus up to 100 MHz of spectral chunks can be aggregated for data transmission. Besides, cellular networks using small cells drawn much attention recently due to their potential gain to bring network closer to user. Both small cells and CA technique are viewed as a key revolution to enhance the capacity and coverage in next generation wireless networks.

While the conventional approach of providing more resources for Long Term Evolution-Advanced (LTE-A) user equipments (UEs) can result in more diversity gain for scheduling, the complexity in processing such as energy consumption and interference management might lead to the degradation of energy efficiency (EE). Therefore, the energy saving is investigated as well as data capacity enhancement in this paper. Considering the heterogeneous network (HetNet) structure with multiple low power base stations (BSs), a joint BS association and radio resource allocation method is discussed in [1] to choose suitable serving BSs and maximize network throughput. Since the energy conservation is quite as important as spectrum efficiency, the literatures [2] and [3] paid attention on EE optimization problem. The above studies tend to translate a combinatorial optimization problem with convex relaxation, but the globe optimal solution cannot be found in polynomial time since the computational complexity is still high. Therefore, many heuristic approaches with sub-optimal solutions have been proposed to schedule resources [4][5].

In order to efficiently achieve the potential gain from hyper-dense deployment with CA technology, the design of resource allocation optimization should be considered. The EE maximization problem in this paper is addressed as a joint problem of BS association, CC assignment, resource block (RB) scheduling and power allocation. To realize the EE enhancement under the consideration of computational complexity, a joint operation of heuristic algorithm and convex optimization method is proposed, called weighted EE-based resource allocation (WEERA). The desirability-based heuristic algorithms are employed to make the decisions of BS association and CC assignment. Desirability is proposed as a performance metric in [6] to realize interference reduction and more efficient resource utilization. Furthermore, a convex relaxation is introduced to solve the remaining subset of variables, namely, the policies of RB scheduling and power allocation. Since the combinatorial problem is transformed into a concave one, the dual decomposition approch can be utilized to find the decision strategies for EE maximization. The overall description of our work is illustrated in the following sections.

2. System Model and Problem Formulation

Consider a cellular network with M evolved Node Bs (eNBs) aim to serve K UEs. Both eNBs and UEs are equipped with a single antenna and share the frequency resources for data transmission. The bandwidth in LTE-A system consists of J CCs. The CCs utilized in LTE-A standard can be categorized into primary CC (PCC) and secondary CC (SCC), the former is picked to handle control information as well as quality-of-service (QoS) support, and the latter is used as supplementary spectrum resource for data transmission. Each

[1] This work was in part funded by the Aiming for the Top University and Elite Research Center Development Plan, NSC 102-2221-E-009-018-MY3, the MediaTek research center at National Chiao Tung University, the Industrial Technology Research Institute (ITRI), and the Telecommunication Laboratories at Chunghwa Telecom Co. Ltd, Taiwan.

CC j is divided into R_j RBs with equal bandwidth W. In this section, an overview on the transmission model and the design of EE optimization problem for downlink cellular network with CA are described as follow.

It is assumed that open access is supported for all eNBs, but each UE belongs to exactly one cell. UE can make connection with eNB and handle the network entry process in the control channel through PCC. As LTE-A system can leverage CA for multi-CC transmission, the eNB first configures a suitable CC set for each associated UE. The CC configuration involves two-step procedures to select one PCC and supplemental SCCs. Then RB assignment is introduced to avoid mutual interference among UEs within the same cell, whereas there exists inter-cell interference. To further enhance the network performance with the consideration of QoS requirements, power allocation is adopted for each transmission link.

The main goal of our discussion is to maximize network EE, denoted as ψ, under the data rate requirements and resource limitations. This can be achieved by optimally solving the decision policies of BS association, CC assignment, RB scheduling, and power allocation. The corresponding variables of this problem are listed in Table 1. For simplicity, it is desirable to symbolize these solution sets as $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\rho}$, and $\mathbf{P}$. Note that $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\rho}$ are binary indicators for resource allocation as either assigned (equal to one) or not (equal to zero).

Table 1
DECISION POLICIES

Variable	Definition
$\alpha^{m,k}$	Tendency to associate UE k with eNB m
$\beta_j^{m,k}[1]$	Tendency to configure CC j as PCC to UE k served by eNB m
$\beta_j^{m,k}[2]$	Tendency to configure CC j as SCC to UE k served by eNB m
$\rho_{j,r}^{m,k}$	Tendency to assign RB r within CC j to UE k served by eNB m
$p_{j,r}^{m,k}$	Transmit power for eNB m to UE k on RB r within CC j

Per aforementioned definitions, the QoS-aware EE maximization problem can be modeled mathematically as

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\rho},\mathbf{P}} \quad \psi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\rho},\mathbf{P}) \tag{1a}$$

$$\text{s.t.} \quad \sum_{m=1}^{M} \alpha^{m,k} \leq 1, \qquad \forall k, \tag{1b}$$

$$\sum_{j=1}^{J} \beta_j^{m,k}[1] \leq 1, \qquad \forall m, \forall k, \tag{1c}$$

$$\sum_{l=1}^{2} \beta_j^{m,k}[l] \leq 1, \qquad \forall m, \forall k, \forall j, \tag{1d}$$

$$\sum_{j=1}^{J}\sum_{l=1}^{L} \beta_j^{m,k}[l] \leq J_{(\max)}, \qquad \forall m, \forall k, \tag{1e}$$

$$\sum_{k=1}^{K} \rho_{j,r}^{m,k} \leq 1, \qquad \forall m, \forall j, \forall r, \tag{1f}$$

$$P_{(\mathrm{T})}^{m}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\rho},\mathbf{P}) \leq P_{(\max)}^{m}, \qquad \forall m, \tag{1g}$$

$$C_{(\mathrm{P})}^{k}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\rho},\mathbf{P}) \geq C_{(\mathrm{th})}^{k}, \qquad \forall k, \tag{1h}$$

where (1b) and (1c) restrict that each UE is only allowed to associate with one eNB and can be configured at most one PCC respectively. Constraint (1d) excludes the CC repeatability and (1e) means that each UE can be assigned at most $J_{(\max)}$ CCs, which is defined in standard [7]. (1f) implicitly imposes a fairness constraint since no UE can dominate the RB reuse process within the same cell. The maximum allowed transmit power constraint for each eNB, i.e., $P_{(\max)}^{m}$, is limited in (1g), where the transmit power allocated to eNB m is written as

$$P_{(\mathrm{T})}^{m} = \sum_{k=1}^{K} \alpha^{m,k} \sum_{j=1}^{J}\sum_{l=1}^{2} \beta_j^{m,k}[l] \sum_{r=1}^{R_j} \rho_{j,r}^{m,k} p_{j,r}^{m,k}. \tag{2}$$

The condition in (1h) specifies that each UE is required to satisfy its target data rate $C_{(\mathrm{th})}^{k}$ on PCC according to the QoS requirement. $C_{(\mathrm{P})}^{k}$ in (1h) is the achievable data rate for UE k on PCC, which is formulated as

$$C_{(\mathrm{P})}^{k} = \sum_{m=1}^{M} \alpha^{m,k} \sum_{j=1}^{J} \beta_j^{m,k}[1] \sum_{r=1}^{R_j} \rho_{j,r}^{m,k} c_{j,r}^{m,k}, \tag{3}$$

where $c_{j,r}^{m,k}$ is the capacity for each transmission link. Denoted $I_{j,r}^{m,k}$ as the interference experienced by UE k in a specific channel, $c_{j,r}^{m,k}$ can be expressed as

$$c_{j,r}^{m,k} = W\log_2\left(1 + \frac{p_{j,r}^{m,k} g_{j,r}^{m,k}}{I_{j,r}^{m,k} + WN_0}\right). \tag{4}$$

For downlink transmission, the network EE is conventionally defined as the ratio of average sum rate to the total transmit power consumption. Therefore, the objective function in (1a) is given by

$$\psi = \frac{C_{(\mathrm{total})}}{\sum_{m=1}^{M} P_{(\mathrm{T})}^{m}}, \tag{5}$$

where

$$C_{(\mathrm{total})} = \sum_{k=1}^{K}\sum_{m=1}^{M} \alpha^{m,k} \sum_{j=1}^{J}\sum_{l=1}^{2} \beta_j^{m,k}[l] \sum_{r=1}^{R_j} \rho_{j,r}^{m,k} c_{j,r}^{m,k}. \tag{6}$$

The optimization problem (1) is treated as a mixed integer programming problem that is in general NP-hard for the optimal solutions. For reducing the computational cost and realizing comparatively efficient resource allocation, a weighted EE-based resource allocation method is proposed in next section.

3. Weighted Energy Efficiency-Based Resource Allocation

In this section, a polynomial-time algorithm is proposed for EE maximization under the QoS constraint. We illustrate this algorithm with two main concepts, called desirability-based decision strategy and convex optimization over subset of variables. The former is employed to efficiently utilize the resources on both BSs and CCs, and the latter is introduced to meet the constrained optimization by dynamic allocating RB and power. The decision strategy which results in better EE can be determined by the joint operation of the above-mentioned ideas.

3.1 Desirability-Based Base Station Association (DBSA)

In multi-cell scenarios, the received reference signal received power (RSRP) is viewed as a measurement of channel quality. While a UE performs cell selection or handover, the eNB with highest RSRP is conventionally expected to be associated. However, the interference from neighboring cells might result in serious performance degradation. As a result, the BS desirability is developed into

$$\bar{d}^{m,k} = \frac{\sum\limits_{j=1}^{J} \Upsilon_j^{m,k}}{\sum\limits_{m=1}^{M} \sum\limits_{j=1}^{J} \Upsilon_j^{m,k}}, \tag{7}$$

where $\Upsilon_j^{m,k}$ is the RSRP received by UE k from eNB m on CC j. The eNB m^* with the largest $\bar{d}^{m^*,k}$ will be selected by UE k, that is,

$$\alpha^{m^*,k} = 1 \text{ with } m^* = \arg\max_{m} \bar{d}^{m,k}, \quad \forall k. \tag{8}$$

From the perspective of a certain UE, the BS desirability provides additional information when RSRP is not sufficient to make a reliable BS association.

3.2 Desirability-Based Component Carrier Assignment (DCA)

After the procedure of BS association, eNBs tend to designate a suitable CC set for their corresponding UEs. The CC desirability for UE k served by eNB m is given similarly to (7), namely,

$$\tilde{d}_j^{m,k} = \frac{\Upsilon_j^{m,k}}{\sum\limits_{k=1}^{K} \alpha^{m,k} \Upsilon_j^{m,k}}. \tag{9}$$

It can be observed that the developed CC desirability can prevent the load imbalance problem since it is preferred to configure different CCs to UEs within the same cell.

CC assignment involves two-step procedures to configure one PCC and supplemental SCCs. The PCC is selected for each UE with the highest $\tilde{d}_j^{m,k}$, which can be presented as

$$\beta_{j^*}^{m,k}[1] = 1 \text{ with } j^* = \arg\max_{j} \tilde{d}_j^{m,k}, \quad \forall m, \forall k. \tag{10}$$

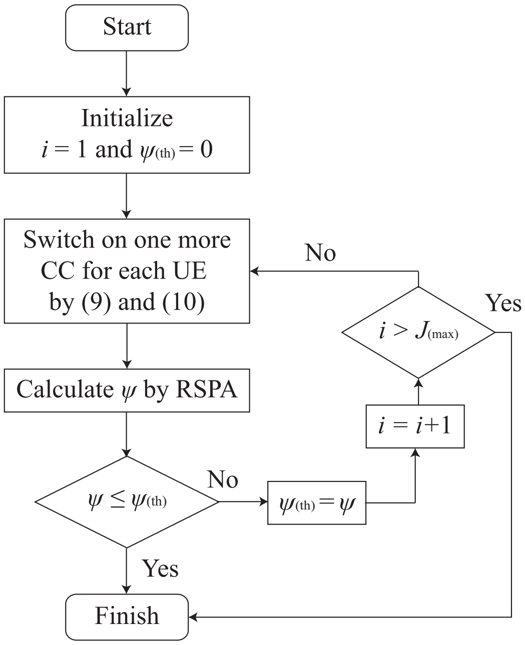

Start
Initialize
$i = 1$ and $\psi_{(\text{th})} = 0$
Switch on one more CC for each UE by (9) and (10)
Calculate ψ by RSPA
$\psi \leq \psi_{(\text{th})}$
No
$\psi_{(\text{th})} = \psi$
$i = i+1$
$i > J_{(\max)}$
No
Yes
Yes
Finish

Fig. 1
FLOW CHART OF DCA SCHEME.

Then the remaining CCs will sequentially turn on as SCCs as the same manner in (10). Each time that one more CC switched on, the RB scheduling and power allocation (RSPA) scheme is performed for EE maximization, which is introduced in next subsection. The remaining CCs will stop switching on as EE limits up or (1e) is unsatisfied.

The detailed procedure for DCA algorithm is described in Fig. 1, where i is the number of active CCs configured to each UE. Note that i of all UEs are set up with the same value in this paper. Dynamically modified i for each UE can get better performance, which is leaving for future work.

3.3 Resource Block Scheduling and Power Allocation (RSPA)

In this section, the non-convex combinatorial problem in (1) is transformed into a concave optimization by introducing fractional programing (FP) [8] and the relaxation of discrete variables. Then, the resulting RB and power allocation problem is solved by Lagrangian dual decomposition method [8].

FP is an iterative algorithm that can be used for solving a non-linear problem. The objective in (1a) expressed in a fractional form can be rewritten as a subtractive form by adopting FP. Without loss of generality, the following holds true at the optimal EE value of ψ^*

$$\max_{\boldsymbol{\rho}, \mathbf{P}} \quad C_{(\text{total})}(\boldsymbol{\rho}, \mathbf{P}) - \psi^* \sum_{m=1}^{M} P_{(\text{T})}^{m}(\boldsymbol{\rho}, \mathbf{P}) = 0. \tag{11}$$

Dinkelbach [9] had proposed an iterative method to find the feasible ψ values by solving the equivalent subtractive concave problems. Since $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be obtained by employing the algorithms of DBSA and DCA, the optimization problem in (1) can be reformulated as

$$\max_{\boldsymbol{\rho},\mathbf{P}} \quad C_{(\text{total})}(\boldsymbol{\rho},\mathbf{P}) - \psi \sum_{m=1}^{M} P_{(\text{T})}^{m}(\boldsymbol{\rho},\mathbf{P}) \tag{12a}$$

$$\text{s.t.} \quad 0 \le \rho_{j,r}^{m,k} \le 1, \quad \forall m, \forall k, \forall j, \forall r, \tag{12b}$$

$$\text{(1f)}, \text{(1g)}, \text{(1h)},$$

where (12b) is generated by relaxing the RB scheduling constraint into a continuous variable.

A dual decomposition method, called Lagrangian, together with a sub-gradient method [8] are then introduced to solve (12). The Lagrangian function is given by

$$\begin{aligned} L(\boldsymbol{\rho},\mathbf{P};\boldsymbol{\sigma},\boldsymbol{\eta},\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\theta}) &= C_{(\text{total})} - \psi \sum_{m=1}^{M} P_{(\text{T})}^{m} \\ &+ \sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{r=1}^{R_j} \sigma_{j,r}^{m,k}\rho_{j,r}^{m,k} \\ &- \sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{r=1}^{R_j} \eta_{j,r}^{m,k}(\rho_{j,r}^{m,k}-1) \\ &- \sum_{m=1}^{M}\sum_{j=1}^{J}\sum_{r=1}^{R_j} \lambda_{j,r}^{m}\left(\sum_{k=1}^{K}\rho_{j,r}^{m,k}-1\right) \\ &+ \sum_{k=1}^{K}\phi^{k}\left(C_{(\text{P})}^{k}-C_{(\text{th})}^{k}\right) \\ &- \sum_{m=1}^{M}\theta^{m}\left(P_{(\text{T})}^{m}-P_{(\text{max})}^{m}\right), \end{aligned} \tag{13}$$

where $\boldsymbol{\sigma}$, $\boldsymbol{\eta}$, $\boldsymbol{\lambda}$, $\boldsymbol{\phi}$, and $\boldsymbol{\theta}$ are the sets of Lagrangian multipliers associated with the resource limitations and the required minimum data rate constraints with non-negative elements $\sigma_{j,r}^{m,k}$, $\eta_{j,r}^{m,k}$, $\lambda_{j,r}^{m}$, ϕ^{k}, and θ^{m} respectively. For notational brevity, denoted $\boldsymbol{\Delta} = \{\boldsymbol{\sigma},\boldsymbol{\eta},\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\theta}\}$ and the actual transmit power on a certain RB as

$$\varepsilon_{j,r}^{m,k} = \rho_{j,r}^{m,k} p_{j,r}^{m,k}. \tag{14}$$

According to standard optimization techniques and the Karush-Kuhn-Tucker (KKT) conditions, the optimal transmit power and RB allocation, denoted by $\hat{\varepsilon}_{j,r}^{m,k}$ and $\hat{\rho}_{j,r}^{m,k}$, should satisfy

$$\frac{\partial L(\boldsymbol{\rho},\mathbf{P},\boldsymbol{\Delta})}{\partial \hat{\varepsilon}_{j,r}^{m,k}} \begin{cases} = 0, \hat{\varepsilon}_{j,r}^{m,k} > 0, \\ < 0, \hat{\varepsilon}_{j,r}^{m,k} = 0, \end{cases} \forall m, \forall k, \forall j, \forall r, \tag{15}$$

$$\frac{\partial L(\boldsymbol{\rho},\mathbf{P},\boldsymbol{\Delta})}{\partial \hat{\rho}_{j,r}^{m,k}} \begin{cases} < 0, \hat{\rho}_{j,r}^{m,k} = 0, \\ = 0, 0 < \hat{\rho}_{j,r}^{m,k} < 1, \forall m, \forall k, \forall j, \forall r, \\ > 0, \hat{\rho}_{j,r}^{m,k} = 1, \end{cases} \tag{16}$$

which are first-order sufficient and necessary conditions for optimality. The optimal allocated power $\hat{p}_{j,r}^{m,k}$ for all m, k, j, and r can be readily obtained as

$$\begin{aligned} \hat{p}_{j,r}^{m,k} &= \frac{\hat{\varepsilon}_{j,r}^{m,k}}{\rho_{j,r}^{m,k}} \\ &= \left[\frac{W\left(\sum\limits_{l=1}^{2}\beta_{j}^{m,k}[l] + \phi^{k}\beta_{j}^{m,k}[1]\right)}{\ln 2\sum\limits_{l=1}^{2}\beta_{j}^{m,k}[l]\left(\psi+\theta^{m}\right)} - \frac{I_{j,r}^{m,k}+WN_0}{g_{j,r}^{m,k}}\right]^{+}. \end{aligned} \tag{17}$$

It can be observed that (17) follows as a multi-level water-filling.

From (16), the following first-order derivation with respect to $\hat{\rho}_{j,r}^{m,k}$ can be obtained by

$$\frac{\partial L(\boldsymbol{\rho},\mathbf{P},\boldsymbol{\Delta})}{\partial \hat{\rho}_{j,r}^{m,k}} = R_{j,r}^{m,k} + \sigma_{j,r}^{m,k} - \eta_{j,r}^{m,k} - \lambda_{j,r}^{m}, \tag{18}$$

where

$$\begin{aligned} R_{j,r}^{m,k} =& \frac{W}{\ln 2}\left(\alpha^{m,k}\sum_{l=1}^{2}\beta_{j}^{m,k}[l] + \phi^{k}\alpha^{m,k}\beta_{j}^{m,k}[1]\right) \\ & \cdot \ln\left(1+\frac{\hat{p}_{j,r}^{m,k}g_{j,r}^{m,k}}{I_{j,r}^{m,k}+WN_0}\right) \\ & -\left(\psi+\theta^{m}\right)\alpha^{m,k}\sum_{l=1}^{2}\beta_{j}^{m,k}[l]\hat{p}_{j,r}^{m,k}. \end{aligned} \tag{19}$$

Each RB r within CC j is allocated to a specific UE k having the highest value of $R_{j,r}^{m,k}$ in each cell m for the purpose of achieving the highest increase in $L(\boldsymbol{\rho},\mathbf{P},\boldsymbol{\Delta})$. The optimal scheduling for RBs is decided as

$$\hat{\rho}_{j,r}^{m,k^*} = 1|_{k^*=\arg\max\limits_{k} R_{j,r}^{m,k}}, \quad \forall m, \forall j, \forall r. \tag{20}$$

Furthermore, a sub-gradient method that exploits iterative approach as in [10] is utilized to update the Lagrangian multipliers, which is given by

$$\sigma_{j,r}^{m,k}(n+1) = \left[\sigma_{j,r}^{m,k}(n) - s(n)\rho_{j,r}^{m,k}\right]^{+}, \tag{21a}$$

$$\eta_{j,r}^{m,k}(n+1) = \left[\eta_{j,r}^{m,k}(n) + s(n)\left(\rho_{j,r}^{m,k}-1\right)\right]^{+}, \tag{21b}$$

$$\lambda^{m}(n+1) = \left[\lambda_{j,r}^{m}(n) + s(n)\left(\sum_{k=1}^{K}\rho_{j,r}^{m,k}-1\right)\right]^{+}, \tag{21c}$$

$$\phi^{k}(n+1) = \left[\phi(n) - s(n)\left(C_{(\text{P})}^{k}-C_{(\text{th})}^{k}\right)\right]^{+}, \tag{21d}$$

$$\theta^{m}(n+1) = \left[\theta^{m}(n) + s(n)\left(P_{(\text{T})}^{m}-P_{(\text{max})}^{m}\right)\right]^{+}, \tag{21e}$$

where n is the iteration index. $s(n) = \frac{\kappa}{\sqrt{n}}$ is the step size and κ is a tunable constant. The detailed procedure for RSPA scheme is described in Algorithm 1.

Algorithm 1: RSPA Algorithm

```
Initialize iteration counter $\tau = 0$ and the maximum number of iterations $T_{(\max)}$
Initialize the maximum energy efficiency $\psi = 0$ and the maximum tolerance $\epsilon$
repeat
    Initialize Lagrangian multipliers $\mathbf{\Delta}$
    repeat
        Maximize: Calculate $\hat{p}_{j,r}^{m,k}$ and $\hat{\rho}_{j,r}^{m,k}$ by (17) and (20)
        Tighten : Update $\mathbf{\Delta}$ with $\hat{p}_{j,r}^{m,k}$ and $\hat{\rho}_{j,r}^{m,k}$ by (21)
    until Lagrange convergence
    Calculate the objective in (12a) with $\hat{p}_{j,r}^{m,k}$ and $\hat{\rho}_{j,r}^{m,k}$
    if Objective in (12a) $< \epsilon$ then
        FP converge = true
        Calculate optimal EE $\psi^*$ by (5) with $\hat{p}_{j,r}^{m,k}$ and $\hat{\rho}_{j,r}^{m,k}$
        return $\psi^*$
    else
        Reset $\psi$ by (5) with $\hat{p}_{j,r}^{m,k}$ and $\hat{\rho}_{j,r}^{m,k}$
        increment $\tau$
        FP convergence = false
    end if
until FP convergence = true or $\tau > T_{(\max)}$
```

4. Performance Evaluation

In this section, simulation results are provided. Consider the HetNet system consisted of five small cells underlying one macro cell with inter-site distance 500 m. The macro eNB can be viewed as three independent BSs since it is equipped with three-directional antennas. The maximum allowable transmit power $P_{(max)}^{m}$ for macro and small BS are 49 dBm and 24 dBm respectively. Moreover, the channel model used for simulation is referred to [11], and the available CCs which can be aggregated for downlink data transmission are referred to [7].

In Fig. 2, the EE of proposed WEERA and RSRP-based scheme versus the number of available CCs is plotted under $K = 10$ and $J_{(\max)} = 5$. The RSRP-based scheme is designed by replacing the desirability in our proposed scheme with RSRP. It can be seen that the proposed algorithm can provide better EE than RSRP-based scheme due to the consideration of interference. As the increase of J, the flexibility of CC assignment becomes higher and the co-channel interference might be lower. Therefore, the EE provided by our method will increase with J, while the EE in RSRP-based scheme will saturate since all UEs are extremely possible to choose the same CCs for data transmission. Also, the outage probability is analyzed in Fig. 3, which is defined as the chance of QoS unsatisfaction for UEs. It can be observed that the RSRP-based scheme is challenged by higher outage probability since the same CC might be configured as PCC for UEs as mentioned previously.

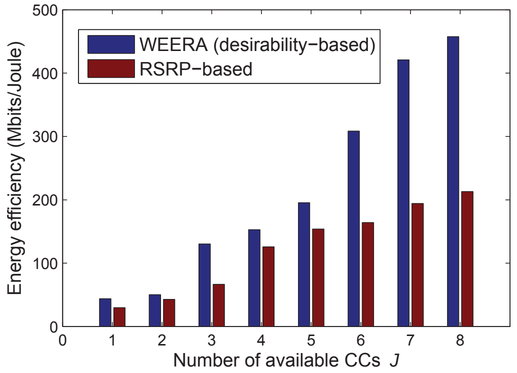


Fig. 2
EE COMPARISON OF PROPOSED WEERA SCHEME TO RSRP-BASED SCHEME UNDER $K = 10$, $J = \{1, 2, 3, 4, 5, 6, 7, 8\}$, AND $J_{(\max)} = 5$.

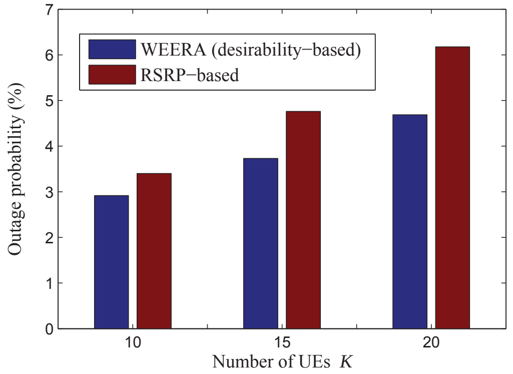


Fig. 3
OUTAGE PROBABILITY COMPARISON OF PROPOSED WEERA SCHEME TO RSRP-BASED SCHEME UNDER $K = \{10, 15, 20\}$, $J = 8$, AND $J_{(\max)} = 5$.

Fig. 4 depicts EE versus $J_{(\max)}$ under $K = \{10, 15, 20\}$ and $J = 8$. While the less CCs can be configured to each UE, the spectrum used for each UE will decrease as K increases. Therefore, the more energy would be consumed to guarantee data rate requirements, which leads to lower EE gain. However, the reversal performance will appear with higher $J_{(\max)}$ since the resources are sufficient enough to realize multi-user diversity gain. The merits of proposed scheme in CA-based cellular network can therefore be observed.

5. Conclusions

In this paper, the EE maximization problem has been investigated to improve network performance under the QoS constraints. The BS association and radio resource allocation problem are modeled as a mixed-integer programming problem, and then a joint operation of both desirability-based heuristic algorithm and convex optimization method is proposed to efficiently assign resources. Contrasted to the

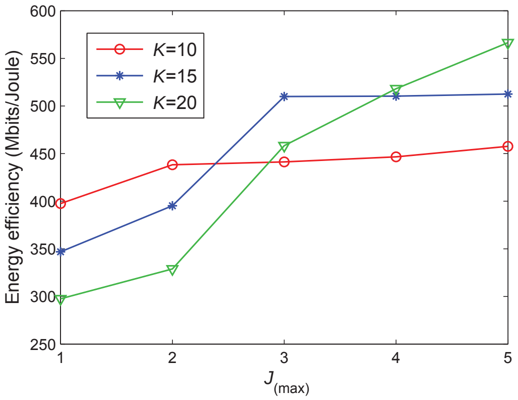


Fig. 4
EE COMPARISON OF DIFFERENT SYSTEM PARAMETERS UNDER $K = \{10, 15, 20\}$ AND $J = 8$.

conventional RSRP-based resource allocation algorithm, the proposed desirability-based scheme can achieve higher EE and finalize more efficient resource utilization. Furthermore, an analysis of outage probability have also provided. Simulation results show that the proposed WEERA is able to provide satisfactory performance in cellular networks with CA technology.

References

[1] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint Cell Selection and Radio Resource Allocation in MIMO Small Cell Networks via Successive Convex Approximation," in *IEEE ICASSP*, May 2014.

[2] S. Wang, C. Feng, C. Guo, and G. Wang, "Energy-Efficient Component Carrier Configuration and Power Control for Carrier Aggregated Systems," in *IEEE PIMRC*, Sept 2013.

[3] K. Sundaresan and S. Rangarajan, "Energy Efficient Carrier Aggregation Algorithms for Next Generation Cellular Networks," in *IEEE ICNP*, 2013, pp. 1–10.

[4] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct 1999.

[5] H. Ahmadi and Y. Chew, "Adaptive Subcarrier-and-Bit Allocation in Multiclass Multiuser OFDM Systems Using Genetic Algorithm," in *IEEE PIMRC*, Sept 2009.

[6] H. Tian, S. Gao, J. Zhu, and L. Chen, "Improved Component Carrier Selection Method for Non-Continuous Carrier Aggregation in LTE-Advanced Systems," in *IEEE VTC*, 2011, pp. 1–5.

[7] *Evolved Universal Terrestrial Radio Access (E-UTRA); Carrier Aggregation; Base Station (BS) Radio Transmission and Reception*, 3GPP TS 36.808, Jun. 2012.

[8] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[9] W. Dinkelbach, "On Nonlinear Fractional Programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.

[10] N. Z. Shor, K. C. Kiwiel, and A. Ruszcayski, *Minimization Methods for Non-Differentiable Functions*. New York: Springer-Verlag, 1985.

[11] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects*, 3GPP TR 36.814 V9.0.0, Mar. 2010.

Energy-efficient Communication Algorithm Using Luminance Control of Ceiling Lighting for Wireless Sensor Networks

Hiroki MURAKAMI[1], **Hiroto AIDA**[2], **Motoi OKADA**[1], **Kento MATSUI**[1] **and Mitsunori MIKI**[2]
[1]Graduate School of Science and Engineering, Doshisha University, Kyoto, Japan
[2]Department of Science and Engineering, Doshisha University, Kyoto, Japan

Abstract—*A wireless sensor network enables monitoring in a wide area by distributing many sensor nodes. The typical method of wireless communication in sensor networks is flooding. In flooding, each node broadcasts data. This method requires each node to transmit data many times and entails such problems as low energy efficiency and packet collisions. An alternative method of indoor communication is visible light communication; however, visible light communication has a low cost-effectiveness for a sensor network which handles small-size data, considering the requirement of devices such as a modulator. Hence, the authors propose a data communication method which does not use wireless communication but uses commonly available lightings and illuminance sensors. This study examines a communication scheme which transmits data to illuminance sensors by causing continuous variations in illuminance within a range not sensible by human eyes in an environment free of external light, and methods to increase the communication speed.*

Keywords: sensor networks, sensor nodes, lights

1. Introduction

In recent years, wireless sensor networks have drawn attention as a technology to control office or home appliances, control crop production facilities in agriculture, traffic monitoring or natural disaster monitoring [1]. A wireless sensor network enables wide area monitoring by placing many sensor nodes in the space which transmit sensing data. Each sensor node has such capabilities as sensing the temperature/humidity status of the target under monitoring, sending and receiving the sensing data and arithmetic processing. Meanwhile, being wireless means that wireless sensor nodes have limited power supply. Hence, to prolong the life and minimize the operating cost of a wireless sensor network, a high level of energy efficiency is needed in sensor nodes control. In sensor networks, flooding is the most typical method of wireless communication for transmitting the status information from sensors. While flooding is the simplest and strongest method of communication between nodes, all sensor nodes comprising the network are subject to communication loads because all sensor nodes broadcast status information. In a sensor network composed of sensor nodes with restrained power supply, flooding causes heavy communication loads. To solve these problems, there have been many research endeavors to develop more efficient methods of communication between sensor nodes based on flooding [2][3][4]. However, as long as using wireless communication, the possibilities of change in the communication environment from obstacles or node arrangement as well as packet losses, cannot be neglected. Hence, we propose a communication method not relying on wireless communication, to alleviate communication loads on a sensor network and increase the energy efficiency of sensor nodes. This study proposes a data communication method which does not rely on wireless communication, but utilizes the luminance control features of commonly available dimmable ceiling lights and luminance sensor nodes. The proposed method transmits data to the sensor nodes within the radiation of a lighting fixture all at once, eliminating the need of broadcasting and realizing high energy efficiency. We will also examine a communication algorithm without regard to effects of external factors other than the luminance of lighting fixtures such as daylight, and approaches to considering a higher communication speed.

2. Related work

2.1 Visible light communication

In recent years, the advancement of communication technologies has made wireless communication ubiquitously available. Meanwhile, for the users of wireless communication, there are no means to know from where data are sent or where they are going. To solve this problem in wireless communication using radio waves, research in visible light communication has been underway [5]. Visible light communication is a communication method which uses lighting for data communication. Visible light communication is considered to have the following three advantages: first, users can visually know that a transmitter is sending data. The sender can recognize by light that the data are being sent while the receiver can know the sender by recognizing the light source. Secondly, it enables secure communication. Since the communication is visible, just physically blocking light is enough to realize secure communication. Thirdly, visible light communication using the light from lighting fixtures can be used safely in areas where the use of radio waves is restricted, such as hospitals and airplanes. Conventional methods of visible light communication use visible light, which is an electromagnetic radiation of 0.4 -0.7 micrometers in wavelength to communicate information. By flashing an LED at a speed too high for human eyes to recognize, it communicates information to terminals equipped with a receiving device. The receiving device is typically a high-resolution photo diode or a high-speed image sensor. Hence, it is difficult to apply these methods to an existing network consisting of illuminance sensors. This study intends to realize a communication method applicable to existing common sensor networks by realizing data communication using illuminance sensors.

2.2 Flooding

Flooding is a method commonly used for transmitting status information obtained by sensors. In flooding, since packet communication to the whole sensor network is enabled by all sensor nodes broadcasting status information, all sensor nodes comprising the network are subject to communication loads. Meanwhile, the power consumption by packet communication shares a large part of the whole power consumption by sensor nodes. Hence, for a sensor network with restraints on power supply, energy efficient control of packet communication is essential. Against this backdrop, researches have been undertaken on ways to increase energy efficiency by improving the methods of route search in flooding [6]. However, power is nonetheless consumed as long as the communication uses a wireless technology. Therefore, we propose a communication method which does not use wireless communication to reduce communication loads on the sensor network to increase the energy efficiency of sensor nodes. In this study, the authors propose a data communication method which uses no wireless communication but uses the luminance control of commonly available dimmable ceiling lighting fixtures and illuminance sensor nodes. In data communication using luminance control, data transmission from the sink node to sensor nodes will be possible: for example, this may be used for sending commands to change the sampling frequency to each sensor node or to operate the power supply. When the luminance of a lighting fixture is altered, the illuminance measured by sensor nodes will change; then each sensor node refers to the luminance trend, and interprets and receives it as a transmitted data. This method, sending data to all sensor nodes within the area of radiation of a lighting fixture at once, eliminates broadcasting to increase energy efficiency.

3. Scheme of data communication using ceiling lighting luminance control

3.1 Overview

In the data communication scheme using ceiling lighting luminance control, the system alters the luminance of lighting fixtures so that the illuminance measured by sensor nodes will change, and sensor nodes calculate luminance changes, which realizes data communication. A possible problem from illuminance changes is that user comfort may be sacrificed if illuminances changes are sensed by users. Hence, the illuminance changes need to be within the range unnoticeable to users. Preceding studies have verified that illuminance changes within 7% from the current value cannot be sensed by human eyes [7]. Hence, assuming an indoor environment, the data communication scheme using ceiling lighting luminance control causes illuminance changes within 7% of the current illuminance to realize data communication. In order to propose a data communication algorithm using ceiling lighting luminance control, the authors investigated the trends of illuminance measured by sensor nodes at lighting luminance changes.

3.2 Illuminance trend experiment using wireless sensor nodes

An experiment was conducted to verify how the illuminance values measured by an illuminance sensor mounted on a wireless sensor node change when changes within about 7% of the current illuminance were given. For the experiment, Crossbow's MOTE MICAz was used as wireless sensor nodes [8]. On a MOTE MICAz node, a general-purpose external sensor board MDA088 was installed and a lead-type NaPiCa [9] illuminance sensor was mounted for obtaining illuminance values. The resistance between the MDA088 and the NaPiCa illuminance sensor here is 430 Ω. The experiment was conducted in the laboratory at Doshisha University, using 28 sets of a fullcolor LED lighting (DLA-016E, SHARP Corporation), a wireless sensor node with NaPiCa and a sink node. The illuminance measurement interval for the illuminance sensor was set to 0.1 second. Fig. 1 shows the plan of the experimental environment. Fig. 2 shows a photo of a scene from the experiment. The distance between a lighting and a wireless node placed perpendicularly below the lighting was 1.9m. As shown in Fig. 1, the wireless sensor node was placed right below the lighting.

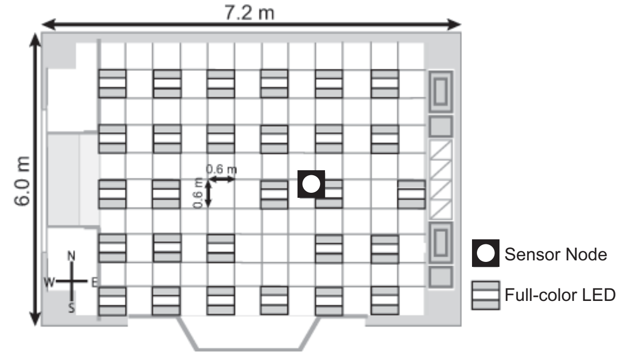


Figure 1: Environment for the illuminance measurement experiment

Figure 2: A scene from the illuminance measurement experiment

It should be noted here that the values obtained by a NaPiCa illuminance sensor do not directly show the illuminance values themselves. Besides, this experiment does not require accurate illuminance values because it is intended to enable the reception of data from a lighting by having a sensor node comparing illuminance values on relative terms. Hence, for the purpose of this experiment, the values obtained by a NaPiCa were corrected against a general-purpose class A, ANA-F11 illuminance meter; using an ANA-F11 illuminance meter capable of measuring illuminance with high precision. The illuminance values obtained from NaPiCa were corrected in this experiment. Equation 1 below is the correction formula.

$$Illana \approx 2.096 * Illnapica + 17 \qquad (1)$$

$Illana$: illuminance obtained by sensor ANA-F11 [lx]
$Illnapica$: value obtained by NaPiCa illuminace sensor

In an environment where the original desktop illuminance is 500 lx, the illuminance was raised by 15 lx, 3% of 500 lx, and then brought down to the original value in 1 second, while the taking illuminance data. The experiment was conducted in an environment in which the illuminance sensor is free from effects of external light such as daylight. Fig. 3 shows the history of illuminance measured at 0.1 second intervals.

As is indicated by Fig. 3, the illuminance changes sharply when a luminance change is detected. In addition, there constantly are small fluctuations in illuminance, which need to be recognized as no change in illuminance. Based on the experiment, a data communication algorithm using ceiling lighting luminance control is proposed.

3.3 Data communication algorithm using ceiling lighting luminance control

Since the illuminance measurement trend experiment using wireless sensor nodes demonstrated the existence of constant illuminance fluctuations, it became clear that there is a need of an algorithm which neglects these fluctuations but regards only large illuminance changes as significant illuminance changes. An algorithm of a data communication scheme using ceiling lighting luminance control satisfying this condition is described below. The proposed algorithm assumes an environment in which illuminance sensors are free from the effects of light other than from the lightings such as daylight.

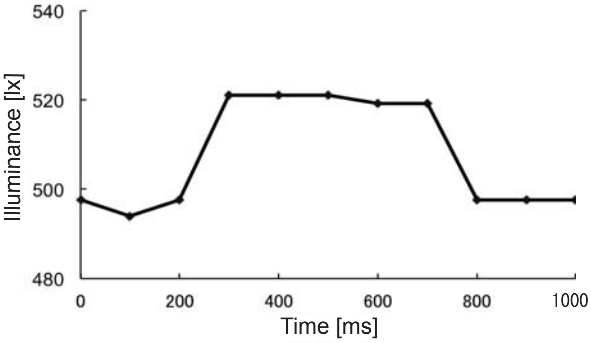


Figure 3: History of illuminance measured at 0.1 second intervals

From Fig. 3, when the sensor detects a luminance change, the illuminance value changes sharply, while some errors in illuminance measurements occur even while the lighting is kept on at a constant luminance, causing fluctuations in illuminance values. Hence, it needs to be ensured that sensors recognize only those illuminance changes occurring upon a luminance change as illuminance changes caused by data communication. For this purpose, in the proposed method, the measured illuminance value is differentiated from the previous illuminance value, and the algorithm judges based on the gradient whether it is an illuminance change caused by a luminance change or not.

The following paragraphs describe the algorithm more specifically. The proposed algorithm is premised on the condition that the data transmission interval is already known to sensor nodes. The data transmission interval here is T[s]. The illuminance sensor obtains illuminance values n times within T[s], and retains the current illuminance value and the previous illuminance value. Then the gradient of illuminance change is calculated using the current illuminance value and the previous illuminance value: if it is not below the threshold α, then it is deemed an illuminance change from a luminance change. Meanwhile, the lighting converts the data to transmit into a binary bit sequence and sends it bit by bit. For the purpose of this method, the bit is "0" if the gradient is below the threshold α or "1" if it is not below the threshold α. To indicate the start of communication to the sensor node, "1" is sent as a start bit. When bits of the same values are sent serially, the lighting is kept on at the same luminance, resulting in no change in illuminance. Hence, it determines whether 1 or 0 at the point of T seconds from the reception of a start bit, using the current illuminance value.

(1) The illuminance sensor obtains the current illuminance value
(2) The lighting raises the luminance by $x_1\%$ (when $x_1 < 7\%$) from the current luminance
(3) When the illuminance sensor detects a change in illuminance, the gradient between before and after the change is calculated.
(4) If the gradient is within the threshold, data communication is started.
(5) The lighting resets the luminance to the original luminance value.
(6) The gradient is calculated at every illuminance measurement interval. Then the system waits for T/2 seconds.
(7) Upon detecting a gradient, the system calculates the amount of change from the maximum illuminance and minimum illuminance in the last T seconds, and determines received bit against the preconfigured threshold.
(8) If the gradient remains below the threshold for T seconds after bit reception and the illuminance value is in the original range, it is deemed 0.

This algorithm enables data communication using changes in illuminance obtained by a sensor node as a communication medium, realized by changing the lighting luminance.

4. Increasing the speed of data communication using ceiling lighting luminance control

4.1 Outline of the strategy for speedup

The speed of communication using the data communication algorithm based on ceiling lighting luminance control depends on the bit transmission interval (T seconds) and the notation of the bits transmitted at T second intervals. Hence, this chapter examines approaches toward increasing the speed of data communication when a data communication algorithm based on ceiling lighting luminance control is used. For the purpose of this study, two approaches to speedup are verified: shortening the transmission interval and sending more information with one luminance control. In the method using a shorter transmission interval, the interval T was shortened from 1 second by 0.1 second in an attempt to realize a higher speed in binary data communication. In the method transmitting more data at a time, illuminance levels graded in multiple stages as shown in Fig. 4 were used to increase the amount of data transmitted by one step of luminance control to realize a higher speed.

4.2 Experiment of higher speed communication using shorter transmission intervals

A data communication experiment was conducted using the data communication algorithm based on ceiling lighting luminance control, varying the transmission interval (T second) between 1.0 second and 0.1 second by 0.1 second increments. The experimental environment was the same as that used in the illuminance trend experiment. The sample data used in the data communication are 100 bits of randomly generated binary data.

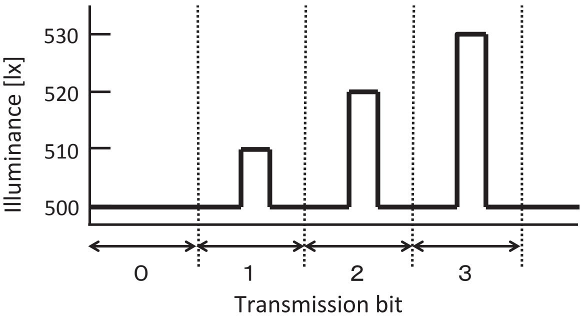


Figure 4: Speedup by graded illuminance changes

Fig. 5 shows the resulting average matching rates. From Fig. 5, with transmission intervals between 1.0 second and 0.2 second, the communication is successful with a matching rate of 99% or more. However, at an interval of 0.1 second, the matching rate is 52.5%, indicating that the communication was unsuccessful. Checking the illuminance measurement intervals in communication with a transmission interval of 0.1 second revealed that the interval between illuminance measurements exceeded 0.01 second, and judgment on one bit took 0.1 second or longer, which disabled successful communication. Because the average minimum interval between illuminance measurements for a NaPiCa sensor is 0.0084 seconds, it was considered that individual differences and application delays may have disabled illuminance measurement at a transmission interval of 0.1 second.

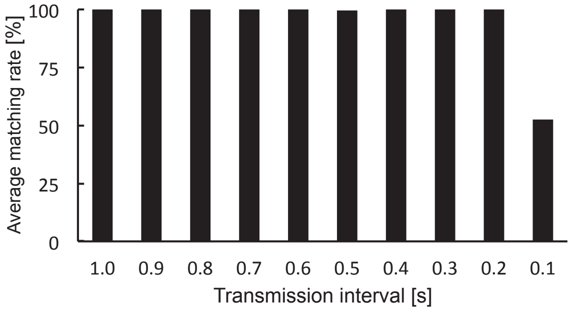


Figure 5: Average matching rates in data communication experiment at different transmission intervals

These results show that the communication speed can be increased by using shorter transmission intervals, of which limit depends on the minimum measurement interval of a NaPiCa illuminance sensor. Also where an illuminance sensor other than NaPiCa is used, the communication speed can be optimized by appropriately setting the transmission interval T. In the case of NaPiCa, since the average minimum measurement interval is 0.0084 seconds, it is considered that the limit of speedup may lie somewhere around 0.0084 seconds; but because a higher speed means a higher error rate, the transmission interval needs to be selected appropriately for the purpose of communication.

4.3 Experiment of higher speed communication using graded illuminance changes

In this experiment, by converting every 2 bits of the binary sequence into a quaternary equivalent to make a quaternary bit sequence, the luminance controlling side transmits more data with each luminance control, to realize communication at a higher speed. The luminance controlling side transmits bits using luminance levels graded in four stages as shown in Fig .4. The sensor node calculates the illuminance gradient after each measurement interval, and then identifies it with one of the four grades referring to three thresholds as shown in Fig.4. The gradient is judged to mean 0 if illuminance remains unchanged from the reference illuminance level L [lx] for T seconds; 1, 2 or 3 when it has changed from L [lx] to L+a, L+2a or L+3a respectively. Fig. 4 shows an example in which illuminance L is set at 500lx.

An experiment was conducted with a configuration to discriminate received bits as shown in Fig. 6. In the experiment, every 2 bits of the 100 bits of a binary bit sequence were converted into a quaternary equivalent to make a quaternary bit sequence, which were transmitted at 1.0 second intervals. The experimental environment was the same as the one used for the illuminance trend experiment using wireless sensor nodes. The data used in communication were randomly generated 100 binary bits. Fig. 7 shows a chart of average matching rates in the cases of two-stage grading and four-stage grading at a transmission interval of 1.0 second. Fig. 7 reveals that communication in four-stage grading was as successful as communication in two-stage grading.

5. Conclusion

This study examined a data communication sheme using a MOTE MICAz wireless sensor node, a NaPiCa illuminance sensor and the luminance control of dimmable ceiling lightings as well as approaches to increasing the communication speed. First, to realize a data communication scheme based on luminance control, the trend of luminance values obtained by a NaPiCa luminance sensor was investigated. The luminance values obtained by the NaPiCa sensor showed some fluctuation even while the lighting luminance was constant. Hence, sensor nodes need to disregard the fluctuation in illuminance value occurring under a constant luminance, but to recognize an illuminance change only when the illuminance trend indicates that it is from a luminance change. An algorithm was developed in which a sensor node calculates the gradient by differentiating the illuminance value obtained, and determines whether there was an illuminance change by comparing the gradient against a preconfigured threshold. As a method to increase the speed of data communication based on ceiling light luminance control using this algorithm, two approaches were proposed – using shorter transmission intervals and transmitting more data at once – for which data communication experiments were conducted.

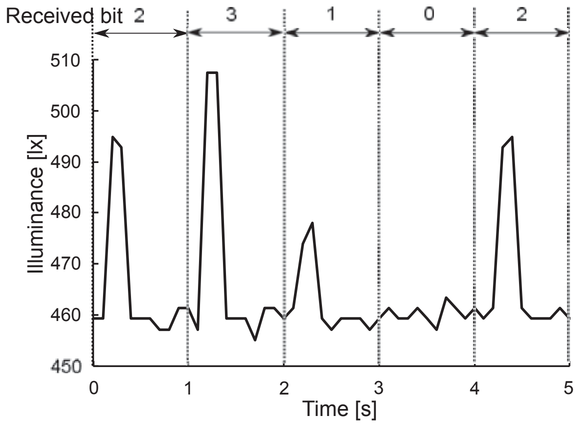


Figure 6: Illuminance values obtained at a sensor node and received bits

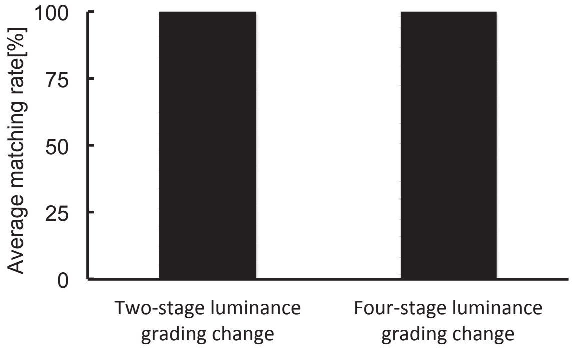


Figure 7: Average matching rates with two-stage luminance grading change and four-stage luminance grading change

The experiment using shorter transmission intervals demonstrated that the interval can be 0.2 seconds at the shortest. The transmission speed (T seconds) depends on the illuminance sensor's performance rating on the minimum illuminance measurement interval. This means that even when using an illuminance sensor other than NaPiCa, the communication speed can be increased by appropriately configuring the transmission interval (T seconds) according to the sensor's minimum illuminance measurement interval.

In the experiment of improving the communication speed by graded illuminance changes, the authors tried to improve the communication speed by increasing the amount of information sent at once by grading the illuminance in multiple stages. The result demonstrated that communication using illuminance changes graded in multiple stages was as successful as the high-speed communication at a transmission interval of 1.0 second.

The future themes for study will include developing communication methods which can cope with changing communication environments. Although the experiments for this study were conducted in an ideal environment free from the effects of daylight, illuminance sensors may be affected by a workers' shadow or display lights: hence, methods with considerations for these influences need to be developed. A disturbance such as a worker's shadow on an illuminance sensor is expected to cause a sharp change in illuminance measurements. A possible solution to his problem is to configure sensor nodes so that they will disregard such sharp changes, judging that no change in illuminance has occurred. Incorporating a process to cope with disturbances like this is expected to enable data communication based on ceiling lighting luminance control taking account of disturbances.

References

[1] I. F. Akyildiz and W. Su, Y. Sankarasubramaniam, "Wireless sensor networks: a survey", Computer Networks Journal, vol.38, no.4, pp.393-422, 2002.

[2] J. Nagashima, A. Utani and H. Ymamoto, "Advanced Particle Swarm Optimization Algorithm Computing Plural Acceptable Solutions and Its Application to Wireless Sensor Networks -Forwarding Power Adjustment of Each Sensor Node for Query Dissemination-", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, vol.23, no.1 pp.65-77, 2002.

[3] Y. Ohta, K. Yanagihara and S. Hara, "A Clustering Method Taking into Consideration of Wireless Ling Characteristic in Wireless Sensor Network", The IEICE Transactions on Information and Systems Technical Report.USN, no.107 pp.85-90, 2007.

[4] K. Suzuki, B. Mandai and T. Watanabe, "Dispersive Packets Transmission to Multiple Sinks for Prolonging Network Lifetime in Sensor Networks", The IEICE Transactions on Information and Systems.B, vol.J91-B, no.8, pp.831-843, 2008.

[5] S. Haruyama, "Visible Light Communication", The IEICE Transactions on Information and Systems.A, vol.86, no.12, pp.1284-1291, 2003.

[6] M. Noto, J. Arikawa and M.Matsuda, "Low-Power Flooding Method in Ad-Hoc Networks", The IEICE Transactions on Information and Systems.D, vol.J91, no.5, pp.1252-1260, 2008.

[7] T. Shikakura, H. Morikawa and Y. Nakamura, "Research on the Perception of Lighting Fluctuation in a Luminous Offices Environment", Journal of the Illuminating Engineering Institute of Japan., vol.85, pp.346-351, 2001.

[8] Crossbow MOTE - Wireless Sensor Networks MTS/MDA Sensor Board User's Manual, http://www.xbow.jp/mtsmdaj.pdf.

[9] Panasonic: NaPiCa, http://www3.panasonic.biz/ac/download/control/sensor/illuminance/catalog/bltn_jpn_ams.pdf.

SESSION

POSTER PAPERS

Chair(s)

TBA

Design and Implementation of Vehicular Network Platform using Wi-Fi Direct

Beomjun Kim, Yongsu Jeon, Seyoung Park, Yunju Baek
School of Electrical and Computer Engineering, Pusan National University, Busan, Republic of Korea
beomjun.kim@eslab.re.kr, yongsu.jeon@eslab.re.kr seyoung.park@eslab.re.kr, yunju@pusan.ac.kr

Abstract - *The purpose of this paper is to design and implement a vehicular network platform based on Wi-Fi Direct. The platform can provide variety of services for vehicles such as information-sharing or safety service. In the platform, Wi-Fi Direct is a communication method of the network. However, Wi-Fi Direct is not suitable for the vehicular environment because it has difficulty of making a connection at high speed. So, we propose two adaptive methods to environment; first is fixed channel allocation with constant search technique (FAST) for device discovery and another is preparation for group formation in vehicular environment (PRE) for group formation. Experimental results show that the FAST and PRE respectively reduce the delay of device discovery and group formation.*

Keywords: vehicular network platform; wi-fi direct; fast connection; scanning; grouping

1 Introduction

According to the advances in technology, there is a growing interest in a smart vehicle such as a Google's driverless car. Because the smart vehicle is based on the communication and computing technology, management for safety and information on the smart vehicle is required. However, the infrastructure or platform for vehicle-to-vehicle (V2V) network is not sufficient. The technology for this has to be developed. We design and implement a platform for these services. Based on the characteristic of vehicular environment, we propose two methods; first is the method to quickly search the communication devices on vehicles and another is the group formation method minimizing the delay of group owner selection and establishing network group simultaneously. Through the experiments, we will verify the performance of these methods.

2 Vehicular Network Platform

In the section, we describe structure of the vehicular network platform and rapid network formation methods, called FAST and PRE.

2.1 Structure of the platform

Vehicular network platform is an infrastructure collecting useful information via the sensors and delivering it to the wireless network. Each vehicle may have to send the information extracted from the sensors to a management and service application. The communication between vehicles is based on the Wi-Fi Direct [1], also called Wi-Fi P2P [2]. As shown in Fig. 1, each vehicle has dual communication interfaces. Wi-Fi Direct network consists of group owner and group client. One interface is set to group owner and another interface is set to group client. Through the connection between the interfaces, vehicles make a series of networks. If there are many group owners, the vehicle must select the group owner in consideration of the distance, wireless signal strength, relative speed and direction. In addition, each vehicle should not form the cycle through a proper connection.

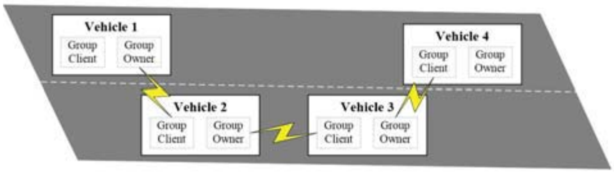


Fig. 1. Multi-hop communication with dual interfaces

2.2 Rapid network formation methods

2.2.1 Fixed channel Allocation with constant Search Technique (FAST)

FAST is a way to simplify the device discovering process through the usage of a fixed channel and continues searching. In conventional method, the way of the device searching uses a number of channels and random delay. It is based on probability; the two devices may take a long time to search for each other. It is one of the reasons why the traditional searching mechanism is not suitable for vehicular environment. Reducing the time to searching devices is very important to communicate each other, because the vehicles are driving on the road at high speed. So, FAST is designed to optimize searching the target device. Each device determines an agreed single channel, and it searches the channel without the random delay.

2.2.2 Preparation for gRoup formation in vehicular Environment (PRE)

PRE is a method to reduce a group formation time and group owner selection time, and to form a pre-group. By fixing the role of each interface, the predetermined interface is possible to form a pre-group. In conventional method, there is a process of determining group owner with negotiation before creating the group. In the negotiation process, there are a lot of communications in order to exchange information with each other. Unlike traditional methods, the proposed method can accommodate clients immediately. In terms of group formation time, it is obvious that PRE is more efficient than the traditional method.

3 Performance evaluation

In the section, we evaluate the performance for device discovery time and group formation time using open source hardware: Raspberry Pi Model B. The evaluation was performed with two devices based on Debian wheezy OS. And a wpa_supplicant [3] was used; it is IEEE 802.11x supplicant for implementation of FAST and PRE methods. Proposed methods were evaluated with default IEEE 802.11n setting of wpa_supplicant without channel scanning method and group formation order.

We performed evaluation of device discovery time with records of the 100 experiments based on UNIX time. Fig. 2 shows the cumulative delay of device discovery. Since FAST method is better than conventional one, a time difference is growing increasingly.

We also evaluate the group formation time with 50 records with actual measurement through the experiments. Fig. 3 shows the result of evaluation for PRE method. Similarly, the time difference is growing because average delay of PRE is shorter than average delay of conventional method.

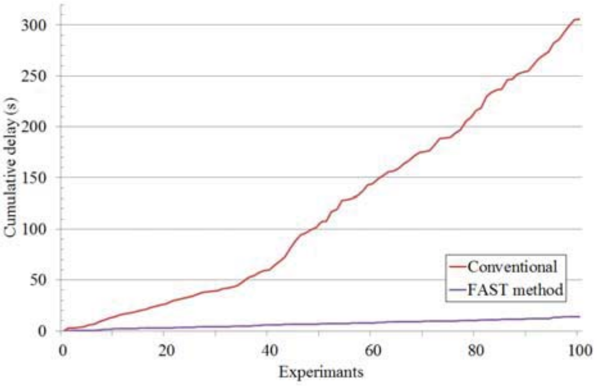


Fig. 2. Cumulative delay of device discover

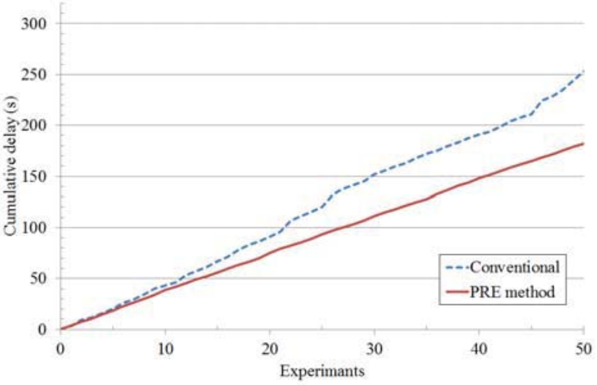


Fig. 3. Cumulative delay of group formation

We calculate the average of these evaluations at the Table 1. Average discovery time of FAST was much shorter than conventional method, and average formation time of PRE was shortening than conventional method about 1.42 seconds.

TABLE I
RESULT OF EVALUATIONS OF FAST AND PRE

	Avg. discovery time	Avg. formation time
Conventional	3.06s	5.06s
FAST	0.14s	-
PRE	-	3.64s

4 Conclusions

In this paper, we have carried out the design and implementation of the vehicular network platform using the Wi-Fi Direct devices. The platform is designed to use a specific network formation method, FAST and PRE. FAST is rapid device discovery method, and PRE is the early group formation method by forming a group in the preceding time.

We evaluate both methods through open source hardware. Through the results, we confirmed that the device discovery time is reduced 2.92 seconds when using the FAST, and group formation time is reduced 1.42 seconds when with PRE method.

The rapid network formation of our platform provides variety of services in vehicular environment. We have a plan to expand the research to combining our platform with cellular networks.

5 Acknowledgement

This work was supported by the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT & Future Planning as Global Frontier Project (CISS-2011-0031863), Corresponding author: Yunju Baek (yunju @pusan.ac.kr)

6 References

[1] Wi-Fi Alliance, “Wi-Fi Certified Wi-Fi Direct”, White paper, 2010.

[2] Wi-Fi Alliance, “Wi-Fi Peer-to-peer (P2P) Technical Specification 1.1”, 2010.

[3] wpa_supplicant, Retrieved Feb., 12, 2014, from http://wireless.kernel.org

Trustworthy and Communal Social Classifieds using HTTP and SMS

Yung-Ting Chuang
Department of Information Management, National Chung Cheng University, Chia-Yi County, Taiwan

Abstract - *As ubiquitous networked devices continue to play an increased role in the daily lives of most people, there is a growing desire to share ever more information and perspectives from across the world. However, two major problems lurk behind both social networking and community-based classified systems: privacy and security. To address this need, we propose a trustworthy and communal social classifieds, named TCSC using HTTP and SMS, allowing users to access TCSC network using a variety of devices from traditional computer desktops with wired networking to mobile ad-hoc wireless devices such as mobile phones. We hope that such infrastructure will ultimately encourage more contributions to the community, and allow users to get to know their neighbors by increasing their level of contact with users in their geographic areas.*

Keywords: Distributed System, Social Networks, Peer-to-Peer Networks, information retrieval, Privacy.

1 Introduction

Traditional social networking services, such as those of Facebook and Twitter, allows people to share and exchange information with their friends. Similarly, online classified systems like Craiglist and ebay, helps people to buy, sell, or trade items with others [9]. However, both privacy and security problems lurk behind social networking and community-based classified systems. First, there is no similar service or method that enables people to share personal information directly and easily in a distributed and de-centralized manner. Furthermore, the dissemination of information in centralized networks can be subverted or restricted by governments or corporations if the information is deemed undesirable [11]. According to [1], some countries, such as China, block or restrict access to Facebook and Twitter in order to curtail protests and political discussions. Second, the anonymity inherent in the community-based classifieds does not provide any safeguard to the users. According to [13], the anonymity feature on Craigslist has led to many crimes such as kidnapping, threats, and prostitution. The problem is that there is no way for the buyer to know the seller, or vice versa, and thus the entire transaction contains an element of risk.

2 Related Work

2.1 Peer-to-Peer Network

[10] provide comparisons of distributed search methods for peer-to-peer networks. The structured approach, like [2][5], requires the nodes to be organized in an overlay network based on distributed hash tables (DHTs), trees, rings, which is efficient but is vulnerable to manipulation by untrustworthy administrators. The unstructured approach, like [3][6], is typically based on gossiping, uses randomization, and requires the nodes to find each other by exchanging messages over existing links. Our TCSC uses the unstructured approach, which is less vulnerable to manipulation.

2.2 Combining Social Network and Communal Classifieds

Yang [14] proposes a search mechanism for unstructured peer-to-peer networks based on interest groups, formed by nodes with similar interests. Tiago [12] describe a system for mobile search in social networks based on the Drupal content site management system using the network of social links formed from the mobile phone's address book. Rather than integrating social network searches with Web searches, our TCSC utilizes social networking services with the community-based classifieds to provide users to perform searches on a particular advertisement.

Some social-networking websites have recently begun providing a framework for third-party classified applications to combine their efforts with its existing social graph. For example, Facebook Marketplace simply connects Facebook and eBay. Similarly, [13] present an interesting application called Serefind, which combines social networking and online classifieds. However, none of the above applications addresses the problem of distrust of the centralized site - not the way our TCSC does. In TCSC, we don't tackle the security problems by having an administrator to trace individuals. Rather, we utilize feedback mechanisms for users to acquire more information about others before a transaction is started.

2.3 Mobile Services over Cellular Networks

The Mobile Agent Peer-To-Peer (MAP2P) system [8] supports mobile devices in a Gnutella file-sharing network using mobile agents which acts as a proxy for the mobile device. Mobile social networking (MSN) applications, like [7], emerges social communication infrastructures, have attracted great attention recently and have been implemented pervasively. They help users to find old or new friends through similar interests, through location, through mutual friends, or through similar topics of conversation.

Some existing systems try to limit privacy leaks of social networking applications by decentralizing the social network and providing users more control over their data. Examples include decentralized social services on personal mobile devices [4]. Similarly, TCSC is based on a decentralized online social network where we aim to make distribution and requests through a subset of randomly chosen nodes to get rid of censorship, filtering, and subversion.

3 Research Method and Research Plan

In this paper, we proposed a Trustworthy and Communal Social Classifieds (TCSC) using HTTP and SMS, where users may access the system using a variety of devices from traditional computer desktops with wired networking to mobile ad-hoc wireless devices. We will first construct our TCSC such that it allows users to use mobile phones to connect to TCSC over HTTP via the Short Message Service (SMS). We will design the SMS interface and connect the TCSC SMS-HTTP bridge, which would allow any SMS-capable mobile phones to communicate with and obtain information from HTTP nodes in the TCSC network. In addition, we will design an Android user interface that builds on the basic SMS capabilities of mobile phones and that offers a user-friendly way of accessing TCSC using HTTP or SMS. After that, we will consider an environment in which nodes join or leave the network rapidly. We will design TCSC membership protocol, determine the parameters of the membership protocol, conduct performance evaluations, and discuss performance metrics for this membership protocol.

4 Acknowledgment

This research is supported by MOST 103-2410-H-194-064 of Ministry of Science and Technology, and NSC 102-2410-H-194-118 of National Science Council, Taiwan.

5 References

[1] D. Bamman, B. O'Connor, and N. Smith. Censorship and deletion practices in chinese social media. First Monday, 17(3), 2012.

[2] S. Bianchi, P. Felber, and M. Gradinariu. Content-based publish/subscribe using distributed r-trees. In Proceedings of Euro-Par, pages 537-548, Rennes, France, August 2007.

[3] I. Clarke, O. Sandberg, B. Wiley, and T. Hong. Freenet: A distributed anonymous information storage and retrieval system. In Proceedings of the Workshop on Design Issues in Anonymity and Unobservability, pages 46-66, Berkeley, CA, July 2001.

[4] B. Dodson, I. Vo, T. Purtell, A. Cannon, and M. Lam. Musubi: disintermediated interactive social feeds for mobile devices. In Proceedings of the 21st international conference on World Wide Web, pages 211-220. ACM, 2012.

[5] A. Gupta, O. Sahin, D. Agrawal, and A. El Abbadi. Meghdoot: Content-based publish/subscribe over P2P networks. In Proceedings of the 5th ACM/IFIP/USENIX International Conference on Middleware, pages 254-273, Toronto, Canada, October 2004.

[6] Gnutella. http://en.wikipedia.org/wiki/Gnutella.

[7] B. Han and A. Srinivasan. Your friends have more friends than you do: identifying influential mobile users through random walks. In Proceedings of the 13th ACM symposium on Mobile Ad Hoc Networking and Computing, pages 5-14, 2012.

[8] H. Hu, B. Thai, and A. Seneviratne. Supporting mobile devices in gnutella file sharing network with mobile agents. In Proceedings of the 8th IEEE Symposium on Computers and Communications, pages 1035-1040, Kemer-Antalya, Turkey, 2003.

[9] M. Luchs et al. Toward a sustainable marketplace: Expanding options and benefits for consumers. Journal of Research for Consumers, vol. 19, pp. 1-12, 2011.

[10] J. Mischke and B. Stiller. A methodology for design of distributed search in P2P middleware. IEEE Network, 18(1):30-37, 2004.

[11] L. Story and B. Stone. Facebook retreats on online tracking. The New York Times, 30, 2007.

[12] P. Tiago, N. Kotiainen, M. Vapa, H. Kokkinen, and J.K. Nurminen. Mobile search social network search using mobile devices. In Proceedings of the 5th IEEE Consumer Communications and Networking Conference, pages 1201-1205, LV, Neveda, Jan. 2008.

[13] P. Verma. Serefind: a social networking website for classifieds. In Proceedings of the 22nd international conference on World Wide Web companion, pages 289-292. International World Wide Web Conferences Steering Committee, 2013.

[14] J. Yang, Y. Zhong, and S. Zhang. An efficient interest-group-based search mechanism in unstructured peer-to-peer networks. In Proceedings of International Conference on Computer Networks and Mobile Computing, p. 247-252, Shanghai, China, Oct. 2003.

SESSION

LATE BREAKING PAPERS: COMMUNICATION SYSTEMS AND NOVEL APPLICATIONS

Chair(s)

TBA

Model for action prediction and inference of risk situation in smart environments

A. Alfredo Del Fabro Neto[1], **B. Bruno Romero de Azevedo**[1], **C. Rafael Boufleuer**[1],
D. Iara Augustin[1], **E. João Carlos D. Lima**[2]
[1]Informatics Graduation Program, Federal University of Santa Maria, Santa Maria/RS, Brazil
[2]Department of Languages and Computer Systems, Federal University of Santa Maria, Santa Maria/RS, Brazil

Abstract— *The risk involved in human activities is an issue not widely addressed by the ubiquitous computing community. So, this paper proposes an approach to detect the risk in activities considering the actions composing them. In order for the system to be able to aid the user, it must predict his future actions/activities and the associated risk. For that, we developed a model for action prediction based on the user's normal behavior. Our work has its foundations in the Activity Theory [1] for modeling actions and activities and determining the normal and abnormal subject's behaviors. Experiments were made for the action prediction, with an accuracy of 78,69%, and for the detection of risk situations, with an accuracy of 98,94%.*

Keywords: Action Prediction, Activity Theory, Context-awareness, Risk prediction.

1. Introduction

The growth in the availability of low cost and reduced size sensors as well as the established mobile devices technologies has enabled advances in researches in the ubiquitous and pervasive computing area [2]. Currently, mobile devices are equiped with sensors that can gather different types of data that can be used to represent the user's state and the environment in which he is inserted into. Advances in microeletronics contributed to researches in the creation of smart environments, such as smart houses, healthcare systems and service recommendation systems. This way, for providing customized services such as the ones previously cited, it is necessary to comprehend how the user interacts with the environment. For such, the systems must have access to data related to the context where the users are inserted into and, especially, analyse the interaction between a user and the environment by the detection of the activities executed by the person [3].

Therefore, the detection of activities is a crucial point in smart environments, since it is the context in which the user is inserted into [4][5]. In this sense, the detection of risk situations from an activity presents a role equally important, due to the fact that besides allow to obtain contextual information about the situation of the user while he performes an activity. The recognition of risk activities permits determining the quality of the interaction between the user and the environment. Besides that, this recognition allows the contextual information itself to be improved in a way to establish a cycle, where each activity performed by the user improves the understanding of the personal context and its relationship with the environment [3].

This work presents a model for determining risk situations based on the actions performed by the user. By analysing the user's normal behavior and based on the Activity Theory (AT) [1] and the contextual model Hyperspace Analogue to Context (HAC) [6], the proposed model is capable of infering the probability of execution of the next actions to be performed. Whereas the determination of risk situations is performed by the monitoring of physiological data while the user is performing the activity. This way, it is possible to associate a risk situation for each action or activity to be developed based on the user's normal behavior. This work falls under the project for the detection of risk situations *Activity Project*, which has a context-aware middleware (Figure 1) that encompasses the proposed model in its layer *Activity Manager* [7], [8].

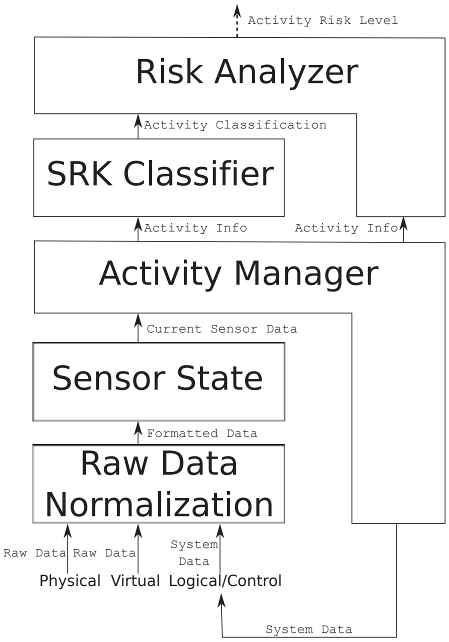


Fig. 1. Architecture from the Activity Project.

This paper is structured as follows: section 2 presents the background needed for the understanding of the proposed model; section 3 describes the proposed model and its functioning; section 4 relates the evaluation for the action prediction and the detection of risk situations; section 5 outlines related works to action prediction and risk detection and compares them with our work; and section 6 draws our final considerations and future works.

2. Background

2.1 Activity Theory (AT)

Based on the AT, an activity is composed by actions that are considered atomic units, that is, they can not be divided in more actions and its basic notion is that the subject is participating in a activity and wants to achieve some specific goal [1]. The subject's interest is focused on the object of an activity that he wishes to use and/or modify in order to achieve some expected result. The interaction between the subject and the object is mediated by artifacts. Thus, a basic triangle between subject, object and the mediation artifacts is formed (Figure 2).

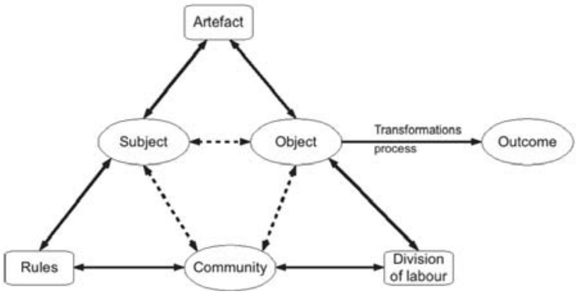


Fig. 2. Cultural-Historical Activity Theory (CHAT) [9].

Although it is possible to recognize activities based on the basic model of the AT, it does not encompasses all the components that influence this recognition. The subject also has social and cultural contexts that must be taken into consideration in the mediation with the community in which he is inserted into. For this reason, the Cultural-Historical Activity Theory (CHAT), presented in Figure 2, covers the the community component mediated by rules and the division of labor [9]. Due to the difficulty in representing computationally these concepts, the context taxonomy (Figure 3) was developed in a way to provide a pragmatic view in the construction of artifacts and incorporates general concepts of the AT to context-aware systems [10]. Thus, the CHAT components can be related to the knowledge context taxonomy, as presented in Table 1.

Table 1: Basic Aspects of an Activity and their Relation to a Taxonomy of Contextual Knowledge [11]

CHAT aspect	Category
Subject	Personal Context
Object	Task Context
Community	Spatio-Temporal Context
Mediating Artefact	Environmental Context
Mediating Rules	Task Context
Mediating Division of Labour	Social Context

2.1.1 Composition of Activities

This work uses the AT to contextually model an activity. However, since the relationship among actions and how they compose an activity for the AT are not defined from the computational point of view, it is necessary to define the structure that composes an activity. This way, a hierarchical model of activities for the activities of daily living (ADL) is used [12] which can be extended to any system that defines an activity as a composition of actions. The ADLs are composed of many actions and an ADL can compose another ADL of a higher level. For such, they are elaborated in a hierarchical plan-representing language, called Asbru, task specific and intention oriented. In Asbru, an ADL can be classified as mandatory or optional. If an ADL has sub-goals (or sub-activities, or actions) that are classified as mandatory, these actions must be performed before labeling the ADL as executed. If it is optional, the action does not have to be necessarily executed for an activity to be recognised.

Besides that, actions can be ordered in different ways. For example, in a sequential manner with a strict order of execution, in parallel when they are performed simultaneously, in any order with only one action being performed at a time and not ordered that is performed without synchronization. So, it is possible to develop a model for each of the activities that the system detects and, this way, define if an activity has restrictions in the order of exection of its actions or if they are mandatory. Lastly, it is important to highlight that the model allows the addition of temporal restrictions, enabling the implementation of a time window for the detection of the activity. This temporal restriction can be applying to the ADL or for each action that composes it.

2.2 Hyperspace Analogue to Context (HAC)

In order to detect human activities, it is necessary to detect the context attributes that occur in a pervasive environment. For this purpouse, we opted for utilizing the HAC model [6], which uses multiples dimensions for the characterization of the context with values that can vary between pre-defined thresholds. Thus, if a value is outside the threshold for some context attribute, an action can be performed. The HAC presents a well defined sintax and defines operations that allow dealing with the context data.

However, its main advantage is the possibility of capturing every context change, allowing the understanding of the user's behavior. This way, the context historical information of the user and the environment can be used to determine the risk of each action that composes the activity being developed, as well as the risk in the possible infered future actions. Thus, if a user performes an action that yields a specific context change, it is possible to evaluate if such change would yield a risk situation. Performing the simulation of context changes permits unknown scenarios to be verified with the intention of probabilistically predict if a certain action will yield a risk situation from the generated context. This work uses the context changes to infer if the change in a certain context

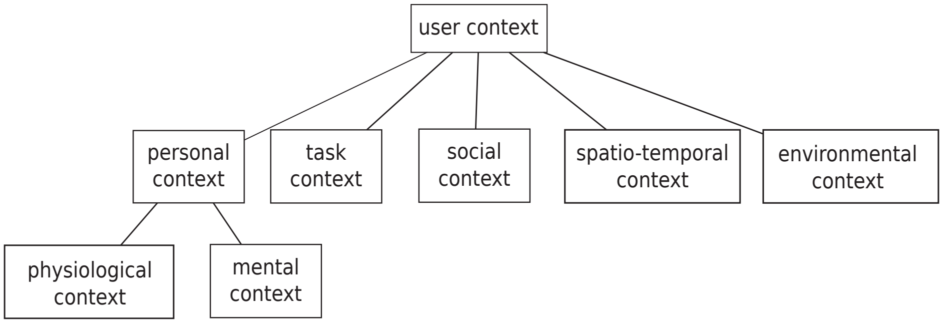


Fig. 3. Context Taxonomy. Source: [10]

attribute can result in a future context change that can harm the user, that is, if it puts him in a risk situation.

3. Proposed Model

The proposed model in this paper represents the layer *Activity Manager* of the *Activity Project* (Figure 1) and has as goal the realization of different task, which are: (i) the detection of actions and activities; (ii) the assignment of a risk situation for each action; (iii) and the inference of future actions and activities, so it is possible to predict risk situations. The structure of our model is presented in Figure 4 and it works in the following way: after receiving the aggregated sensor data (1), the first step is to recognize the action being performed (2) and infer the next action (3) to be performed from this action and based on the history of actions and activities executed by the user.

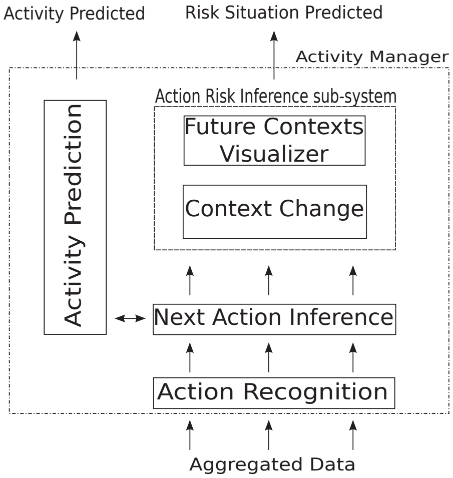


Fig. 4. Proposed model for the Activity Manager layer.

With the next action to be detected, two distinct process are initiated in parallel order: (i) the activity prediction and (ii) the action risk inference.

For the activity prediction (i), the Activity Prediction component (4) receives the probable next action and assumes it indeed happened, and requests to the Next Action Inference component (3) a new future action, using as the current action the one that was previously detected. The Activity Predicition Component (4) repeats this process until a sequence $(3-4)$ of actions that represents an activity is recognized (6).

In order to clarify this process, one can imagine that the Action Recognition component (2) recognized the current action a_1, and the Next Action Inference component (3) detects that the probable next action to be performed by the user is the action a_2. The Activity Prediction component (4) requests to the Next Action Inference component (3) the probable next action using as the base current action the action a_2. The process used to make this inference is explained in section 3.3; (ii) besides that, the Action Risk Inference subsystem (5) associates a risk situation to each infered action based on the context changes and on the future contexts yielded by these changes (7). Each of these componentes is explained in the next subsections.

3.1 Activity Recognition

Our model assumes that the activities are composed of actions and that the releationship between these actions determine how an activity happens. This way, if a certain set of actions is performed in a established time window, it is said that an activity has happened. In order to model the registered activities in the system, we used an approach similar to the one in reference [12], which permits to define if actions are or not mandatory and if they must be performed in a defined order.

With this, two types of actions are defined: (i) core actions and (ii) secondary actions. The former represents the actions that are essential in order to achieve the goal of the activity, and, therefore, necessary for its recognition. The latter are actions that are related to some activity, but are not essential for its recognition, they are useful for adding meaning to the activity. For an activity to be recognized, each action that composes it must be performed in a pre-defined time window. This way, an initial time window is determined and in the preliminary phase it is adjusted for the system's calibration. Since this window depends on the frequency of each activity, it varies according to the user's behavior.

Figure 5 presents an example for the activity recognition. In the example, there are three activities that the system is able to recognize: X, composed by the actions a_1 and a_5, with a time window of 3 periods; Y, composed by the actions a_2, a_3 and a_4, with a time window of 5 periods; and Z, composed

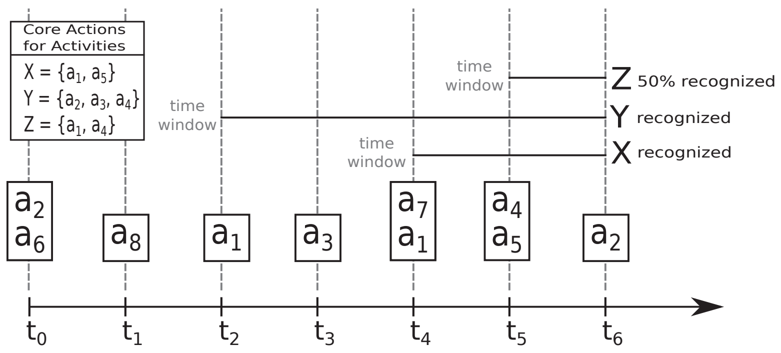


Fig. 5. Example of the used method for the activity recognition based on the AT.

by the actions a_1 and a_4, with a time window of 2 periods. The detected actions in a certain period t_i are arranged inside a box. For example, in the period t_4 the detected actions are a_1 and a_7, because the time window is only considered with the last detected actions. This way, it is possible to note that the activity X was detected, since during its time window the actions that composes it were performed and detected (a_1 in t_4 and a_5 in t_5). The same happens for the activity Y, but not for the activity Z.

An activity is initiated when an action is detected during the time window and it is considered as finished with the detection of the last pending action that composes the activity. If not all of the actions are detected during the time window, the activity is said as not complete, meaning that only some of its parts were performed. This is the case of the activity Z in Figure 5. In the shown example, X, Y and Z are happening simultaneously, althought the activity Z is not complete.

3.2 Action Recognition

The component Action Recognition receives the aggregated data from each type of sensor and, based on classification algorithms, compares the received data with the already classified data for each action registered in the system. In short, the set of sensor data will be classified according to the similarity that it has with the registered data for each action. In reference [13] are presented six machine learning algorithms for classification based on raw sensor data for the recognition of actions: *Multilayer Perceptron*, *Naive Bayes*, *Bayesian network*, *Decision Table*, *Best-First Tree* and *K-star*. The authors used the software WEKA in order to apply these algorithms over the raw data, with default patterns associated with each of the classifiers and applied to the data after the feature extraction process. The algorithm that was able to classify the highest number of samples correctly was the *K-star* and was the choosed algorithm for our model.

3.3 Next Action Inference

The inference of the next action to be performed by the user is based on the historical data of his already performed actions in order to reflect his usual behavior. For such, the algorithm 1 is based on the search for patterns of activities in the history H of the user, that is, the search for certain sequences of activities with the objective of discovering which action is the next one to probably be executed after these patterns.

Algorithm 1 Algorithm for the action prediction.

```
Require: Max Pattern Lenght $MPL$
Require: History of performed actions $H$
$N_a^p \leftarrow initZero()$
$A \leftarrow H.getLastActions(MPL)$;
$P \leftarrow getPatterns(A)$;
for all $a \in A$ do
    for all $p \in P$ do
        $N_a^p \leftarrow getNumOccurrences(H, a, p)$;
    end for
end for
$ap \leftarrow max(N)$
return $ap$
```

This way, the algorithm 1 searches the list A of the last MPL performed actions, where MPL is the window size or the quantity of actions to be analyzed. Afterwards, a search is made for the list of patterns p in the user's history for each of the actions in A previously found. Thus, the number of occurrences for each of the actions a is updated for each pattern p found and related to at most MPL periods with each action a. The highest occurence found is the probable future action.

3.4 Action Risk Inference

The determination of risk situations in actions and activities proposed assumes that each user has his own bahavior pattern, since people are considered beings of habits [14]. In order to determine the risk, the changes that each activity causes in the context has to be analized. It is worth mentioning that this analysis has to occur before an action is executed by the user. That is, it is necessary to predict the actions and, consequently, activities (composed of actions) that could be performed and this way identify if the user will be in a risk situation when he performs a certain action or activity. This approach implies in the need of (i) capturing the current user context, (ii) infering which is the next action to be executed, (iii) applying the context changes resulting from this action in the current context and (iv) evaluating the resulting context while looking for risk situations.

In order to achieve this goal, we proposed a solution to predict actions with a (i) component to discover the context changes (Context Change component) and a (ii) component to simulate future contexts (Future Contexts Visualizer) resulting from these changes (Action Risk Inference sub-system, Figure 4). From the predicted action, the Context Change component searches in the user's historical data which context changes were previously caused by it. Therefore, the Future Contexts Visualizer applies the context changes found in the current user context in order to generate a new context that represents the future state of the current context if the predicted action is performed. Based on this future context, the Future Contexts Visualizer looks in the user profile to see if it is not outside the safety tresholds preset for the user. If it is not, the user is considered to be in a risk situation.

4. Evaluation

The detection of risk situations proposed in this work is based on the prediction of actions performed by an user's previous behavior, as well as in context changes yielded by such action. In this sense, it is necessary the system to be (i) able to correctly predict the probable next actions to be executed, as well as (ii) estimate if the context changes yielded by an action will imply in a risky context for the user. In order to validate the proposal of this work, we conducted two distinct experiments. The first one intends to analyze the accuracy of the presented prediction model, while the second one intends to validate the model for the detection of risk situations based on the context changes yielded by actions in the user's context.

4.1 Evaluation of Actions Prediction

In order to evaluate the proposed model in this work, we opted for the usage of public dataset, called Aruba Dataset [15], because it allows the results of the model of actions predictions to be compared to other correlated proposals, since it is a widely used dataset in researches of activity recognition in the ubiquitous computing area. The dataset has 11 different activities registered using 42 sensors. Thus, in this work, the obtained accuracy from the used dataset was 78.69%.

4.2 Evaluation of the Risk Situation in Actions

The evaluation of risk in actions was made in a dataset of our own, since we did not find public datasets with relevant information, that is, with annotated actions and some user's physiological data. In this sense, the dataset is composed by the actions walking, sitting, running, lying and standing, which where captured from accelerometer and gyroscope data coupled in a smartphone. The physiological data gathered was the heart rate obtained from a sensor connected to an Arduino.

This way, from the 2455 entries, the model detected 49 risk situations and had an accuracy of 98.94%. This accuracy was measured based on the values true-positives (36), true-negatives (2393), false-positives (13) and false-negatives (13). These values were obtained from the analysis of the comparison between the values for the current heart rate, the predicted thresholds for the current action and the real thresholds for the current action. The predicted thresholds are determined based on the preceding action, such that it is used the median of the context changes performed by it and the value of the heart rate while it was being developed. The real thresholds are measured from the average of the historical values of the heart rate for the current action, using 3 times the standard deviation for these values, since thresholds that consider the average of the values subtracted and add to 3 times the standard deviation represents almost the entirety of the values in a normal distribution.

5. Result Discussion

The discussion of our results is made by comparing our approach with works that address similar problems. These works are briefly described in subsection 5.1 as well as the results each of them obtained. In subsection 5.2 the comparison is made.

5.1 Related Works

A system for classification of emergency situations for people that risk their lives in the line of duty, such as the firemen and the Civil Protection rescuers is presented in [16]. The operatores are equipped with two sensors in their protection clothes, an accelerometer and a ECG sensor. The system is composed by a classifier capable of recognizing many user states that correspond to many ADLs in real time. Tests were conducted in laboratory and the presented system had about 88.8% of accuracy in the activities classification.

A distributed approach that employs the computing and storage resources in each node of the wireless sensors network to detect abnormal activities is presented in [17]. In the work, an activity is defined as the combination of the trajectory and duration and an abnoraml activity is defined as the activity that deviates significantly from the trajectories and durations of the normal activities. In order to determine the normal behavior of the user, the authors performed a frequent pattern mining to find the patterns of normal activities considering their duration and trajectory. This way, if the frequency of an itemset (in this case, it is considered as an activity) exceeds the minimum threshold defined, it is classified as a normal activity. In an environment simulated by software, the accuracy was 96.2% [17].

A proposal for the prediction of household activities in a smart home is presented in [18]. The goal of the authors is to adapt the behavior of the house applications from the predicted human activities, in order to correct the behavior of devices and prepar the rooms to receive people in a pleasent condition to them. The proposal for the activities prediction is based on the construction of a directed graph for each occupant from the statistical analysis of the activities performed by him. The graph nodes represent the tasks and the edges represent the sequence of execution between two tasks, where they have the probability of execution of their sequence assigned to themselves. This way, since each task is performed in a given

Table 2: COMPARISON OF THE RELATED WORKS WITH OUR APPROACH.

Work	Accuracy	Dataset	Category		Algorithm	Attributes
			Act. Pred.	Risk Det.		
[16]	88.8%	Own (in lab)		✓	Rule Based	Accelerometer, ECG
[17]	96.2%	Own (by software)		✓	Distributed	Trajectory, Duration
[18]	61.28%	Aruba [15]	✓		Directed Graph	Action Sequence
Our Work	78.69% (action prediction) 98.94% (risk detection)	Aruba [15] (act. pred.) Own (risk det.)	✓	✓	Patterns (act. pred.) Thresholds (risk det.)	Action Sequence, Physiological Context

environment, it is possible to predict the next displacement in the graph from the current task, characterizing the prediciton of activities.

5.2 Comparison

In order to compare our proposal with the related works presented in subsection 5.1, we considered some aspects, such as: accuracy of the approach, dataset used, category (for risk detection or action prediction), algorithm used and attributes used for the risk detection. Table 2 summarizes this comparison.

In reference [16] the detection of risk situations is made by the usage of a pre-determined set of combinations between known activities, similar to a rule system. For this system to be able to identify new kind of risk situations, it is necessary the addition of new possible combinations between activities. In our approach, we consider the variation in the user's physiological data while he is performing some activity, which allows the definition of adjustable thresholds (for the risk situation detection) based on the user's history. That is, the system is able to adapt itself to changes in the user's execution of his activities, making the system more flexible.

The work presented in reference [17] uses an approach similar to ours, which considers deviations in the user's normal behavior as a risk situation. However, in such work, the authors only consider the trajectory and the duration of the activity's execution. This way, they do not account for the physiological aspects of the users that are performing activities, thus, risk situations related to changes in such physiological aspects are not detected.

The approach based on the usage of a directed graph for action prediction that considers as parameter for the measurement of the probability the ratio between the number of times that a sequence (two actions) was performed by the person and the number of times that he performed the initial action in the same edge [18], was worse than our approach because it only considers the last action performed for the inference of the next action. In this sense, we obtained better results by using an approach that allows the discovery of an appropriate pattern length for each case (based on the user's historical data), which can be used to consider not only the last action, but also a higher number of previously performed actions.

This way, considering the algorithm 1, the best value for the MPL is 2 with an accuracy of 78.69%, higher than the result obtained from the directed graph algorithm proposed in [18], which was 61.28% for the same dataset. This represents a gain of 28.41% in the future actions inference.

6. Conclusion

The prediction of risk situations is important in order to allow context-aware system to act in a preventive manner, aided in the user's decision making. Thus, this work presented a model for the action prediction and detection of future risk situations based on the Activity Theory and on the Hyperspace Analogue to Context. The used techniques were superior to other related works, since in the actions prediction we obtained an accuracy of 78.69% and in the evaluation of risk situations we obtained a accuracy of 98.94%. In future works we intend to improve the approach for actions prediction, considering the evaluation of the performance of the algorithms. Besides that, we intend to perform tests in public datasets with a higher number of the user's physiological information in order to allow a more complete evaluation of the calculated risks.

References

[1] L. S. Vygotsky, *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.

[2] B. J. d'Auriol, J. Yang, X. Wu, H. Xu, Y. Niu, J. Wang, R. A. Shaikh, M. Meng, S. Lee, and Y. Lee, "A research framework model to guide both broad and focused research into ubiquitous sensor networks," in *Proceedings of the 2007 International Conference on Wireless Networks, June 25-28, 2007, Las Vegas, Nevada, USA*, 2007, pp. 468–473.

[3] I. Mocanu and A. M. Florea, "A model for activity recognition and emergency detection in smart environments," in *AMBIENT 2011, The First International Conference on Ambient Computing, Applications, Services and Technologies*, 2011, pp. 13–19.

[4] C.-H. Lu and L.-C. Fu, "Robust location-aware activity recognition using wireless sensor network in an attentive home," vol. 6, no. 4. IEEE, 2009, pp. 598–609.

[5] C. F. Crispim-Junior, F. Bremond, and V. Joumier, "A multi-sensor approach for activity recognition in older patients," in *The Second International Conference on Ambient Computing, Applications, Services and Technologies-AMBIENT*, 2012.

[6] K. Rasch, "Smart assistants for smart homes," Ph.D. dissertation, Royal Institute of Technology, 2013.

[7] A. Del Fabro Neto, R. Boufleuer, B. Romero de Azevedo, I. Augustin, J. Carlos D. Lima, and C. C. Rocha, "Towards a middleware to infer the risk level of an activity in context-aware environments using the srk model," in *UBICOMM 2013, The Seventh International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2013, pp. 38–42.

[8] A. D. F. Neto, B. R. de Azevedo, R. Boufleuer, J. C. D. Lima, I. Augustin, and M. Pasin, "An approach based on activity theory and the srk model for risk and performance evaluation of human activities in a context-aware middleware," in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 2014, pp. 40–47.

[9] K. Kuutti, “Activity theory as a potential framework for human-computer interaction research,” 1996, pp. 17–44.

[10] M. Mikalsen and A. Kofod-Petersen, “Representing and reasoning about context in a mobile environment,” 2004, pp. 25–35.

[11] A. Kofod-Petersen and J. Cassens, “Using activity theory to model context awareness,” in *Modeling and Retrieval of Context*. Springer, 2006, pp. 1–17.

[12] U. Naeem, J. Bigham, and J. Wang, “Recognising activities of daily life using hierarchical plans,” in *Smart Sensing and Context*. Springer, 2007, pp. 175–189.

[13] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, “Simple and complex activity recognition through smart phones,” in *Intelligent Environments (IE), 2012 8th International Conference on*. IEEE, 2012, pp. 214–221.

[14] C. C. da Rocha, J. C. D. Lima, M. Viera, M. A. Capretz, M. A. Bauer, I. Augustin, and M. A. Dantas, “A context-aware authentication approach based on behavioral definitions.” in *IKE*, 2010, pp. 178–184.

[15] D. Cook, “Learning setting-generalized activity models for smart spaces.” *IEEE Intelligent Systems*, 2011.

[16] D. Curone, A. Tognetti, E. L. Secco, G. Anania, N. Carbonaro, D. De Rossi, and G. Magenes, “Heart rate and accelerometer data fusion for activity assessment of rescuers during emergency interventions,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 3, pp. 702–710, 2010.

[17] C. Wang, Q. Zheng, Y. Peng, D. De, and W.-Z. Song, “Distributed abnormal activity detection in smart environments,” *International Journal of Distributed Sensor Networks*, vol. 2014, 2014.

[18] J. Gil-Quijano and N. Sabouret, “Prediction of humans’ activity for learning the behaviors of electrical appliances in an intelligent ambient environment,” in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, vol. 2. IEEE, 2010, pp. 283–286.

A Study on the NFC based Livestock Management System for Traceability

Byeongbeom Kang[1]**, kyungjin Lee**[2]**, Hyun Yoe**[*]
[1, *] Department of Information and Communication Engineering, Sunchon National University, Suncheon-si, Jeollanam-do, Republic of Korea
[2] Agrifood Convergence ICT Research Center, Sunchon National University, Suncheon-si, Jeollanam-do, Republic of Korea

Abstract - *The livestock record tracking system to manage the record of domestic livestock transparently discloses overall distribution process from livestock production to sales. In addition, barcode printed on ear-tag of livestock is used to monitor livestock to prevent false sales such as false labeling of origin. In the case of one-dimension barcode, however, it has the weakness of not being able to store data and restore data upon contamination or damage. As a new technology to substitute such bar code, the attention is paid to NFC, a near field communication device, with the highest potential to replace bar code and RFID. This paper will proposed an information of livestock traceability managing method in the server environment using NFC (Near Field Communication). By utilizing smart devices equipped with NFC function, scanning NFC tag, and using wireless connection method such as 3g, LTE, and Wifi, a user can monitor the information as it is synchronized with the traceability information of server server. NFC based livestock tracking system is divided into raising, butchery, processing, and sale and the information collected in each stage is saved to traceability server. Then, the management information saved can be monitored through smart devices of user or web portal. Also, XML (eXtensible Markup Language) is used and managed in order to convert livestock information into electronic documents. Based on this, the information being sold is provided to consumers to allow safe purchase and when sanitation and safety issue occurs, it allows swift response such as collecting or discarding beef by tracking livestock record information*.

Keywords: NFC; Livestock; Traceability; XML; Mobile

1 Introduction

A lot of issues are raised in regards to livestock industry recently including the outbreak of livestock diseases such as BSE, foot-and-mouse disease, and others in and outside countries and fraudulent sale of livestock product including false notification of country of origin. Also, in order to resolve tracking problem in distribution process of imported beef which has been a political issue between Korea and United States or prevent safety related accidents caused by contaminated food, production information and distribution path shall be opened to the public. In addition, the tracking of distribution path of food shall be available when the safety accidents occur. Thus, there is a necessity to introduce systematic distribution management system for livestock products in relevance to such safety issues[1][2]. For the livestock tracking management system, an ear tag is attached to the ear of livestock for individual identification. Individual information of livestock can be confirmed through bar code printed to ear tags. However, in case there is no ear tag on the ear of livestock or ear tag has been damaged, it is banned from entering the butchery and it cannot be sold. A bar code printed to the ear tag is one dimensional bar code and it has shortcomings in that data cannot be read when the part of bar code is contaminated or damaged and it cannot be restored[3][4][5].

As a new technology to substitute such bar code, a great attention is paid to RFID and NFC. NFC has been an issue in smart device market together with server and LTE (Long-Term Evolution) and it is a near field communication device with the highest potential to replace RFID. In addition, it can support Card mode, RFID Reader mode for external information acquisition, and P2P mode which can transmit and receive information between devices [6].

The purpose of this study is to propose NFC based livestock traceability information management system utilizing smart devices equipped with NFC function in order to resolve such problems. NFC based livestock information management system in proposal saves and manages individual and traceability information. The data saved to traceability server can monitor the information through smart devices and web portal. Individual information of livestock in each stage from production to butchery, processing, and sale, class information based on butchery, and information on processing and sale are input and output by the manager of

[*] Corresponding author

each stage. The information based on data update can be confirmed through smart devices and web portal.

In case of utilizing smart devices, the information can be monitored by scanning the NFC tag attached to the body of livestock or sold products and searching for individual identification number. Web portal can monitor the information in more detail compared to smart devices by searching individual identification number. Also, the management XML technology is to be utilized for the livestock information. XML is a technology which supplements the shortcomings of HTML that are the expression of document, structural problem, and others. In addition, it has advantage in that it is easier to produce than HTML and data can be exchanged on the web with the same method. The web document can be monitored by converting and providing individual identification information of livestock from the smart devices with the utilization of XML[7][8].

Through the livestock traceability management system in proposal, the problems of previous method can be resolved, information from production in livestock farm to sale stage can be provided to the consumers, and the safety and reliability can be assured.

This study is composed as following. Chapter 2 gives an explanation on relevant studies and Chapter 3 presents NFC based livestock tracking management system in proposal. Chapter 4 illustrates the implementation result of proposed system and Chapter 5 concludes this study by giving a conclusion and contents of follow-up studies.

2 RELATED RESEARCH

2.1 NFC(Near Field Communication)

With the recent expansion in mobile payment market, PayPal, Telefonica/O2, and Best Buy introduced the payment system which uses NFC technology. Also, Google, ISIS, and Visa used NFC as a main technology to develop solution to call upon the wallet [9][10][11][12].

Fig. 1 illustrates the cloud based NFC payment system utilizing the POS terminal and smart devices equipped with NFC.

The contents of secure storage in smart device of user can be examined after synchronizing NFC smart device to POS terminal and utilizing the application.

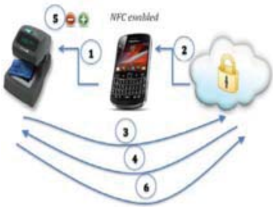


Fig. 1. NFC Service

POS terminal determines the reliability of secure contents through the communication with server supplier and transmits necessary information. Afterwards, the payment status is determined through POS terminal and data saved within server is updated accordingly with the transaction details. The utilization of NFC smart payment and range of utilization in commerce are estimated to expand through such NFC based payment system [13].

3 LIVESTOCK TRACEABILITY MANAGEMENT SYSTEM

The configuration of livestock traceability management system proposed in this study is as Fig. 2.

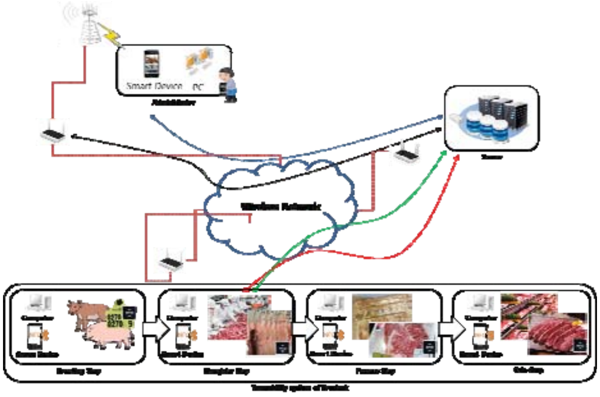


Fig. 2. System conceptual diagram

Individual and traceability information for each stage from the production of livestock to sale is as Table 1.

TABLE I. INFORMATION OF LIVESTOCK STEP

Division	Contents
Breeding Step	Object identification number, Date of birth, Sex, Kind, Area of Production, Origin information, Disease, Vaccination
Slaughter Step	Slaughter house, Date of slaughter, Information of slaughter, Sanitation class
Process Step	Rendering works, Working days, Processing information, Cut of meat
Sale Step	Sale information, Sale dates

In production stage, basic information for raising the livestock is managed and the data saved to traceability server is updated with the utilization of smart devices or PC in case of detecting abnormal symptoms in each individual. In butchery stage, the butchery related information for each livestock and hygiene information are handled and information on processing and sale is managed in processing and sale stage.

The information from production to sale stage is saved to traceability server and saved data can be monitored using smart devices and PC of user. The monitoring is available with the utilization of smart devices with NFC function by scanning the tag attached to the body of livestock or tag attached to product label or searching individual identity number by accessing to web portal.

The information of livestock saved to traceability server is managed as data in XML format and it can be examined by downloading it in document format. Then, it provides alarm accordingly with each individual when abnormal symptoms occur. Also, the infection status and path can be traced by monitoring the information on individual and information on the disease in case of disease outbreak.

The proposed system operates with processes such as Fig. 3.

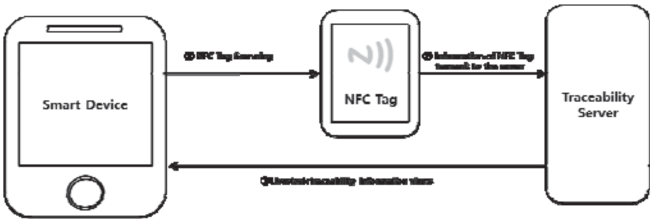

Smart Device
① NFC Tag Scanning
NFC Tag
② Information of NFC Tag transmit to the server
Traceability Server
③ Livestock traceability information view

Fig. 3. Process of the system

A user scans RFID scan or input individual identification information by utilizing NFC function of smart devices in order to search individual information of livestock. Then, the manager inputs or revises individual information of livestock through log-in process.

The collected data is processed and saved to traceability server. Individual information saved to traceability server is output to smart devices. Also, when abnormal symptoms of each individual occur, the alarm function is performed and information of each individual and disease related information are output. Through such livestock traceability management system, a user can monitor the information collected from production of livestock to sale stage and information related to traceability for each stage with the utilization of smart phones.

Elements to convert information related to individual and livestock into electronic documents shall be analyzed and electronic document element and XML tag are defined based on it. Elements to manage the livestock information are divided for each stage from the production to sale. The basic information and information on livestock husbandry is as Table 2. Basic information of livestock and information generated during the husbandry are managed. Basic information is divided into individual information, livestock species, sex, and date of birth and it can be searched through individual identification number.

TABLE II. BREEDING STEP OF LIVESTOCK

Object	XML tag	Information
Object identification number	<id>	Object identification number
Document name	<title>	Document name
Identification information	<Object>	Identification information of livestock
Kind	<Kind>	Kind of livestock
Sex	<Sex>	Sex of livestock
Date of birth	<Birth>	Date of birth
Origin information	<Breed>	Origin information of livestock
Area of Production	<O_House>	Area of Production
Fodder	<Fodder>	Fodder of livestock
Manager	<O_Manager>	Manager of production
Phone number	<O_Phone>	Phone number of Manager
Transfer	<Transfer>	Transfer of livestock
Car number	<O_Car>	Car number

The information on the husbandry is divided into place of husbandry, feed, and manager information of each livestock. Also, vehicle information shall be input when it is forwarded from the livestock farm

TABLE III. SLAUGHTER STEP OF LIVESTOCK

Object	XML tag	Information
Information of slaughter	<Slaughter>	Information of slaughter
Slaughter house	<S_House>	Place of Slaughter
Date of slaughter	<S_Date>	Date of slaughter
Sanitary inspection	<Sanitation>	Sanitary inspection
Label print	<S_Label>	Whether to print labels
DNA analysis	<DNA>	Whether to DNA analysis
Class	<Class>	Class of livestock
Car number	<S_Car>	Car number

In order to manage the information of livestock advanced to the butchery process, the object called butchery information was created and the information on butchery is as Table 3.

It confirms the attachment of label containing the information on place and date of butchery, hygiene test of butchery, and livestock information and determines the class of livestock.

TABLE IV. PROCESS STEP OF LIVESTOCK

Object	XML tag	Information
Information of process	<Process>	Information of Process
Rendering works	<P_House>	Rendering works

Slaughter inspection form	<Application_form>	Slaughter inspection form
Each part number	<Numbers>	Each part number
Parts of packing number	<P_PSN>	Parts of packing number
Whether boxed	<BOX>	Whether boxed
Working days	<P_Date>	Working days
Car number	<P_Car>	Car number

The livestock which went through the butchery process is transferred to processing factory and numbering is conducted for each body part of livestock. It is moved to packaging stage accordingly with the numbers given for each part and it is transferred to the shop after it is put into the boxes.

TABLE V. SALE STEP OF LIVESTOCK

Object	XML tag	Information
Information of sale	<Sale>	Information of sale
Information of Shop	<Sale_House>	Information of Shop
Each part number	<S_Number>	Each part number
Whether to print labels	<S_PSA>	Whether to print labels
Manager	<S_Manager>	Manager
Phone number	<S_Phone>	Phone number of Manager

The livestock which went through stages from production to butchery and processing reaches the shop in a package. The shop checks the sale management number and attachment of label for each part of livestock displayed. The manager and phone number of shop are input as well.

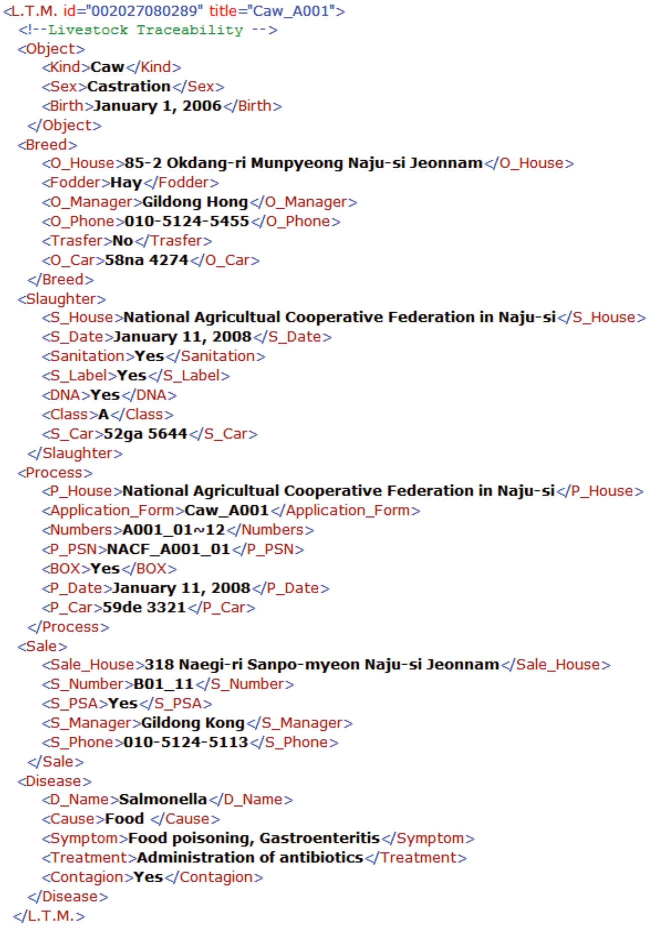
TABLE VI. DISEASE STEP OF LIVESTOCK

Object	XML tag	Information
Information of Disease	<Disease>	Information of Disease
Name of Disease	<D_Name>	Name of Disease
Cause	<Cause>	Cause of Disease
Symptom	<Symptom>	Symptom of Disease
Treatment	<Treatment>	Treatment of Disease
Shipment	<Shipment>	Shipment of Disease

In addition, the information on the disease which generated in the production process and its treatment shall be checked by handling the disease information of livestock.

In case problem occurs for the meat of relevant livestock, prompt correspondence is available such as collection, disposal, or others by tracing the information.

Electronic documents using XML are as Fig. 4.



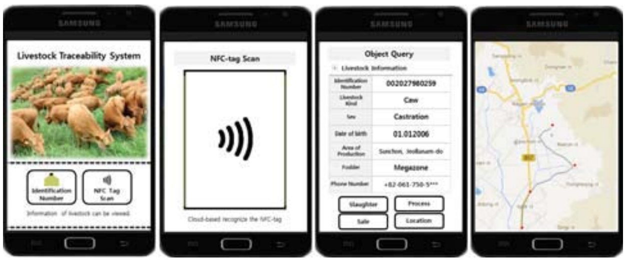
```
<L.T.M. id="002027080289" title="Caw_A001">
  <!--Livestock Traceability -->
  <Object>
     <Kind>Caw</Kind>
     <Sex>Castration</Sex>
     <Birth>January 1, 2006</Birth>
   </Object>
  <Breed>
     <O_House>85-2 Okdang-ri Munpyeong Naju-si Jeonnam</O_House>
     <Fodder>Hay</Fodder>
     <O_Manager>Gildong Hong</O_Manager>
     <O_Phone>010-5124-5455</O_Phone>
     <Trasfer>No</Trasfer>
     <O_Car>58na 4274</O_Car>
   </Breed>
  <Slaughter>
     <S_House>National Agricultual Cooperative Federation in Naju-si</S_House>
     <S_Date>January 11, 2008</S_Date>
     <Sanitation>Yes</Sanitation>
     <S_Label>Yes</S_Label>
     <DNA>Yes</DNA>
     <Class>A</Class>
     <S_Car>52ga 5644</S_Car>
   </Slaughter>
  <Process>
     <P_House>National Agricultual Cooperative Federation in Naju-si</P_House>
     <Application_Form>Caw_A001</Application_Form>
     <Numbers>A001_01~12</Numbers>
     <P_PSN>NACF_A001_01</P_PSN>
     <BOX>Yes</BOX>
     <P_Date>January 11, 2008</P_Date>
     <P_Car>59de 3321</P_Car>
   </Process>
  <Sale>
     <Sale_House>318 Naegi-ri Sanpo-myeon Naju-si Jeonnam</Sale_House>
     <S_Number>B01_11</S_Number>
     <S_PSA>Yes</S_PSA>
     <S_Manager>Gildong Kong</S_Manager>
     <S_Phone>010-5124-5113</S_Phone>
   </Sale>
  <Disease>
     <D_Name>Salmonella</D_Name>
     <Cause>Food </Cause>
     <Symptom>Food poisoning, Gastroenteritis</Symptom>
     <Treatment>Administration of antibiotics</Treatment>
     <Contagion>Yes</Contagion>
   </Disease>
 </L.T.M.>
```

Fig. 4. Electronic document capitalize on XML

It is output of individual information and information of livestock on Explorer and XML is divided into total 6 areas. Id of electronic documents for livestock traceability management is determined accordingly with individual identification number of livestock and it is devised to manage the information for each individual. It manages the individual information of livestock and information from production to butchery, processing, and sale stage. The information input by the manager of each stage and it is output by the device selected by the user. Also, by handling the information related to the disease, the information of relevant livestock can be traced for countermeasure upon the outbreak of disease.

4 MOBILE APPLICATION IMPLEMENTATION

Mobile application for livestock traceability management proposed in this study is as Fig. 6.



Identification Number	002027980259
Livestock Kind	Cow
Sex	Castration
Date of birth	01.012006
Area of Production	Sunchon, Jeollanam-do
Fodder	Megazone
Phone Number	+82-061-750-5***

Fig. 6. Mobile Application for Livestock Traceability Management System

Eclipse 4.2 (Juno) was used as a system development tool for the development of mobile application for livestock traceability management and Android SDK 4.1.2 (Jelly Bean) version was used for Android OS.

Initial screen of mobile application for livestock traceability management is composed of individual identification number and NFC Tag. A user can monitor individual information of livestock through a desirable method. Individual information of livestock is divided into individual identification number, type of livestock, sex, date of birth, owner, place of husbandry, butchery, date of butchery, and processing factory.

5 Conclusions

Domestic livestock industry insists the attachment of label which contains the livestock information in order to prevent false notification of livestock distribution market and place of origin.

A NFC based livestock traceability management system proposed in this study can monitor individual information of livestock with the utilization of NFC tag. Individual information is saved and managed to traceability server with division into production, butchery, processing, and sale stage. The data saved to server can be monitored through smart devices or web portal. Also, the disease related information can be monitored by providing alarm function together with data update when abnormal symptoms occur for each livestock. By supplementing the problem of previous system and monitoring the information on each livestock through proposed system, the information of livestock sold to the consumer can be monitored and it makes consumer feel at ease of purchasing. Also, information can be traced for countermeasure when hygiene related problems occur.

6 Acknowledgment

"This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the the C-ITRC(Convergence Information Technology Research Center) (IITP-2015-H8601-15-1007) supervised by the IITP(Institute for Information & communications Technology Promotion)"

7 References

[1] Keunjeong Jeong, “Application of cattle production traceability system”, Hankyoung Nation University, 2009, p9~16

[2] Juyoung Choi, “Traceability Management Model Supporting Safety Critical Transaction of Livestock Products”, Chunbuk National University, 2009, p8~13

[3] Minwook Hong, “Analysis of DNA Identity for Traceability in Various Cooking Methods of Beef”, Kangwon National University, 2012, p7-8.

[4] Temin Jeong, “Understanding the comparison with Barcode, RFID, QR-Code”, Korea University, 2011, p1-5

[5] Byeongbeom Kang, Hyungi Kim, Hyun Yoe, “A Study on the Mobile Application for Korean native cattle Traceability Management“, 2013, p1-2

[6] Dongwon Choi, Jae Wook, “Data Sharing Service using NFC Application in the Cloud Environnent”, “Human-computer interaction 2012”, 2012, p275~277

[7] Hongbae Son, “A Study on a Smartphone Application for the Warning and Prevention of Crime using Extensile Mark-up Language”, Inje University, 2010, p69

[8] Jinwook Lee, “Design and Implementation of Android-based Total Weather Information Application using XML Parsing Techniques”, Digital Contents Society,. 2011, p611-612

[9] Yarbrough, S., and Taylor, S., “The future of NFC payments: Is it in the Cloud or NFC?”, TSYS, April 2012.

[10] Visa, Visa Mobile Wallet. March 2013.

[11] Ankeny. J, “ISIS, USA Technologies team for mobile wallet loyalty program”, May 2013.

[12] Google. Google Wallet. March 2013.

[13] Pardis Pourghomi, “Ecosystem Scenarios for Cloud-based NFC Payments”, Proceedings of the Fifth International Conference on Management of Emergent Digital Eco Systems 2013, 2013, p116~118

The Link List with Wormhole Mechanism for Wireless Data Collection Network

Jih-Ching Chiu, Wen-Shin Wang, Chien-Lung Chen
Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan

Abstract - *Out of the 50 billion predicted nodes to be connected to data network. Data collection addressed by outdoor low-data-rate, wireless, smart metering utility network requirements is one of important applications in Internet of Things (IOT). How to provide a low-cost way to collect data within a duty time is an important issue in particular on the Low-Rate Wireless Personal Area Networks (LR-WPANs). However, the existing routing protocols are mostly on the opinions for data transmission with complex routing protocols and more hardware cost, and it's inefficient if used in data collection with polling manner. In this paper, we proposed a linked list routing algorithm with wormhole mechanism for data collection application. Using linked list topology to construct a data collecting network, where data collecting behavior can be completed on the same path so that we can gather data at one time by data fusing operation. Because the excessive hop counts of traditional linked list topology, we use the concept of wormhole to construct shortcuts to the destination node to reduce hop counts. Data can travel back to data collector quickly through shortcuts, therefore data collecting time decreases. The simulation results show that our protocol is more efficient than existing mechanisms for data collection application.*

Keywords: Internet of Things, data collection, Liked list network, wormhole

1 Introduction

Wireless sensor networks [1] (WSN) is composed of many small, low cost and low power consumption embedded sensor device. A WSN typically has no infrastructure and nodes communicate with the other node through wireless. It is originally used in military intelligence gathering. In recent year, WSN is gradually widely used in many ways such as industrial automatic control, data collection and monitoring in many different kinds of field like environment, ecology and electricity. The architecture is shown in figure 1.

Generally there is a data collector to give a command or gather data and a sensor field consists of many wireless sensor devices. Users or managers can control the data collector via internet to communicate with another node, gather data from each node and monitor its condition. Thus, the range of data collection and monitoring can be extended to difficult environment and it is unessential that users or managers must be in the place where the data collector is. To achieve this purpose, sensor devices must be capable of self-organizing themselves and there must be a good routing protocol, which can let sensor nodes automatically organize a communications network between sensor nodes. Nodes can send its data through this communications network to data collector if they receive command.

Therefore, a routing protocol is an important research in wireless sensor network. The architecture of a wireless ad-hoc network (WANET) is the most similar architecture to a WSN. They both have no infrastructure. There are many proposed routing protocol like AODV, DSR and so on used in WANET. With this paper, we discuss problems we may meet if we use these routing protocols for data collection in section 2. Then we propose a new routing protocol for data collection in section 3 and simulate it by NS2 in section 4. Section 5 concludes the paper and the future work.

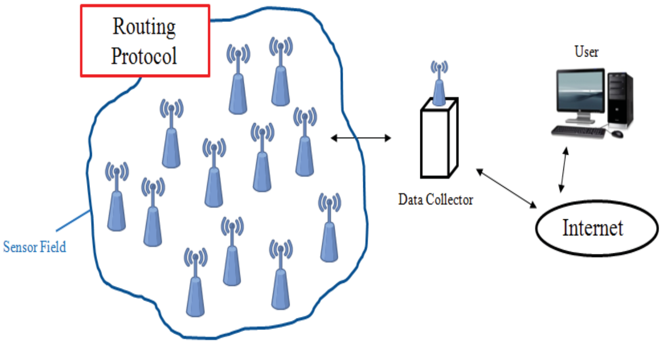


Fig. 1. WSN architecture

This study is conducted under the "Project for Research on International Energy Policy and Technologies Development on AMI Value-Added Services" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China.

2 Related research

2.1 Overview of the existing routing protocol and comparison

In the application of the data collection, due to the range of network is wild and nodes are deployed far from the sink node, we need a routing protocol which supports multi-hop manner to fit in with this topology. The existing routing protocol used in wireless sensor network and supports multi-hop manner includes mesh routing and tree routing. In this section, we discuss the difficulty we may meet when those routing protocols are used for data collection and compare with the linked list routing protocol.

In traditional tree routing [2], it is not easy to define parameters like *Cm* or *Rm* in advance because the number of sensor nodes may change. Though it takes less time in path discovery, a simple selection of the next hop address that is either parent node or child node may cause a lot of delay. Thus, many improved tree routing is proposed such as shortcut tree routing [3] and RPL [4], [5]. A shortcut tree routing is a tree routing that contains neighbor table in each node. Whenever a source node needs to transmit data to another node, every source node's neighbors calculate the route cost to the destination and choose a node which has the less route cost as source node's next hop. Therefore, the next hop address is not just parent node or child node, it can be every node in its neighbor which has the less hop count. However, if the quantity of nodes is huge, it takes too much time in calculating the route cost. RPL routing protocol is a protocol used in directed acyclic graph. RPL is a routing protocol that each node contains its own neighbor table and routing table. But there is no appropriate transmission protocol used for data collection. It takes too long to collect each node's data by using polling manner.

A mesh routing protocol is used in WANET and it is generally divided into three types. These three types of routing protocol are proactive [6], reactive [7], [8], [9], [10], [11] and hybrid routing protocol [12], [13], [14], [15]. In proactive routing protocol, each node has to maintain a routing table which contains route information to another node and it needs to upgrade the routing table frequently. It is not appropriate using in wireless sensor network because the architecture of the WSN is not stable and the bandwidth is small. A reactive routing protocol whereby routes are created on-demand like AODV is better than proactive routing protocol using in WSN. Thus, we compare AODV routing protocol with linked list routing protocol when using in a data collecting wireless network. The comparison is shown below：

Protocol Comparison	AODV	Linked list routing protocol
Construct route in data collection.	It takes some time in path discovery whenever a sink node needs to collect data	It doesn't take additional time in reconstruct route because once a linked list network is constructed. It won't change unless there is a new node.
The different in when a new node want to join an existing network.	When a new node wants to join an existing network, data collector needs to construct a path to this node.	When a new node wants to join an existing network, it only communicates with nodes nearby.
The stability of data collecting time.	The path is unstable, resulting in an unstable data collecting time.	The path is simple and the data collecting time is stable.
The manner of data collection.	Using polling manner in data collection to avoid collision, resulting in worse overall network performance.	Collecting data at one time by using only one command.
Application	Mostly used in data transmission.	

2.2 General discussion

Based on the above comparison and discussion, we know that the characteristics and advantages of a linked list network is shown below：

- To avoid unnecessary time of the route construction, the route doesn't change after finishing constructing a route unless there is a new node.
- A new node joins a network quickly and doesn't interfere with other irrelevant nodes.
- The routing table and the neighbor table are simple and stable.
    - It will not take too much time on searching the route.
    - Due to the memory capacity of the routing table is stable, it will not limit to hardware memory capacity. The routing table doesn't become larger if the quantity of nodes rise. The expansion of the network is high.
    - The path is simple and it is suitable for data collection. As long as a node can join to this path, we can collect its data.
- Using the concept of the wormhole to construct the shortcut between nodes in order to reduce the hop count, reducing data collecting time.

3 Linked list routing algorithm with wormhole mechanism

We know about the architecture and its operation of the wireless sensor network and the problem we may have when using in data collection through the analysis of the previous section. Hence, we propose a linked list routing protocol with wormhole mechanism.

Figure 2 shows the architecture of the Linked list routing algorithm with wormhole mechanism. We have one data collector and several wireless sensor nodes. Data collector sends a command to collect data. The command goes through nodes according to the direction of linked list topology path and finally reaches the final node. Each node prepares its data when receiving this command. When the final node receives command and finishes its data preparation, it initiates a data returning packet. As the data returning packet travels back from the final node to data collector, it fuses each node data along the path. The more nodes the data returning packet goes through, the greater data size it will be. When the size of data reaches the limit of packet, it travel back to data collector through wormhole and informs that the second data can begin collecting (see section 3.5.2).

In this section, first, we show that how to construct our linked list architecture and how to construct a wormhole mechanism. Second, we show that how a new node joins an existing linked list system. Third, we show that how to maintain a route when link failure.

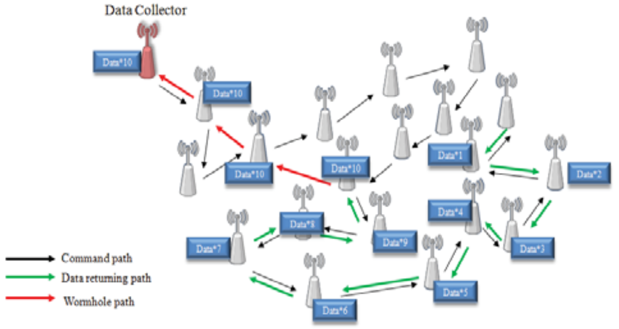


Fig. 2. The architecture of the linked list routing algorithm with wormhole mechanism

3.1 The construction of the Linked list architecture

Whenever a node starts, it requests to join an existing network by broadcasting a packet called Ask_Pkt to its neighbors. An Ask_Pkt contains its own address. Each neighbor searches its routing table when receiving an Ask_Pkt. If routing table exists that node's information, it responds a packet to let the node join network. And the node joins the network in the same place as before. The Ask_Pkt is used when a node restart, it joins the same place of the network because we do not want a node changes its place whenever it restarts, in particular the master node.

If a node doesn't receive any response after broadcasting the Ask_Pkt, then it broadcasts another packet called Join_Pkt so that it can connect with other nodes. The following steps show that how to construct a link with other nodes when a node is turned on：

When Node1 is turned on, then：

1. Node1 broadcasts Join_Pkt to its neighbor.
2. Receiving nodes respond a packet back called Join_r_Pkt that contains its own address and Next_addr.
3. Node1 receives the Join_r_Pkt which is the first packet and takes two addresses out from the packet as its Prev_addr and Next_addr, respectively. Now Node1 knows who its next node and previous node are.
4. Node1 sends Join_0x08 to its previous node. Its previous node updates its Next_addr as Node1 when receiving Join_0x08 and there is a connection between Node1 and Node1's previous node.
5. Node1 sends Join_0x0a to its next node if Node1 isn't a final node. Its next node updates its Prev_addr when receiving Join_0x0a and there is a connection between Node1 and Node1's next node.

Every node follows the above steps after it is turned on to make a link connection with other nodes.

3.2 The construction of the wormhole mechanism

After finishing constructing the linked list architecture, then we construct the wormhole mechanism.

Data collector initiates a wormhole construction command. When the next address of the data collector receives the command, it broadcasts the Worm_Pkt which contains its own address to its neighbors. Node which receives the Worm_Pkt takes the address out from the packet as its own wormhole address (Worm_addr) and continues broadcasting Worm_Pkt to its own neighbor if receiving node is a next node of source node. Its neighbors receive Worm_Pkt and update the Worm_addr if its Worm_addr is 0x00. If receiver's Worm_addr is not 0x00, it discards the Worm_Pkt because its Worm_addr has already been updated. The rest can be done in the same manner. Broadcasting the Worm_Pkt one by one followed the sequence of the linked list architecture.

After finishing constructing the wormhole mechanism, data can go back to data collector through the Worm_addr as figure 3.

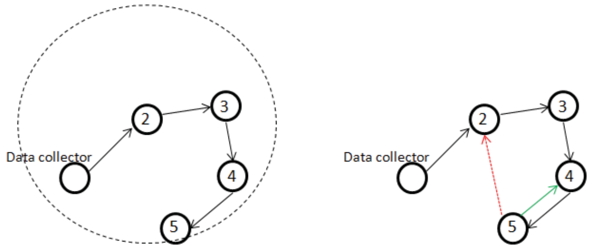
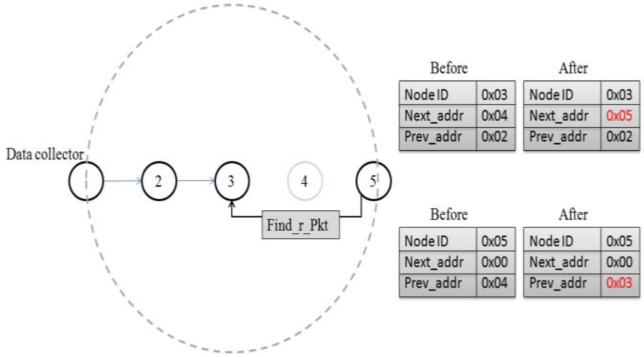


Fig. 3. Node2 broadcasts Worm_Pkt. Node3, 4, 5 upgrades its Worm_addr as node2. Thus, node5 originally has a path to node4 (green arrow) and there is a new path to node2 (red arrow).

3.3 Data jumping

If one of a node in linked list gets broken, the operation of the whole network also gets broken. Thus, we must adopt some method to avoid this condition happening.

Whenever a node sends packet to another node, the source node waits a short time in order to receive the acknowledge packet (ACK_Pkt) from destination node. If the source node does not receive the ACK_Pkt, the data jumping mechanism starts up. Source node broadcasts a packet called Find_Pkt that contains broken node's address to search another reachable node, which Prev_addr is the same as broken node's address if it is a forward direction (from data collector to final node) or Next_addr is the same as broken node's address if it is a reverse direction (final node travels back to data collector). Receiving nodes that satisfies the condition updates its Next_addr or Prev_addr (depends on the transmission direction) as source node address and then responds Find_r_Pkt to source node. The source node updates its Next_addr or Prev_addr(depends on the transmission direction) as the address in packet after receiving Find_r_Pkt. Figure 4 gives an example of data jumping mechanism.



Before

Node ID	0x03
Next_addr	0x04
Prev_addr	0x02

After

Node ID	0x03
Next_addr	0x05
Prev_addr	0x02

Before

Node ID	0x05
Next_addr	0x00
Prev_addr	0x04

After

Node ID	0x05
Next_addr	0x00
Prev_addr	0x03

Fig. 4. Suppose node4 is unreachable so node3 broadcasts Find_Pkt. Node5 responds Find_r_Pkt because its Prev_addr is node3. Node5 updates its Prev_addr as node3 after receiving Find_Pkt and node3 updates its Next_addr as node5 after receiving Find_r_Pkt.

3.4 Node restart and a new node

Once a broken node restarts, it broadcasts an Ask_Pkt and a Join_Pkt as we said in the previous section. Except the Next_addr and Prev_addr need to be updated, its Worm_addr must be updated too. In the previous section, node broadcast Worm_Pkt one by one followed the sequence of the linked list architecture. If we use the same manner to deal with a node whenever it restarts, it is inefficient. So we use another way to deal with a node which restarts. After finishing joining a network, it requests a wormhole address from its previous node. If it can reach it, then its Worm_addr is the same as its previous node. If it cannot reach it, it requests a wormhole address from its next node and updates its Worm_addr if it is reachable. If both of the wormhole address of the next node and the previous node is unreachable, its Worm_addr is the same as Prev_addr.

3.5 Data transmit and data collecting

3.5.1 Modbus extended command

We aim at the modbus protocol as application layer, developing two modbus extended command named Gather and Scatter used for data collection(see section 3.5.2) and overall data write(see section 3.5.3), respectively. The format of the Gather is the same as the format of the modbus read. The difference between them is the function code. The destination address in Gather is set to 0x00 because it doesn't matter who the destination is. The destination node is always the final node.

The format of Scatter command is shown in figure 5. It can write data to all of nodes or write data to part of nodes at one time. Scatter command carries modbus write command, using station number to determine whether the modbus read command should be executed or not.

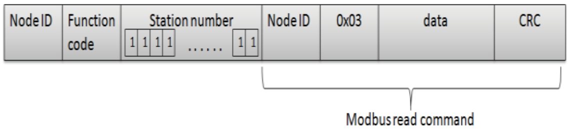


Fig. 5. The format of the Scatter command

3.5.2 Data collection using Modbus extended command

As shown in figure 6. Data collector initiates a packet of Gather command. Node which receives command prepares its data. Eventually, a Gather command will arrives at the last node. After the last node finishes preparing its data, it propagates data returning packet that contains its own data to its previous node. As the returning packet travels back, it fuses every node data into the returning packet along the path.

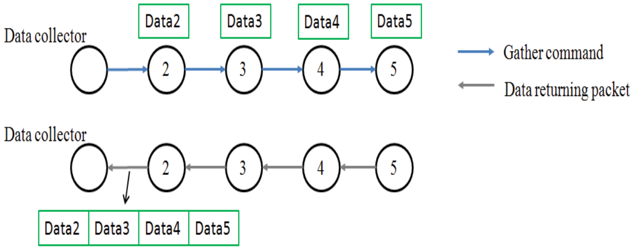


Fig. 6. Example of data gathering

When the amount of data reaches the limit of packet size, we send this first full data to data collector through wormhole and inform that the second data can start to be collected. An example is given in figure 7. Assume each node has 10 bytes data to be read and the maximum size of an IEEE802.15.4 packet is 100 bytes. As a result, node15 receives a 100 bytes packet and its own data can't fuse into it. Thus, node15 sends this first full data returning packet to node4 (assume node15's Worm_addr is node4) and then node4 sends this packet to node1 (data collector). After node4 sends packet to node1, it informs node15 that the second data returning packet can start. Similarly, the second data returning packet fuses node data from node15 to node5 and node5 propagates the second data returning packet to node1. Node1 informs node5 that the third data returning packet can start after receiving the second data returning packet. Finally, node1 receives the third data returning packet that contains data of node5, node4, node3 and node2.

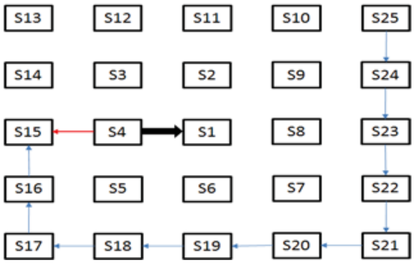


Fig. 7. After node4 sends the first returning packet (the black arrow) to node1, node4 sends a message (the red arrow) to node15 that the second returning packet can start.

4 Simulations and results

We have simulated the linked list routing protocol with wormhole mechanism using a simulator called NS2. The main objective of our simulations is to show that our routing protocol is much more efficient than AODV polling by the analysis of the data collecting time.

Nodes are placed fixedly as shown in figure 8 left. Data collector is placed at the middle of square. Arrows represent the direction of the transmission of linked list routing protocol. Each node has 10 bytes data to be read and the limit of the packet size is set to 100 bytes. In AODV polling, fist the data collector constructs the path to all nodes and collects data by polling. In our routing protocol, we construct the link and wormhole address first and then use the Gather command to collect data. The result is shown in figure 8 right. Originally, the data returning packet travels back to data collector through wormhole when its size reaches the maximum. We also simulate some conditions that the data returning packet travels back to data collector when fusing 9 node data, or 8 node data, or 7…... even 1 node data as figure 9.

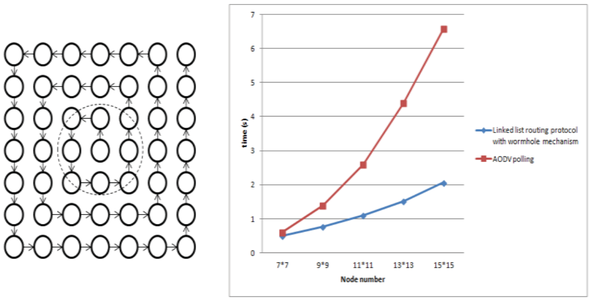


Fig. 8. Simulation result

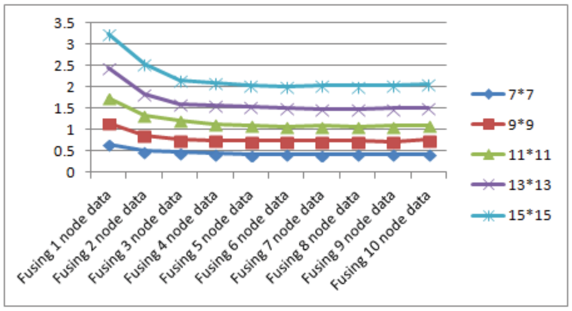


Fig. 9. A simulation that traveling back to data collection through wormhole in different timing. Generally they are almost the same, except the condition of fusing 1 node data and 2 node data.

5 Conclusion and future work

Our proposed routing algorithm aims at data collection, proposing a linked list routing algorithm with wormhole mechanism. Based on data collection, wish that it can be developed to a routing protocol that is used for Internet of Things in the future.

In this paper, we use the advantage of linked list network's simple path to gather data on the same path and then we can gain information about each node's data at one time by using Gather command. The construction of the wormhole mechanism indeed achieves the purpose of data transmission quickly, making the data collection much more efficient. The simulation results also verify that our routing protocol is efficient than the existing routing protocol. By using the advantage of linked list routing algorithm with wormhole mechanism in data collection, we wish that it can be applied to power consumption monitoring of streetlights and home appliances or any wireless sensor network fields that need data collection or monitoring. In the future, we keep researching and doing some improvements which may support mobile objects and further reduce the data collecting time without collision.

6 References

[1] Lotf, J.J.;Nazhad, S.H.H.;Alguliev, R.M.: A Survey of Wireless Sensor Networks. In Proc. Application of Information and Communication TechnologiesConf.AICT'11,Baku, AZE,12~14 (2011)

[2] Ran Peng, Sun Mao-heng, Zout You-min.: Zigbee Routing Selection Strategy Base on Data Services and Energy-balanced Zigbee Routing. APSCC, pp.400-404, 2006 IEEE Asia-Pacific Conference on Services Computing (APSCC'06) (2006)

[3] T. Kim, D. Kim, N. Park, S. E. Yoo, T. S. Lopez.: Shortcut tree routing in ZigBee networks. In Proc. IEEE International Symposium on Wireless Pervasive Computing, San Juan, Puerto Rico (2007)

[4] M. Dohler: RFC 5548-Routing Requirement for Urban Low-Power and Lossy Networks. Internet engineering task force (2009)

[5] T. Winter: RFC 6550-RPL：IPv6 Routing Protocol for Low Power and Lossy Networks. Internet engineering task force (2012)

[6] Perkin, Charles E. and Bhagwat, Pravin: Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers. In proc, ACM SIGCOMM Conference (SIGCOMM 94), pages 234-244 (1993)

[7] David B. Johnson, David A. Maltz, and Yih-Chun Hu: The Dynamic Source Routing Protocol for Mobile Ad Hoc etworks(DSR). <draft-ietf-manet-dsr-10.txt> Internet-draft (2004)

[8] C. E. Perkins and E. M. Royer: Ad Hoc On-Demand Distance Vector Routing. Second IEEE Workshop on Mobile Computing Systems and Applications, pp. 90-100 (1999)

[9] C. E. Perkins, Royer, and S. Das: Ad Hoc On-Demand Distance Vector (AODV) Routing. Internet Draft, draft-ietf-manet-aodv -13.txt (2003)V.D. Park and M.S. Corson: Temporally-Ordered Routing Algorithm (TORA) Version 1. IETS Internet draft (draft-ietfmanet-tora-spec-04.txt) (2001)

[10] Leslie Lamport. “LaTeX: A Document Preparation System”. Addison-Wesley Publishing Company, 1986.

[11] V.D. Park, J.P. Macker and M.S. Corson: Applicability of the temporally-ordered routing algorithm for use in mobile tactical networks[C]. Proceedings of MILCOM 98. Boston Marseilles: IEEE (1998) 426-430

[12] Z. J. Haas and M. R. Pearlman: The Zone Routing Protocol(ZRP)for Ad Hoc Networks. Internet draft, draft-zone-routing-protocol-01.txt (1998)

[13] M. R. Pearlman, Z. J. Haas and Syed I. Mir: Using Routing Zones to Support Route Maintenance in Ad Hoc Networks. Wireless Communications and Networking Conference, 2000. WCNC. 2000 IEEE (2000)

[14] S. J. Lee and M. Gerla: AODV-BR: Backup Routing in Ad Hoc Networks. Proceedings of IEEE WCNC 2000, Chicago, IL (2000)

[15] Tsung-Chuan Huang, Sheng-Yu Huang and Lung Tang.: AODV-Based Backup Routing Scheme in Mobile Ad Hoc Networks. In proc. IEEE 2010 International Conference on Communications and Mobile Computing (2010)

IT-based ventilation control system for optimal management of single span greenhouses

SeongJin Kim[1], **Hyun Yoe**[*]
[1, *] Department of Information and Communication Engineering, Sunchon National University, Suncheon-si, Jeollanam-do, Republic of Korea

Abstract - *The progress of civilization and technical improvement changes the systems in many fields of study at its own breakneck speed. In agriculture of these fields, many studies are proceeded to prepare the future food problem. Especially, the studies such as glasshouse and plant factory are vigorous to overcome the land question and the environmental problem by the progress of civilization. However, general vinyl greenhouse that has been used since the past is very much part of the present environment of domestic facility horticulture, and early cost of glasshouse and plant factory is expensive, so it is quite difficult to use glasshouse and plant factory in general farm. This study suggests a ventilating system based on IT in general vinyl greenhouse used at present to solve such problem. The suggested system maintains the best temperature to grow crops in general vinyl greenhouse, so collects environmental data by installing environmental sensor within vinyl greenhouse, and ventilates automatically through collected data. In addition, a ventilating control system based on web is developed not to activate an automatic ventilating system and to make user ventilate in person. The study established the system in an attached farm of Sunchon National University, and then verified the system through a challenge test. This study proceeded an experiment targeting 2 kinds of vinyl greenhouse, so the first building collected environmental information by sticking environmental sensor in vinyl greenhouse used in general farm in reality, and the second building collected environmental information by sticking environmental sensor in vinyl greenhouse with a ventilating system based on IT. As a result that compares and analyzes 2 data, temperature falls by 23%, and humidity falls by about 13%. Carbon also falls by 31%, and is with the largest drop.*

Keywords: Greenhouse; ventilation; ventilation control; Optimal environment;

1 Introduction

Domestic vinyl greenhouse grows up as the cultivation area of controlled horticulture started to be expanded from the open culture oriented agriculture in 1980s. With the government supported facility modernization project, in early 1990s, the domestic vinyl greenhouse area expanded further and now the total area of controlled horticulture is 76,580 ha, 1.8 times larger than 41,744 ha in 1990, however, it reached its peak of 93,963 ha in 2000s and has been stand still. Especially, due to its energy-intensiveness and high-cost, domestic horticulture business contracted by increase of oil price and agriculture material cost. Domestic consumption and export situation are very unstable, too. [1].

To solve the problems above, modernization of production facility is essential. However, mechanization/automation of ventilation system including automatic switchgear of astrodome and tunnel is insufficient while other basic works including plowing and soil preparation are fairly well mechanized for domestic controlled horticulture in present.

The most crucial works for the cultivation in vinyl greenhouse are temperature control and ventilation. However, most of domestic vinyl greenhouses are using motor-driven astrodome switchgear or manual open/close door by laborer. [1][2].

In this paper, a vinyl greenhouse management system with IT based ventilation system which provides the optimized environment control for cultivation is suggested.

The suggested system stores optimum temperature for cultivation on database and, through the unified server, notifies laborer when changes of temperature occur so more immediate response relative to existing system is available. Therefore, higher productivity and quality of horticultural products are achievable by reducing the products' stress and can decrease production cost by reducing energy waste.[2].

This paper is composed as follows. In chapter 2, related researches are explained. In chapter 3, the design of system is explained. Finally, conclusion wraps up this paper.

2 Related Research

2.1 Status of domestic horticulture

Domestic horticultural facilities by area is presented in Table 1.

[*] Corresponding author

TABLE I
Area of domestic horticulture.

(Unit : ha)

	2000	2005	2010
Greenhouse	51,967 (99.5%)	51,517 (98.7%)	52,850 (97.9%)
Automated greenhouses		4,466	7,790
General greenhouses	51,967	47,051	43,326
Glass greenhouse	235 (0.5%)	330 (1.3%)	329 (2.1%)
Total	53,388 (100.0%)	53,388 (100.0%)	53,388 (100.0%)

Domestic controlled horticulture can be roughly divided into automated vinyl greenhouse general vinyl greenhouse and glass greenhouse and general greenhouse occupies 81% of total domestic controlled horticulture but in downtrend while the other two in upward trend. However, the most of domestic horticulture facilities are general vinyl greenhouses not automated/mechanized. Since general vinyl greenhouse occupies the biggest portion of domestic controlled horticulture, modernization of environment control system which is suitable for domestic climate, in present, most of controlled horticulture facilities in the countries with advanced technology are modernized in opposite to domestic situation. Therefore, researches for modernized ventilation system and optimized environment management will be required. [3]

2.2 Greenhouse ventilation panels refraction development

Every year, large scale facilitated cultivation which requires stable production, shows lowering of the quality and quantity of product due to the high temperature and the ventilation of the interior/exterior of greenhouse.

Since forced ventilation system using air cooling apparatus consumes a lot of energy, maximizing natural ventilation is preferable but, most of ventilation windows of span-roof greenhouses are installed in lengthways along the ridge of roof and ventilation effect is insufficient.

For this reason, the newly developed greenhouse with folding panels is designed to have more than 2 times better ventilation capacity than existing greenhouse by bending middle of unit panel attached from the ridge of roof to eaves to open/close the roof. [5]

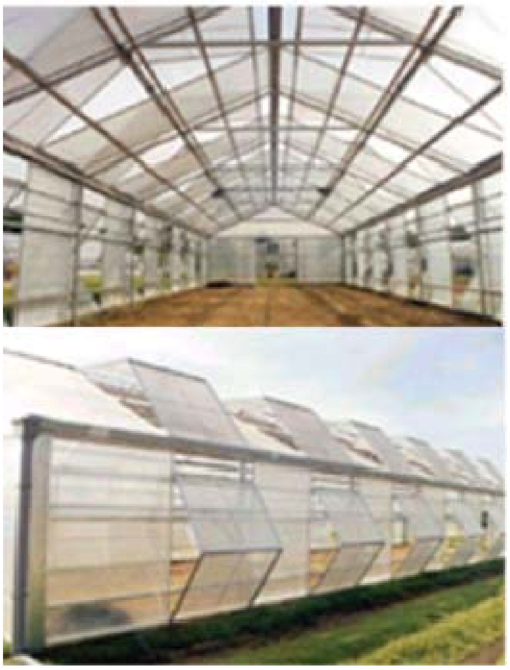


Fig. 1. Panel articulated greenhouse

A lot of research for the design of ventilation system for greenhouse is conducted in overseas countries. The future greenhouse which can be opened by lifting up whole side panels using the upper part of the side panel as a hinge and opening the ridge of roof and the Caprio greenhouse which can fully open roof side by moving along inner side and using the ridge of roof as a hinge were developed as a result of such research. [5]

One of the main objectives of greenhouse ventilation using ventilation window is all year round use of the greenhouse by maintaining similar temperature between interior/exterior of greenhouse while minimizing the energy consumption in the high temperature season.

2.3 Priva system

Priva system opens the ubiquitous ear of horticulture business.

In controlled horticulture including glass greenhouse, automation of all the system and the support for the internet based network function can provide monitoring/controlling of the greenhouse at anywhere in the world. Also, by linking with nutrient solution supply system, temperature and humidity of facilitate horticulture products can be check and controlled with ease. This was achieved by unification of atmosphere control, ventilation control, heating, energy management and water management by Priva system.

With this, the Priva system can control environments from small to middle size facilities to large scale facilities over 100 thousand m^2 and keeps pace with the farm scale-up project. On the other hand, the Priva system is compatible with existing usual desktop PC and usual farmers or users can control the system easily. [4].

Fig. 2. Priva system.

Figure 3, shows actual display for the priva system's multi-variable controller software. Temperature, humidity, CO2, the amount of water and etc. can be controlled through sensor by direct input of data values with PC. And if inputted value has a problem, the user-defined control program solves the problem by sounding three times of alarm[4].

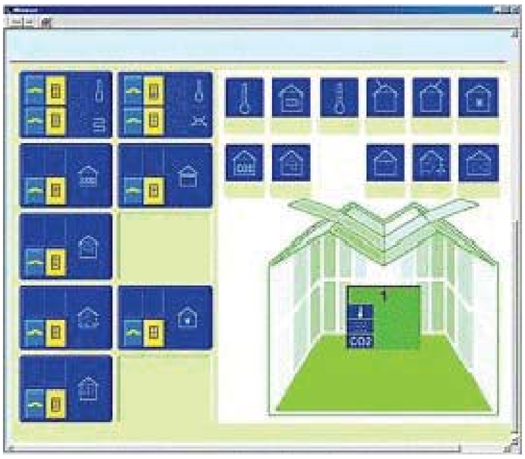


Fig. 3. Priva system operation screen.

3 IT-based ventilation system for optimal environment management of single span greenhouses

The biggest problem for domestic vinyl greenhouse environment management is the imprecise management since the farm owners check temperature and ventilate by their own decision. It leads inefficient management of vinyl greenhouse and decrease of productivity and quality of horticulture products due to the stress.

To solve the problem, the new system which is different from existing system collects exterior/interior temperature data through sensors and transmits the data collected to user while carrying out open/close of astrodome window and ventilation and solves the problem of the existing system.

Length

The maximum allowed number of pages is seven for Regular Research Papers (RRP) and Regular Research Reports (RRR); four for Short Research Papers (SRP); and two for Posters (PST).

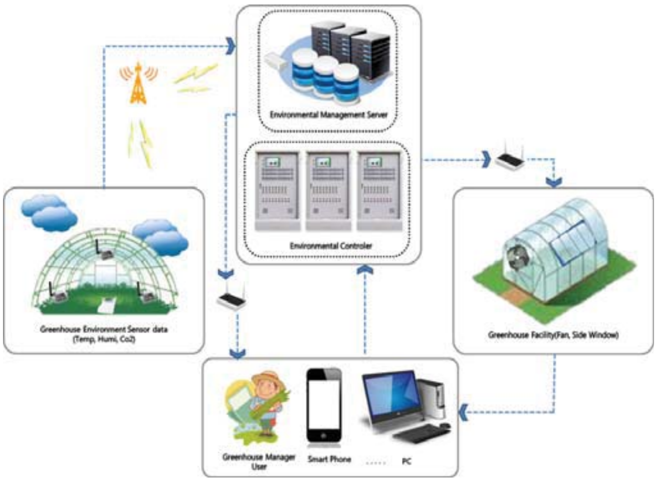

Environmental Management Server
Environmental Controler
Greenhouse Environment Sensor data (Temp, Humi, Co2)
Greenhouse Facility(Fan, Side Window)
Greenhouse Manager User
Smart Phone
.....
PC

Fig. 4. System organization.

Figure 4 shows the organizers of the paper. The system manages ventilation effectively by collecting exterior/interior temperature data through sensors and transmitting the data collected to the farm owners, immediately. The farm owners set the required temperature for the cultivation and the system is comprised of PC or smart phone device which can control open/close of the window and ventilation. The vinyl greenhouse is comprised of sensors collecting the temperature of inside of the greenhouse and control sensors which controls with the data collected. Environment management server informs the farm owners with the data collected after comparison/analysis.

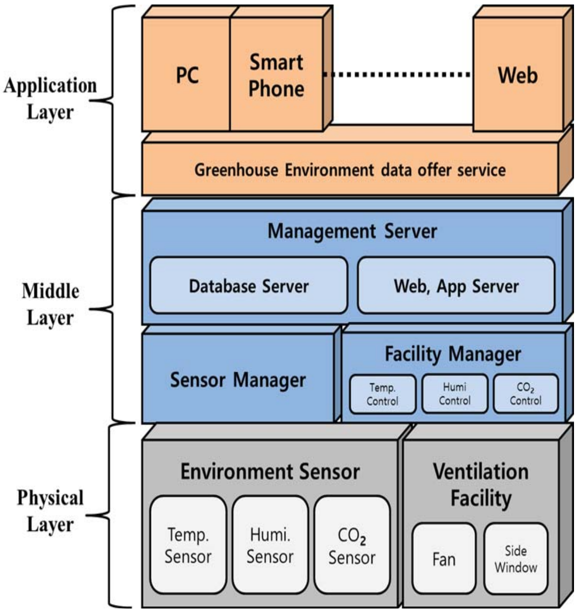


Fig. 5. architecture of system.

Architecture of this paper is shown in Figure 5.

The system is divided into physical layer, middle layer and application layer. The physical layer is comprised of sensors which can measure inner greenhouse's environment information and control part which can control environmental apparatus with the data collected. The middle layer stores/manages the data collected and is comprised of server which transmits the data to farm owners and facility manager which controls according to the data collected. The application layer is comprised of equipment which can confirm the farmer owners with the environmental information like PC, smart phone device, web and etc. and service which provides the greenhouse environmental data.

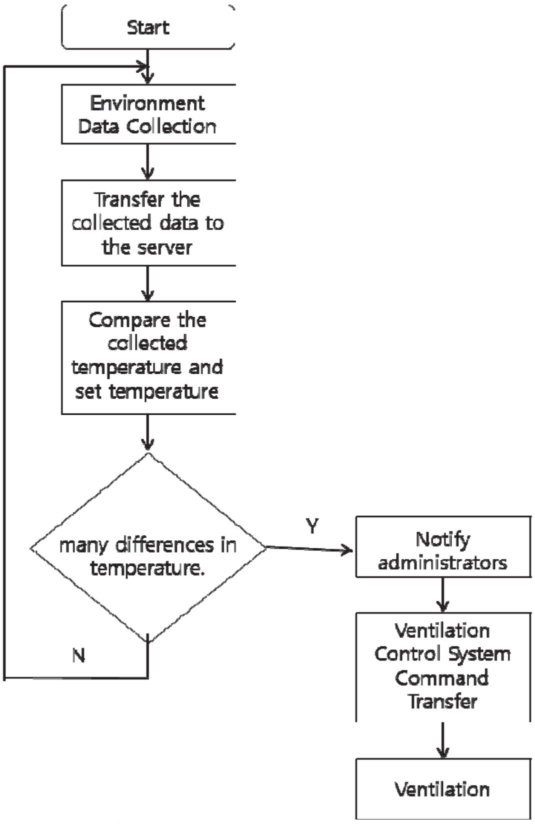

Start
Environment Data Collection
Transfer the collected data to the server
Compare the collected temperature and set temperature
many differences in temperature.
Y
N
Notify administrators
Ventilation Control System Command Transfer
Ventilation

Fig. 6. Processes in the system.

Figure 6 shows the operation process of the system suggested. Data collected by environment sensors in vinyl greenhouse are transmitted to server. And then, the server compares the data with user defined temperature. After the comparison, the sever keep collecting unless any difference detect. When temperature difference is big enough to harm the cultivation, the server notifies the farm owner to activate ventilation system through the ventilation control sensor.

The productivity/quality of horticultural product which is not suitable for domestic climate can be improved with the effective temperature control by using the system. The application of IT is the best way to maintain the suitable cultivation status in domestic farms which use vinyl greenhouse a lot. With the system suggested, increase of production and cost reduction are achievable by carrying out ventilation method which is fit for domestic climate.

The database for the proposed system was constructed as follows using Mysql Community Server 5.6.19

table 1 Configure the manager table.

Table name		TB_USER				
No	Column name	Type	Length	Null	PK	FK
1	USER_IDX	VAR	10	N	Y	
2	USER_NUM	INT		N	Y	
3	USER_PW	VAR	10	N		
4	USER_NAME	VAR	20	N		
5	GREENHOUSE_IP	CHAR	20	N		
6	GREENHOUSE_NAME	VAR	20			

Manager table can be configured with the IP of the user number, username, password, and greenhouses. In contrast to later users to increase the user was configured to be able to add subscribers.

table 2 Configure the Sensor table.

Table name		TB_SENSOR				
No	Column name	Type	Length	Null	PK	FK
1	SENSOR_NUMX	VAR	20	N	Y	
2	SENSOR_NAME	VAR	20	N		
3	SENSOR_TYPE	VAR	20	N		Y

Sensor table is composed of number and type of sensors, were configured to be able to distinguish the control for each sensor. By specifying the type of sensor in the collected data, see table FOREIGN KEY.

table 3 Configure the data collection table.

Table name		TB_DATA_TYPE				
No	Column name	Type	Length	Null	PK	FK
1	SENSOR_IDX	VAR	20	N		Y
2	SENSOR_NUM	INT		N		
3	DATA_TYPE	VAT	20		Y	
4	DATA_DT	DATE				

Collecting data table was configured to store data received from the sensor to the environment over time. Refer to the data in the setting data table by specifying the sensor ID to the FOREIGN KEY.

table 4 Configure the sensor data table.

Table name						
No	Column name	Type	Length	Null	PK	FK
1	DATA_NUM	INT		N	Y	
2	SENSOR_ID	VAR	20	N		Y
3	DATA_TEMP	INT				
4	DATA_HUMI	INT				
5	DATA_CO2	INT				
6	DATA_INTM	DATE	10			

The sensor data table stores data received by broken down from the sensor environment temperature, humidity. The configuration data set with reference to the table via a sensor data table.

4 Conclusions

The most important factor in vinyl greenhouse cultivation is the maintenance of suitable cultivation status and improvement of productivity/quality by increasing of the vitality and energy of horticulture products. The objective of the system suggested in this paper is to maintain suitable temperature through the ventilation system with sensors. The existing manual ventilation method needs unnecessary labor and due to its inefficient temperature control, stress of horticultural products increases which leads decrease of productivity/quality. With the system suggested, providing suitable environment for the cultivation and improvement of productivity/quality will be achieved.

For the future, application of the IT based vinyl greenhouse for actual farm and further improvement is expected.

5 Acknowledgment

"This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the the C-ITRC(Convergence Information Technology Research Center) (IITP-2015-H8601-15-1007) supervised by the IITP(Institute for Information & communications Technology Promotion)"

6 References

[1] Namkyu Yun, Seonghyeon Lee, Kyeongwon Kim, seonghyeon Yum, Inbok Lee, “Analysis on natural ventilation of single span greenhouse with insect screen”, National Institute of Agricultural Engineering, RDA, 441-857, pp.123-126

[2] Kyung-Hwan Yeo,In-ho Yu, Han-Cheol Rhee, Jae-Woan Cheong, Gyeong Lee Choi, ″A field survey on roof ventilation system of single-span plastic greenhouse in cucurbitaceae vegetable cultivation″ CNU Journal of Agricultural Science, Vol. 40, No. 4, pp. 317-323, December 2013.

[3] Jong-Won Lee, ″Analysis of Safety Wind Speed and Snow Depth for Single-Span Plastic Greenhouse according to Growing Crops″, Current Research on Agriculture and Life Sciences, 31(4), pp.280-285, 2013.

[4] Choi, Dong-Ho, Huh, Jong-Chul, Lim, Jong-Hwan, Suh, Hyo-Duk, ″Analysis of Indoor Thermal Environment and Cooling Effects by Ventilation Condition, and Spray irrigation or Nonspray of Single Span Plastic Greenhouses″, Journal of Bio-Environment Control, 9(1), pp.27-39, 2000

[5] Ohyoung Choi, Hanwoo Shin, Taehui Kim, Gwanghee Kim, “A Study on current Status and the problem of Agricultural Facilitie”, Journal of the Korean Geotechnical Society, Vol.8 p.147-154, December 2013

[6] Suncheol Kim, “Handbook of Protected Horticulture”, Ministry of Agriculture, Food and Rural Affairs, 2003

[7] priva, http://www.priva-international.com/en

[8] Chilgoo Choi, “Nation of horticulture and agricultural energy consumption status” Rural Development Administration, http://www.rda.go.kr/ , 2013

Energy Aware Bottleneck Nodes Avoidance Data Gathering Tree for Wireless Sensor Networks

Md. Morshedul Islam & Md. Mahfuzul Islam [1]**, and Golam Sorwar** [2]
[1]Department of Computer Science and Eng., Bangladesh Uni. of Eng. & Tech., Dhaka, Bangladesh
Email: mdmorshed.islam@ucalgary.ca; mahfuz@cse.buet.ac.bd
[2]School of Business and Tourism, Southern Cross University, NSW, Australia
Email: Golam.sorwar@scu.edu.au

Abstract – *Efficient data gathering via a heterogeneous wireless sensor network (WSN) is a challenging research area due to the limited energy of the battery used inside the tiny sensor nodes. This paper proposes a novel tree-structure based data gathering technique for WSNs called Energy aware Bottleneck-nodes Avoidance Data Gathering (EBA-DG) tree minimizing energy wastage to ensure maximum utilization of each node energy before terminating the exploration process under the network. Theoretical analysis and simulation results confirm the superior performance of the proposed EBA-DG over an existing data gathering tree-based algorithm.*

Keywords: WSNs; data gathering tree; load balancing tree; bottleneck node avoidance

1 Introduction

Recent advances in electronics and communication technologies have led to the development and deployment of WSNs with small-sized, inexpensive finite power battery-operated sensor nodes that are capable of data sensing, aggression and communication in diverse applications. Due to energy constraint with battery-operated sensors [1] and impracticality of replacing them frequently, energy efficient data gathering architecture and protocol is critical to preserve the energy and to extend the network operational lifetime [2].

Various co-operative mode data gathering protocols have been proposed to effectively utilize the resources of sensor nodes [3-7]. In each round-time, data is collected and transmitted to a base station where an end user processes those data based on user queries [8]. For the simplicity and minimum graph structure of the tree [9], a tree based topology has gained more attention than others in WSN for efficient data gathering in a continuous monitoring application with a periodic traffic pattern [10]. In this topology, a tree is constructed after initial node deployment, and is rebuilt upon significant topology changes. In this scheme, all nodes in a network form a tree by being elected/selected either as intermediate or leaf node. The intermediate nodes sense data from the environment, aggregate them with the data received itself from their descendants' neighbors, and forward the aggregated data to their predecessor neighbors. On the other hand, the leaf nodes only sense and transmit the sensed data to their corresponding intermedia (called parent) node. This process continues until all sensed data reach to the root which finally forwards them to a sink (base station). The time required to complete this process is called round-time of a network.

A number of tree construction and data gathering algorithms [11-15] have been proposed to efficiently gather and transmit data to the base station. Among them, Meghanathan [15] proposed an efficient data gathering tree, called Energy Aware Maximum Leaf node Data Gathering Tree (EML-DG), considering the residual energy of a node and their uncovered neighbors. This scheme calculates the weight of a node as a product of the residual energy and uncovered neighbors, and then computes the tree based on the weight values. In this scheme, a node with higher residual energy and larger node-degree has more weight and higher probability to be elected as an intermediate node. The proposed scheme increases the energy efficiency of a network, however, a node with relatively lower residual energy and higher node-degree (i.e., higher weight) has higher probability of being elected as an intermediate node and may drain its energy quicker than the predefined round-time.

This paper analyses the problems identified in [15] and proposes an efficient tree computing scheme, called Energy aware Bottleneck-nodes Avoidance Data Gathering (EBA–DG), by utilizing the available energy resource at nodes effectively while maintaining the tree height in limit. To compare the residual energy of a node with the energy required to survive for a predefine round-time, the proposed scheme classifies nodes into three distinct categories: rich, bottleneck and poor. The EBA-DG algorithm then constructs a tree ensuring a higher energy node with the predefined round-time survivability being elected/selected as intermediate node and handling more traffic than others. If any overburdened node is required to be elected as an intermediate node, a proper load balancing is applied to optimize their load level. Thus the proposed scheme ensures connectivity of the network for a predefine round-time and reduces the height of the tree for optimizing the data gathering delay.

The rest of this paper is organized as follows. Section 2 describes the proposed EBA-DG with system model and formulates the problem definition. Section 3 includes both simulation results and the performance evaluation of the proposed scheme in terms of network lifetime, constructed tree height and network throughput while Section 4 concludes the paper.

2 Energy aware Bottleneck-node Avoidance Data Gathering Tree (EBA-DG)

This section details the proposed EBA-DG in regards to constructing an efficient load balancing data gathering tree. The EBA-DG consists of two phases: 1) node categorization and 2) tree construction. In first phase, all nodes are categorized into three distinct categories comparing each node's residual energy with the energy required to survive for a pre-define round-time τ . In second phase, a Data Gathering (DG) tree is computed on the basis of node category ensuring higher energy nodes handle more traffic than lower ones. Before introducing formal EBA-DG algorithm, next section briefly introduces system model and some assumptions highlighting the basic characteristics of the network.

2.1 Network Model

Assume that $V = \{v_1, v_2, ..., v_n\}$ is a set of sensor nodes uniformly deployed across a field to continuously monitor the environment. The nodes form a connected graph $G(V, E, D)$ with the vertex set V. Here, E and D are the set of bidirectional wireless link and path-distance between those nodes respectively. If two nodes, v_i and v_j, are within the communication range, the link represents as $e(v_i, v_j)$ and their distance as $d_{ij} \in D$. The network is assumed to have the following characteristics:

- The network is centralized and static, i.e. the sink has information for all the nodes which are stationary after deployment.
- Sensor nodes are heterogeneous, i.e. they possess different levels of initial energy and the sink has infinite power supply.
- The root is the highest energy node in the network responsible for direct communication to the sink.
- Nodes possess finite communication ability and are within a constant range.

Fig 1 shows a sample network scenario consisting of 15 nodes with node ID 'a' to 'o'. The integer values outside the circle represent the residual energy $E_r(v_i)$ of each node v_i, where *I* = a, b,…, o. The line edge indicates the possible link between a pair of nodes. The path-cost of all links are assumed constant in the scenario.

2.2 Nodes Categorization Scheme

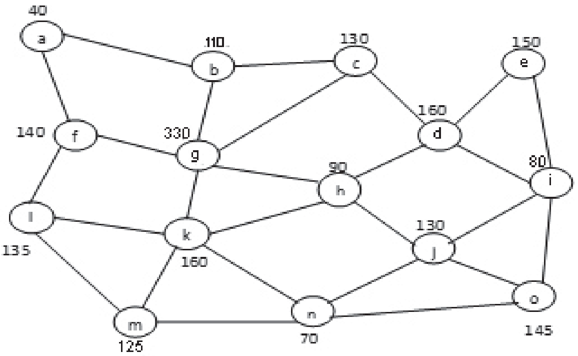


Fig 1: Network scenario with constant transmission-cost.

Constructing a balanced tree *T* using Data Gathering Sequence (DGS), the load of the nodes (except the farthest node in DGS) requires to be balanced enabling them to survive for almost the same round-time, τ either as an intermediate or leaf node. The minimum energy required to survive for an expected lifetime, round-time τ , either as leaf or intermediate node, is calculated by [16]. Energy estimated for a node with node degree one and $D(T, v_i)$ represents a leaf-node and intermediate-node energy requirement of $EL(T, v_i)$ and $EI(T, v_i)$ respectively. Comparing $EI(T, v_i)$ or $EL(T, v_i)$ with residual energy of a node, $E_r(v_i)$ nodes are categorized into three distinct categories of rich nodes (v_r), bottleneck nodes (v_b) and poor nodes (v_p) as:

$$V = \begin{cases} v_r & if\ E_r(v_i) \geq E_I(T, v_i) \\ v_b & if\ E_I(T, v_i) > E_r(v_i) > E_L(T, v_i) \\ v_p & if\ E_L(T, v_i) \geq E_r(v_i) \end{cases} \qquad (1)$$

In order to maximize the network survivability, the rich nodes, v_r , are to be selected as candidate nodes to being an intermediate node handling maximum traffic in the network. On the other hand, due to the limited energy resource, a poor category, v_p nodes function as a leaf node which is turned into sleep mode periodically to conserve their energy. Moreover, the farthest nodes in any category are considered as leaf nodes in the new constructed tree. As per (1), a 'bottleneck-node' v_b can operate either as intermediate or leaf node in a new constructed tree. The network performance heavily depends on the lifetime of this category of nodes. During tree construction, the loads on them are balanced in such way that assures their survivability for maximum round-time if acting as an intermediate node, otherwise turned into leaf nodes.

2.3 Tree Construction

In tree construction phase, the proposed EBA-DG initially selects the highest energy rich node as a root of a new constructed tree. The lifetime of root node is greater than the expected round-time of the network. The root is the first covered node in the network with level zero connected with the tree *T* and is stored in a list called covered-node-list. The root then forwards a connection request to all of its uncovered neighbors (which are yet to connect to the constructed tree *T*) with level zero. Through request-response protocol, the uncovered neighbors connect themselves to the tree and change their status to covered node. Their level is updated by increasing the level of parent node one. The newly connected nodes are also added to the covered list. The requested node becomes an intermediate node and is removed from the covered list. For example in Fig 2, node g is the highest energy rich node selected as a root in the new constructed tree *T*. Node g, then sends a connection request to its all uncovered neighbors b, c, f, h and k with level zero. These uncovered neighbors respond to the connection request and are connected to the tree showing as dotted line with arrow in Fig 2, and set their level one by increasing the level of requested node g. Nodes b, c, f, h, and k are then updated their parent list with g. The newly connected nodes b, c, f, h and k are converted into covered nodes and added to the covered list with the exclusion of g from the list due to turn into an intermediate node.

After first iteration, the algorithm recalculates the energy requirement for each node on the covered list by dropping their descendent neighbors already connected to the tree *T* and changing their status (as necessary). A node on the covered list with the highest energy status is then selected as the next potential candidate to being an intermediary node. The selected node, defined as intermediate node, then sends connection request to all of its uncovered neighbors. The newly connected nodes are added to the covered list with an increase in their level by one from their parent. The new intermediate node is removed from the covered list and the algorithm recalculates the energy requirements of all the nodes of the covered list by dropping the newly covered nodes from their neighbor list. The iteration process discussed above continues until all nodes but the leaf one(s) have completed their connection request to their descendent neighbor nodes. Fig. 3 shows final data gathering tree *T* constructed in the proposed scheme. Fig. 3 also shows that the highest and the total level of the constructed tree is 3 and $\sum level\ (v_i) = 26$, ensuring the tree height is in limit to minimize the data gathering delay that might occur in the network.

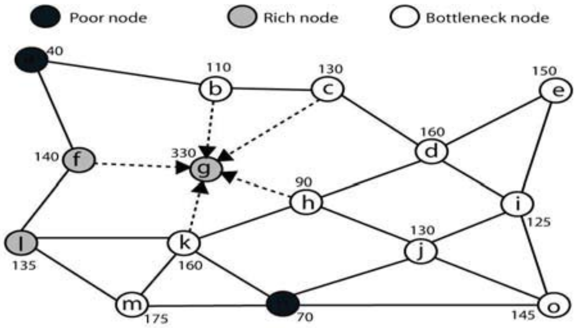


Fig 2: Tree construction (iteration 1).

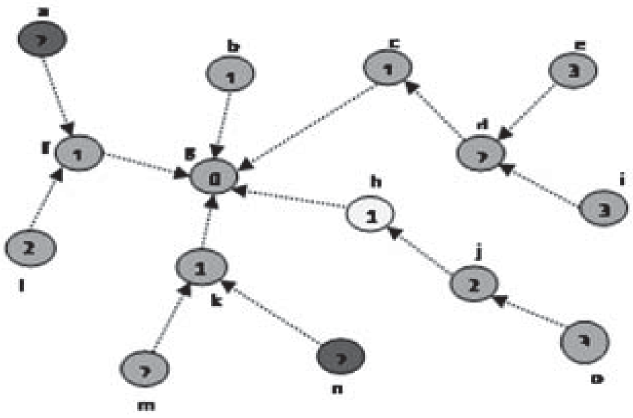


Fig 3: Final data gathering tree.

The proposed EBA-DG algorithm is formally presented in Fig. 4.

Algorithm EBA-DG (G, E$_r$)

Apply EBA-DG algorithm to the graph $G = (V, E, D)$, where V is the sensor nodes with edges$E$. D is the path distance between tow nodes which is constant. $E_r(v_i)$ represents the residual energy of the nodes with level zero. The algorithm executes as:

Step1: Calculates the energy requirements of all nodes; leaf-node energy requirement $E_L(T, v_i)$and intermediate-node energy requirement $E_I(T, v_i)$ by considering the node degree one and $D(T, v_i)$ for the round time τ. Categorize the nodes into three distinct set as, *rich node* (V_r), *bottleneck node* (V_b) *and poor nodes* (V_p) by comparing the energy requirements with their residual energy. Rich nodes get the highest priority than bottleneck nodes and poor nodes have the lowest priority. For the nodes of same priority, the tie is break by considering their level.

Step2: Selects the highest energy rich node as the root for the tree. The root becomes intermediate node by connecting all its uncovered neighbors (nodes that are not connected to the new constructed tree is called uncovered neighbor). The newly covered nodes are added into a list called covered list by increasing their level one.

Step3: Recalculate the energy requirement of all nodes in the covered list by dropping the neighbors that are covered and change status if possible.

Step4: Selects the next node with highest status and lower level from the covered list. The selected node becomes the intermediate node by connecting all its uncovered neighbors. All the newly connected nodes are added into the covered-list by increasing their level one and continue the Step3 and Step4 until all the uncovered nodes in the network become covered.

Fig 4: The EBA-DG algorithm.

3 Experimental Results

This section presents simulation results to analyze the performance of the proposed EBA-DG algorithm compared against the existing most popular scheme EML-DG [15].The experiments were carried out using well-known robust discrete-event, open source component based sensor network simulator and emulator J-sim [17].

To conduct the experiments, a 500×500 2-D free space was chosen to deploy sensor nodes randomly. The major simulation parameters with their corresponding values listed in Table I have been adopted from [18]. The following assumptions were taken into account in the experiment:

- All nodes in the network continuously generate a constant size data packets per-round and transmit them to their parent nodes.
- An intermediate node must wait to receive all data packets from their descendent child and aggregate them before transmitting them to its parent along the routing path.

As per the above sections, the appropriate selection of a round-time plays critical role in analysing the performance of a network. Experiments were conducted to estimate their empirical values before analysing the performance of the proposed algorithm. The performance of the EBA-DG scheme was compared against the EML-DG in terms of average lifetime of the network in regards to its survivability, the height of the tree related to data gathering delay of the network, and the throughput of the network related to the packet loss of the network as follows.

Table I: Simulation parameters and their values.

Parameters	**Values**
Number of Nodes N	300 ~ 500
Initial Energy of the Nodes $E_r(v_i)$	80J ~ 200J
Transmission Electronics E_e	50nJ
Transmission Amplifier ε_{mp}	100pJ
Data Packet size l	2000bits
Aggregation Cost $E_{ax}(v_i)$	5nJ
Node Deployment	Random
Simulation Area	500 × 500

3.1 Average lifetime of the network

The performance of the proposed EBA-DG and the EML-DG for network lifetime is presented in Fig 5. The figure shows that the network longevity has improved by about 31% and 40% over the existing EML-DG with 300 and 500 nodes respectively. This is due to the fact that the proposed scheme selects the higher residual energy nodes as an intermediate node and balances the load of any overburden node while constructing a new data gathering tree. Figure also shows that the life time of the network decreased with an increase in its size. Due to a change in the number of unmanaged overburden nodes and load balancing operation, the longevity of network may not vary proportionally with the network size. For example, Fig 5 shows that, in the existing scheme, there was a sudden fall, about 20%-40% lower than the initial stage, in the lifetime of the network as the node increases. On the other hand, the life time variation in EBA-DG is more consistent with the variation of network size due to an effective node categorization and load balancing technique. However, the proposed scheme always outperformed the existing scheme.

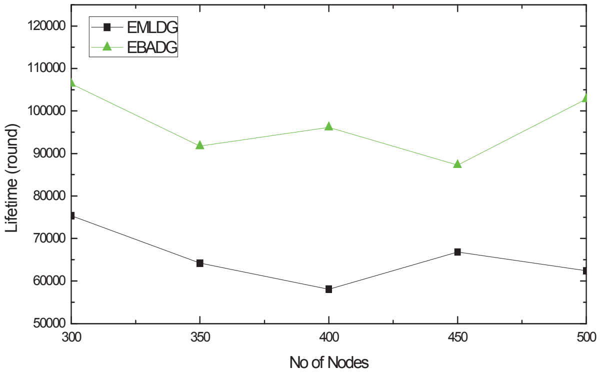


Fig 5: Average Lifetime of the network.

3.2 Height of the Constructed Tree

Fig 6 shows the performance of the EBA-DG and the EML-DG for spanning tree height. From the figure it can be seen that the proposed scheme always outperformed the EML-DG in different sized networks. Increasing the number of nodes steadily increases the load of the network, so does the height of the tree for load balancing. However, the proposed scheme always outperformed the EML-DG. For example, the height of the tree of EBA-DG and EML-DG is increased by 23.8% and 27.27% than their initial height. Therefore, the experimental results prove that the modified DGS structure in the proposed scheme maintains a tree height in a limit.

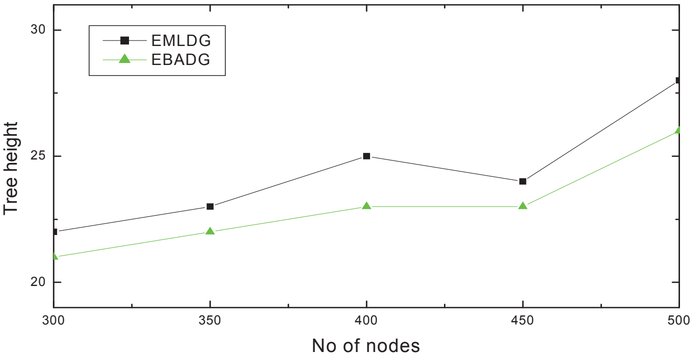


Fig 6: Height of the tree.

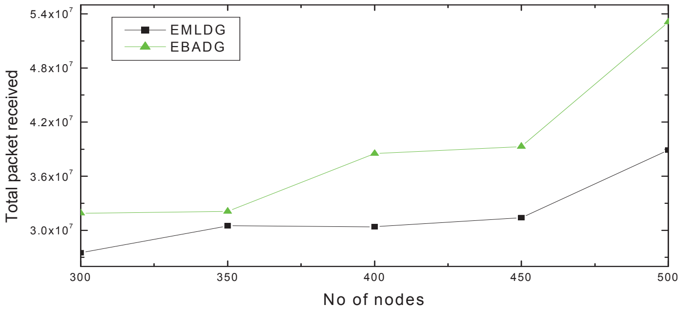


Fig 7: Throughput of the network.

3.3 Throughput of the Network

Fig 7 shows the throughput performance of the EBA-DG and EML-DG with different number of nodes. The figure evidences that the proposed scheme achieved greater throughput than the existing scheme with any number of nodes. For example, the EBA-DG achieved approximately 13.6 % higher throughput compared to EML-DG with 300 nodes. With an increase in network size, the throughput of the network increased exponentially in both schemes. Figure also evidences that some data losses occurred in both schemes due to a path interruption. However, data loss was much less in the proposed scheme due to a more effective load balancing that results in an increase of nodes strength, i.e. longevity of all intermediate nodes, of the data transmission path to a root. With 500 nodes, the throughput of the EBA-DG increased by 100% (approximately) to its initial stage compared to those in the existing EML-DG. This indicates that the EBA-DG performs better in a larger network because it selects the highest energy node as root node and increases the round-time of the overburden intermediate nodes by decreasing their load which increase the overall lifetime of the network. In addition, some of the overburden nodes become leaf nodes in newly constructed tree to generate more data packets in network lifetime.

In summary, it is evidenced from the experimental results that the proposed Energy-Aware Bottleneck-Nodes Avoidance Data Gathering (EBA-DG) scheme outperformed the existing EML-DG in terms network lifetime improvement and the overall throughput of the network.

4 Conclusions

The paper has presented a novel EBA-DG technique to address some major drawbacks identified in existing popular tree based schemes. The performance of the proposed EBA-DG has been compared to existing efficient EML-DG technique. Simulation results have proven that the EBA-DG outperforms the EML-DG by extending network life time and the throughput of the network. Hence, the proposed technique is expected to make the tree based algorithm robust and reliable in different WSNs applications.

5 References

[1] M. Zou and S. Zheng, “Energy balancing routing algorithm based on HGACA in WSNs”, Second International Conference on Computer Engineering and Technology (ICCET), vol. 2, pp. 637-640, 2010.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks”, IEEE Communi-cations Magazine, Vol. 40, No. 8, pp. 102-114, 2002.

[3] A. A. Abbasi and M. Younis, “A survey on clustering algorithms for wireless sensor networks”, Journal of Computer Communication, vol. 30, no. 14-25 pp. 2826-2841, 2007.

[4] O. Younis and S. Fahmy “HEED: A hybrid energy efficient distributed clustering approach for ad hoc sensor networks”, IEEE Transactions on Mobile Computing, vol. 3, no. 3, pp. 366-379, 2004.

[5] A. Taherkordi, R. Mohammadi, and F. Eliassen, “A communication-efficient distributed clustering algorithm for sensor networks,” 22nd IEEE International Conference on Advanced Information Networking and Applications, pp. 634-638, 2008.

[6] S. Jung, Y. Han, and T. Chung, “The concentric clustering scheme for efficient energy consumption in the PEGASIS,” 9th IEEE International Conference on Advanced Communication Technology, pp. 260-265, 2007.

[7] S. Lindsey and C. S. Raghavendra, “PEGASIS: Power efficient gath-ering in sensor information systems,” In: Proc. of the IEEE Aerospace Conference, IEEE Computer Society, pp. 1-6, 2002.

[8] J. Norman, J. P. Joseph, and P.P. Roja “A faster routing scheme for stationary wireless sensor networks - a hybrid approach,” International Journal of Ad Hoc, Sensor and Ubiquitous Computing, vol. 1, no. 1, pp. 1-10, 2010.

[9] S. Kwon, J. Kim and C. Kim “An efficient tree structure for delay sensitive data gathering in wireless sensor networks,” International Pro. of 22nd IEEE International Conference on Advance Information Networking and Applications (AINA), pp. 738-743, 2008.

[10] B. Hohlt, L. Doherty, and E. Brewer, “Flexible power scheduling for sensor networks,” In Proc. IPSN: ACM/IEEE International Conference on Information Processing in Sensor Networks, pp. 205-214, 2004.

[11] Z. Wang, and Y. Liu, ”Data gathering routing algorithm based on energy level in wireless sensor networks,” IEEE International Conference on Future Computer and Communications (ICFCC), vol. 2, pp. 160-164, 2010.

[12] Y. Wu, S. Fahmy, and N. B. Shroff, “On the construction of maximum lifetime data gathering tree in sensor networks: NP-completeness and approximation algorithm”, IEEE 27th Conference on Computer Communication (INFOCOM), pp. 356-360, 2008.

[13] V. Deepali and S. Jain, “Centralize lifetime maximizing

tree for wire-less sensor networks," International Journal of Computer and Electrical Engineering, vol. 1, no. 5, pp. 529-534, 2009.

[14] G. Jin,C. Shin, and B. Kim, "Energy-aware data gathering in wireless sensor networks," IEEE Consumer Communications and Networking Conference (CCNC), pp. 1-4, 2009.

[15] N. Meghanathan, "An algorithm to determine energy-aware maximal life nodes data gathering tree for wireless sensor networks", Journal of Theoretical and Applied Information Technology(JATIT), vol. 15, no. 2, pp. 96-107, 2010.

[16] Y. Zhu, W. Wu, J. Pan, and Y. Tang, "An energy-efficient data gathering algorithm to prolong lifetime of wireless sensor networks", Computer Communications, vol. 33, pp. 639-667, 2010.

[17] SimJava: http://www.dcs.ed.ac.uk/home/hase/simjava, last visited: 24th August, 2014.

[18] Y. Wu, S. Fahmy, and N. B. Shroff, "On the construction of maximum lifetime data gathering tree in sensor networks: NP-completeness and approximation algorithm", IEEE 27th Conference on Computer Communication (INFOCOM), pp. 356-360, 2008.

A Study of Greenhouse Meteorological disasters Prevention System using Sensor

Soonyong Kim[1], **Hyun Yoe**[*]

[1, *] Department of Information and Communication Engineering, Sunchon National University, Suncheon-si, Jeollanam-do, Republic of Korea

Abstract - *Recently, Korea greenhouse area has sharply increased 52,393ha in 2011. The meteorological factors damaging the structure of agricultural facilities are mostly strong wind and heavy snow. The damage due to a heavy rain has an effect on the structure of agricultural facilities sometimes, but has a large effect mostly on the crops and domestic animals growing inside the facilities. In greenhouse cultivation, the most economical ventilation method is to maximize the natural ventilation performance. Greenhouse industry and farmers recognize the necessity of ceiling ventilation, and the number of farms installing the ceiling ventilation device increases. In order to solve these issues, this paper is intended to study of Greenhouse Meteorological disasters Prevention System using Sensor. The greenhouse ceiling automation system perceives that snow accumulates on the ceiling of greenhouse using a pressure sensor, and informs a user of this, and makes a hot wire work at the same time. The system suggested by this paper can reduce cost necessary for restoration through reduction in damage to facilities by rapidly providing measures to cope with disasters when the manager of greenhouse cannot manage the greenhouse.*

Keywords: Greenhouse, Skylight, Sensor, Automation

1 Introduction

Recently, many greenhouses were installed so as to improve the competitiveness and productivity of agricultural products in Korea. Their area has sharply increased from 25,450ha in the early 1990s, to 52,393ha in 2011. And it was shown that double layer plastic film greenhouses occupied the largest area of 51,754ha.[1]

However, due to the socio-economic development all over the world, energy consumption has rapidly increased. Accordingly, the concentration of greenhouse gases causing the global warming has increased in the atmosphere. As a result, climate change has occurred in temperature, precipitation, and sea level, and has produced profound effect on ecosystem and economic areas. Recently, in Korea, the number of natural disasters and their effects have continued to increase. According to recent studies, it is reported that the number and intensity of disastrous meteorological phenomena increases as climate changes. The meteorological factors damaging the structure of agricultural facilities are mostly strong wind and heavy snow. The damage due to a heavy rain has an effect on the structure of agricultural facilities sometimes, but has a large effect mostly on the crops and domestic animals growing inside the facilities. Most of damage to plastic greenhouse is caused by a heavy snow and strong wind. Most of damage due to a heavy snow occurs when snow falls more than the safety snow depth on the installed plastic greenhouse. Snow is generally classified into dry snow and wet snow. Dry snow doesn't easily accumulate on the plastic greenhouse, and easily runs off. However, wet snow accumulates on the greenhouse as it is. In the last 5 years, the extent of damage to greenhouses due to meteorological disasters is 12,110ha. And damage repeatedly occurs in 2,422ha of greenhouses on average each year. A total of KRW 1,113 trillion is paid for repairing the damage to greenhouses due to these meteorological disasters. And it costs KRW 226 billion each year.[2]

In greenhouse cultivation, the most economical ventilation method is to maximize the natural ventilation performance. And in order to induce sufficient natural ventilation, it is necessary to install a ceiling ventilation system. However, a single-span plastic greenhouse has a structure that makes it difficult to install the system on the ceiling. And it is not easy to install a ceiling ventilation device in assembling, dismantling, and moving the greenhouse. Accordingly, there are many farms that install and operate a side wall window. Recently, greenhouse industry and farmers recognize the necessity of ceiling ventilation, and the number of farms installing the ceiling ventilation device increases. However, there is no standard for its installation. So, there are many issues in installation interval, the number of systems, and the specification and operation of ventilation device (ventilation window, ventilation funnel, ventilation fan, etc.).[3]

In order to solve these issues, this paper is intended to design a greenhouse ceiling automation system using a sensor. The greenhouse ceiling automation system perceives that snow accumulates on the ceiling of greenhouse using a pressure sensor, and informs a user of this, and makes a hot wire work at the same time. Besides, when rain falls and

[*] Corresponding author

when a strong wind blows, it is intended to detect this through a rain sensor and wind speed sensor, and is intended to automatically open and close the ceiling in order to reduce the risk of damage to the ceiling and in order to prevent rainwater from coming in. Through this, it is intended to automatically detect a dangerous condition by utilizing a sensor from a remote place at night or dawn when a manager is absent, and it is intended to reduce the risk of damage to the greenhouse by operating various control systems. Besides, it is intended to avoid reduction in the amount of daylight by rapidly clearing snow that has accumulated on the ceiling, and it is intended to protect the inside of greenhouse from the outside environment through automatically opening and closing the ceiling when rain falls

2 Related research

The sensor used in this paper is a pressure sensor and rain sensor. A pressure sensor is a device and element that detects the pressure of liquid or gas and converts pressure into an electric signal easy to use in instrumentation and control, and transmits the electric signal. It means the same thing as a pressure transducer in a broad sense. A pressure sensor is one of 4 major sensors that support the process automation with flow rate, liquid level, and temperature sensor. Pressure sensors are widely used from 105 bar level for making a synthetic diamond to 10-10Torr for a mass spectrometer or electron microscope. With regard to the principle of measurement, so many types are used such as what uses thermal conductivity due to a change in the density of molecules, including displacement or strain. Putting the measurement principles and transducers together, those are as shown in Table. Recently, a strain gauge-type pressure sensor made of silicon has been developed and used for precision pressure measurement. Besides, an integrated pressure sensor that also carries out signal processing by installing an integrated circuit into the same substrate has been developed.[4][5]

A rain sensor is a sensor that generates an electric signal after detecting the start and intensity of rainfall. And electric resistance type and piezoelectric type is representative type. In case of electric resistance type, a counter electrode is installed on an insulated substrate, and electric resistance is measured between them. It detects electric resistance rapidly decreasing due to short between two electrodes by a raindrop. Though the method is simple, there are many malfunctions due to stain. Besides, once it gets wet in the rain, it is slow to be restored to the original condition. In case of piezoelectric material type, a raindrop impacts a piezoelectric material such as what is used in a piezoelectric microphone, and then the generated voltage pulse is measured. It is possible to distinguish between rain and other factor by the magnitude or frequency of voltage pulse, the integrated voltage pulse, and so on. And it is possible to know the intensity of rainfall to some extent. A rain sensor of this type is used for the control of windshield wiper, and the low and high speed and stop of windshield wiper is controlled according to the amount of rain that contacts with the rain sensor. It is also used for warning that it starts to rain so that the washing doesn't get wet in the rain.[6][7]

2.1 Data Format & Sensor Network

2.1.1 Data Format

The data format used was compliant with Korea standard (TTAK.KO-06.0288-Part1). And following is a detailed specification of the data format used to the system. Structure of a typical frame transmitted between the sensor node and the server is shown in Figure 1. The first transmission frame is leftmost field. The length of the sensor node between the greenhouse server and the greenhouse to data frames is changed variably.

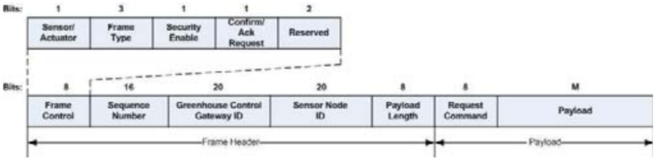


Fig. 1 Datea Structures

"Request" message is "Request-Command-Type" of 1 byte, 20-bit "Sensor Node ID", is made up of "Sensor Value" of variable length.

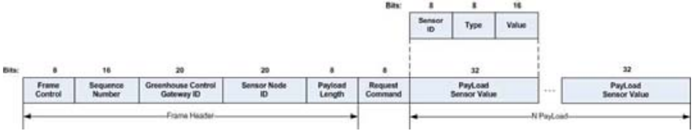


Fig. 2 Request Message

The sensor node in a response message to the "Request" message sent by the sensor node in the server transmits a "Response" message to the server. It is passed through the status information of the sensor node.[8]

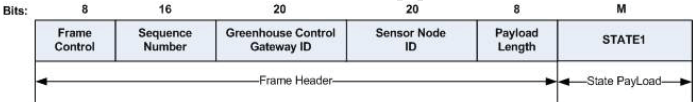


Fig. 3 Response Message

2.1.2 Sensor Network

The system used a mesh network technology. A mesh network is a network topology in which each node relays data for the network. All nodes cooperate in the distribution of data in the network. Every node in a mesh network is called a mesh node. Mesh networks can relay messages using either a flooding technique or a routing technique. With routing, the message is propagated along a path by hopping from node to node until it reaches its destination. To ensure all its paths' availability, the network must allow for continuous connections and must reconfigure itself around broken paths, using self-healing algorithms such as Shortest Path Bridging. Self-healing allows a routing-based network to operate when a node breaks down or when a connection becomes unreliable. As a result, the network is typically quite reliable, as there is often more than one path between a source and a destination in the network. Although mostly used in wireless situations, this concept can also apply to wired networks and to software interaction.[9][10]

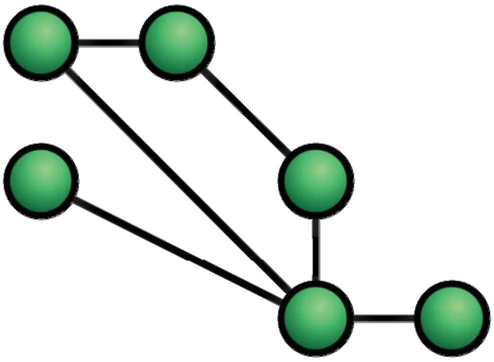


Fig. 4 Illustration of a mesh network

3 Greenhouse Meteorological disasters Prevention using Sensor System

3.1 Configuration of system

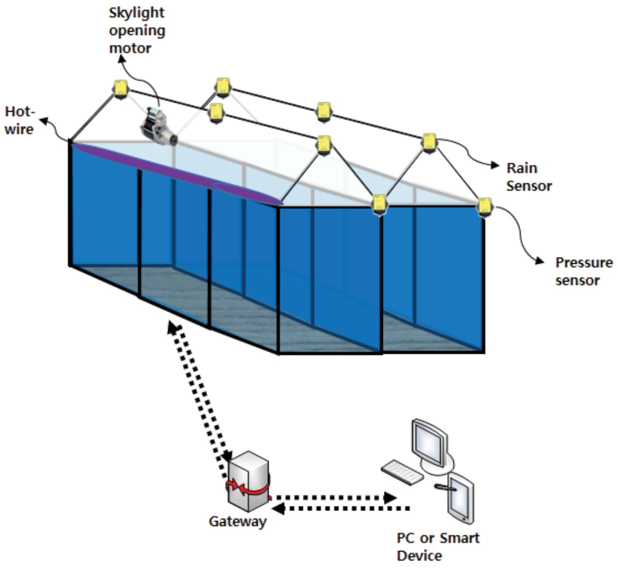
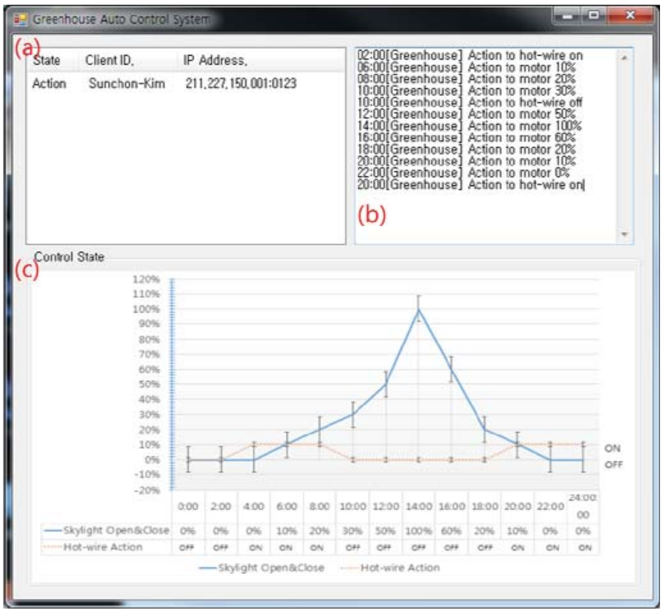


Fig. 5 configuration of system

A ceiling automation system using a sensor is a system that makes a hot wire and motor work, which is collaterally installed on the ceiling through information collected from the sensor. Fig. 5 shows the whole system configuration of this system. If a pressure sensor detects the weight of snow on the ceiling, its signal is sent to a gateway, and the gateway transmits information to PC or smart device that controls a hot wire. A controller makes the hot wire work through the gateway, and then its operating time is set according to the incoming information. Besides, when it rains, a rain sensor detects rain falling, and then a signal is sent to the gateway, and the gateway transmits information to the controller. The controller operates a motor that opens and closes the ceiling through the gateway, and then it decides how widely and when the ceiling is opened and closed according to the incoming information.

3.2 System User Interface

Fig. 6 shows a screen working in PC. And a real time incoming value from a receiver is displayed. (a) shows the access status of application, and displays information about a place that it accesses. (b) shows a system operating status displayed in text format. It was so made that a user could be informed of its operating status by displaying a signal sent by each controller and received by the same with the Korean alphabet. (c) shows a system operating status displayed in a graph. Daily operating status is displayed. And a full line indicates whether a motor for opening and closing the ceiling works. Left Y axis shows how widely the ceiling is opened and closed in operating a motor for opening and closing the ceiling in %. A dotted line shows whether a hot wire works, and is displayed in a graph of toggle type. High visibility is provided for a user by displaying its status with "On" and "Off" in the right Y axis.



State	Client ID.	IP Address.
Action	Sunchon-Kim	211.227.150.001:0123

```
02:00[Greenhouse] Action to hot-wire on
06:00[Greenhouse] Action to motor 10%
08:00[Greenhouse] Action to motor 20%
10:00[Greenhouse] Action to motor 30%
10:00[Greenhouse] Action to hot-wire off
12:00[Greenhouse] Action to motor 50%
14:00[Greenhouse] Action to motor 100%
16:00[Greenhouse] Action to motor 60%
18:00[Greenhouse] Action to motor 20%
20:00[Greenhouse] Action to motor 10%
22:00[Greenhouse] Action to motor 0%
20:00[Greenhouse] Action to hot-wire on
```

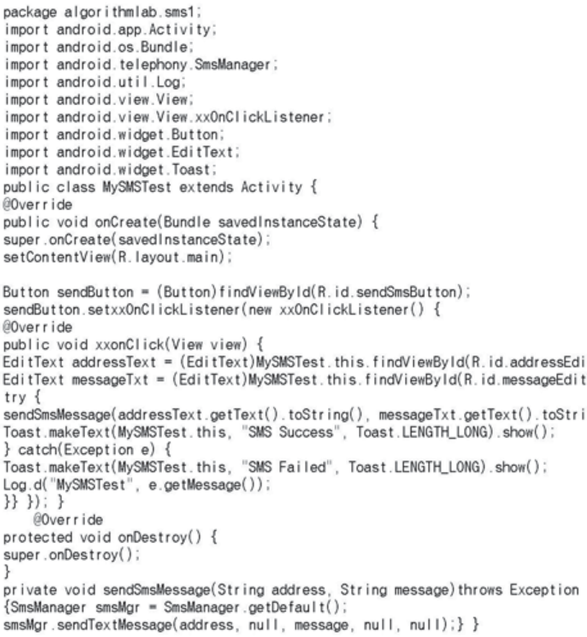
	0:00	2:00	4:00	6:00	8:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	24:00 00
Skylight Open&Close	0%	0%	0%	10%	20%	30%	50%	100%	60%	20%	10%	0%	0%
Hot-wire Action	OFF	OFF	ON	ON	ON	OFF	OFF	OFF	OFF	OFF	ON	ON	ON

Fig. 6 PC User Interface

3.3 Android Message Service

This system used SMS message. The Android SDK is defined function API for sending and receiving text messages.

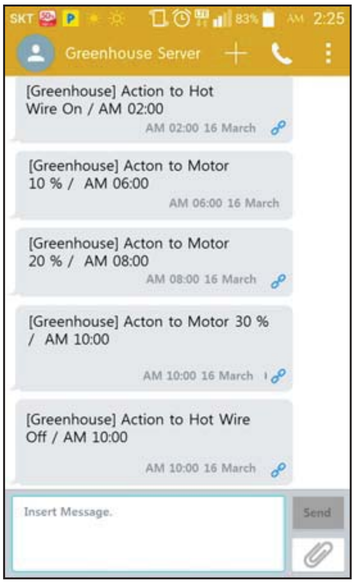
The following source is an example of one of the sources used by the system.



```
package algorithmlab.sms1;
import android.app.Activity;
import android.os.Bundle;
import android.telephony.SmsManager;
import android.util.Log;
import android.view.View;
import android.view.View.xxOnClickListener;
import android.widget.Button;
import android.widget.EditText;
import android.widget.Toast;
public class MySMSTest extends Activity {
@Override
public void onCreate(Bundle savedInstanceState) {
super.onCreate(savedInstanceState);
setContentView(R.layout.main);

Button sendButton = (Button)findViewById(R.id.sendSmsButton);
sendButton.setxxOnClickListener(new xxOnClickListener() {
@Override
public void xxonClick(View view) {
EditText addressText = (EditText)MySMSTest.this.findViewById(R.id.addressEditText);
EditText messageTxt = (EditText)MySMSTest.this.findViewById(R.id.messageEditText);
try {
sendSmsMessage(addressText.getText().toString(), messageTxt.getText().toString());
Toast.makeText(MySMSTest.this, "SMS Success", Toast.LENGTH_LONG).show();
} catch(Exception e) {
Toast.makeText(MySMSTest.this, "SMS Failed", Toast.LENGTH_LONG).show();
Log.d("MySMSTest", e.getMessage());
}} }); }
    @Override
protected void onDestroy() {
super.onDestroy();
}
private void sendSmsMessage(String address, String message)throws Exception
{SmsManager smsMgr = SmsManager.getDefault();
smsMgr.sendTextMessage(address, null, message, null, null);} }
```

Fig. 7 Android Message Send Sources

Fig.8 shows an Android application established for a test. In case an IP address is set, at first, and then 'Connect' button is pressed, the application works. And the control of each device is shown. The user is informed of its operating status through push up message. And "Automatic" or "Manual" is selected in the application. And in case "Automatic" is selected, it is automatically managed by a value set in PC. Besides, in case "Manual" is selected, it is so designed that a user can operate a device with an arbitrary value.



[Greenhouse] Action to Hot Wire On / AM 02:00
AM 02:00 16 March

[Greenhouse] Acton to Motor 10 % / AM 06:00
AM 06:00 16 March

[Greenhouse] Acton to Motor 20 % / AM 08:00
AM 08:00 16 March

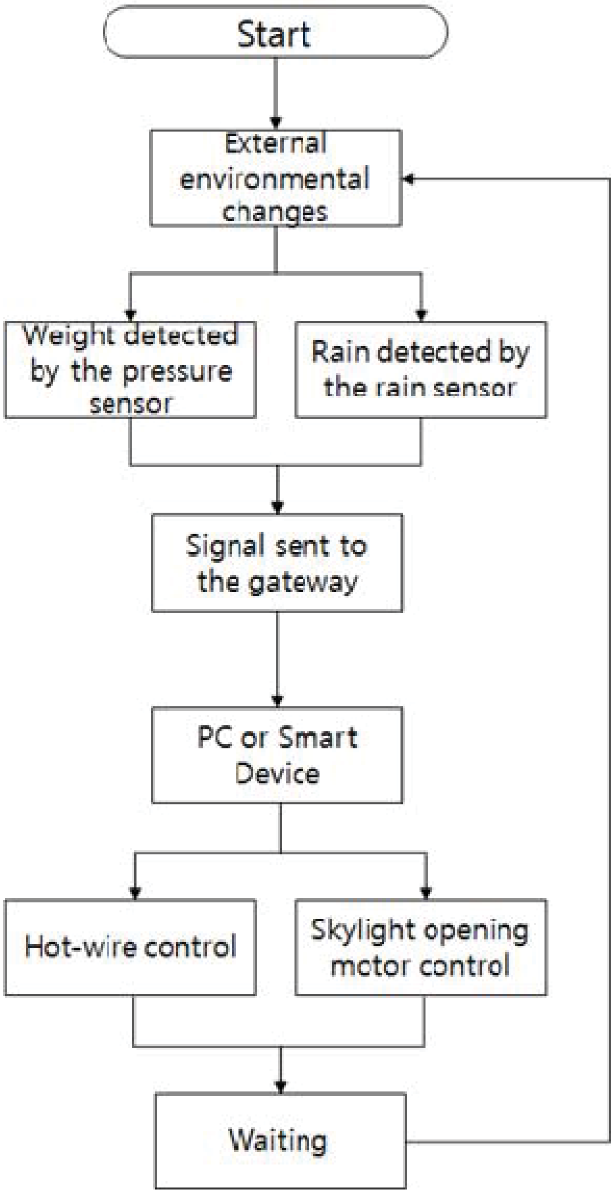
[Greenhouse] Acton to Motor 30 % / AM 10:00
AM 10:00 16 March

[Greenhouse] Action to Hot Wire Off / AM 10:00
AM 10:00 16 March

Fig. 8 Android Message(example)

3.4 System Flowchart

Fig. 9 shows a flow chart of operation in case the outside environment changes. In case of a change in the outside environment such as a tremendous snowfall or rainfall, this is detected by each sensor (pressure sensor, rain sensor), and information about this is transmitted through a gateway to a controller. The controller controls a hot wire or motor for opening and closing the ceiling through the transmitted information.



Start
External environmental changes
Weight detected by the pressure sensor
Rain detected by the rain sensor
Signal sent to the gateway
PC or Smart Device
Hot-wire control
Skylight opening motor control
Waiting

Fig. 9 flowchart of operation in case the outside environment changes

4 Conclusions

This paper designed an automation system on the ceiling of greenhouse by using a sensor, and tested the system through an arbitrary value. A signal making a hot wire work was generated by a controller approximately 3 seconds after the incoming value of snow depth producing an effect on the durability of facilities was changed to 23mm. Besides, it was possible to find that a signal making the hot wire stop working was generated when snow depth went down to 5mm or less. And when the value of rain was set for 3mm and that of wind speed was set for 13.5m/s so as to carry out a test in case of rain, it was possible to find that a signal making the ceiling closed was generated in order to prevent the ceiling from being damaged and rainwater from coming in. The system suggested by this paper can reduce cost necessary for restoration through reduction in damage to facilities by rapidly providing measures to cope with disasters when the manager of greenhouse cannot manage the greenhouse. Besides, management effective in the growth of crops is provided by preventing rainwater from coming into a greenhouse. As future research, we plan to check how widely the ceiling is opened and closed according to wind speed and whether a motor and hot wire is controlled after establishing a test bed, and plan to check the results after applying this system to a real farm.

5 Acknowledgment

"This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the the C-ITRC(Convergence Information Technology Research Center) (IITP-2015-H8601-15-1007) supervised by the IITP(Institute for Information & communications Technology Promotion)"

6 References

[1] In Ho Yu, Eung Ho Lee, Myeong Whan Cho, Hee Ryong Ryu, and Young Chul Kim, “Development of Multi-span Plastic Greenhouse for Tomato Cultivation”, Journal of Bio-Environment Control, 21(4):428-436, (2012)

[2] Kyung-Hwan Yeo, In-ho Yu, Han-Cheol Rhee, Jae-Woan Cheong, and Gyeong Lee Choi, “Field Survey on Roof Ventilation System for Single Span Greenhouse in Cultivation of Several Horticultural Crops”, Kor. J. Hort. Sci. Technol. 31 (SUPPL. II), pp198, October 2013

[3] Jong-Won Lee, “Analysis of Safety Wind Speed and Snow Depth for Single-Span Plastic Greenhouse according to Growing Crops”, Current Research on Agriculture and Life Sciences (2013) 31(4) : 280-285

[4] Michael A. Fonseca, Member, IEEE, Jennifer M. English, Martin von Arx, and Mark G. Allen, Member, IEEE , “Wireless micromachined ceramic pressure sensor for high-temperature applications”, Microelectromechanical Systems, Journal of (Volume:11 , Issue: 4): 337 - 343

[5] Don C. Abeysinghe, Samhita Dasgupta, Joseph T. Boyd, and Howard E. Jackson, “A Novel MEMS Pressure Sensor Fabricated on an Optical Fiber”, Photonics Technology Letters, IEEE (Volume:13 , Issue: 9) : 993-995

[6] Isabelle Bord , Pascal Tardy, Francis Menil, “Influence of the electrodes configuration on a differential capacitive rain sensor performances”, Sensors and Actuators B: Chemical Volume 114, Issue 2, 26 April 2006, Pages 640–645

[7] Lawrence A Klein, “Sensor and Data Fusion: A Tool for Information Assessment and Decision Making” 2004

[8] Telecommunincations Technology Association, “Greenhouse Control System - Part 1:Interface for Between Sensor Nodes and Greenhouse Control Gateway”, 2012.06.12

[9] Wikipedia - The Free Encyclopedia, “Mesh networking”, http://en.wikipedia.org/wiki/Mesh_networking, 2015.03.15

[10] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, “Wireless sensor network survey”, Computer Networks 52 (2008) : 2292-2330

COMPUTER SIMULATION OF A TAPERED HELICAL ANTENNA FOR WIRELESS COMMUNICATION APPLICATIONS

Obidiwe Tochukwu., Nweke Chisom B., Chineke Amechi., Nwachukwu-Nwokeafor Kenneth.

Department Of Computer Engineering Michael Okpara University of Agriculture, Umudike, Abia State Nigeria

ABSTRACT-*Wireless communications presents the platform for remote communication amongst users. It commands applications within spheres such as wireless local area networks (WLANS), Blue tooth (WI-FI, WIMAX), radio frequency identification (RFID), etc, It has subsequently penetrated the consumer electronics market such that they are now the status quo in several industrial, domestic, military as well as aviation oriented wireless communication applications. Consequently, the field of antenna engineering has witnessed a continuous refinement in both theory and practice (Since antennas constitute an integral part of wireless communication systems); this is necessary if the stringent requirements demanded by these 21*st *century cutting edge wireless technologies must be attained. Thus, this work is focused on the simulation of a broadband linearly tapered (conical) helical antenna for wireless local area network applications. Modeling of the antenna behavior and operation are achieved through the use of the Electric Field Integral equation (EFIE) in conjunction with the Method of Moments (MOM). The computer simulation of the antenna with respect to this model is achieved using several efficient MATLAB scripts based on the so called Rao-Wilton-Glisson (RWG) surface patch modeling scheme/strip model and basis functions. The work finally shows that this broadband antenna exhibited suitable characteristics (unidirectional radiation and purity of circular polarization) confined within a fairly broad frequency sweep (900MHz-3.5GHz) and optimal at the design frequency*.

Key words: RWG basis functions, Moment Method, Integral equations, helical antennas, wireless communications.

1.0 Introduction

WLANS and many other wireless technologies usually operate in the microwave region of the electromagnetic spectrum, mostly in the range of 900MHz-2.4GHz as specified by the I.E.E.E 802.11 standard for WLAN applications. The design of an antenna that can effectively radiate and receive signals for such applications basically involves the determination of the pertinent operational dimensions vis-à-vis parameters of the antenna (including feed design [1]) with respect to a particular frequency within the microwave regime.

Hence the computer simulation of the antenna will basically involve the determination of the radiation characteristics with respect to the pertinent parameters which include the turn circumference, turn spacing and total wire length for a conical helix antenna situated on an infinite ground plane[1,2], at an operational frequency of 2.1GHz and an impressed Delta-Gap voltage of 1Volt.The motivation for the design of a helical antenna for WLANS lies in the fact that the helical beam antenna (invented in 1946 by John D.Krauss [2]) and its modified versions (the tapered monofilar and multifilar axial mode helical antennas [7-10]) have proved to be the most acceptable and indispensable set of antenna's in space and terrestrial microwave satellite communications.

heir elliptical polarization property are utilized extensively in space probes and ballistic missiles to transmit and receive signals that have undergone Faraday rotation in the ionosphere [1].

2.0 The Problem Geometry and Edge Elements

Fig.1 below shows the general structure of a tapered helical antenna on a finite ground plane (a monopole helix) with the basic dimensions which are germane to the analysis and design of a conical helix antenna. Fig.2 shows the RWG edge element [3] equivalent of a conical helix antenna with a continuously varying radius 'a' on an infinite ground plane.

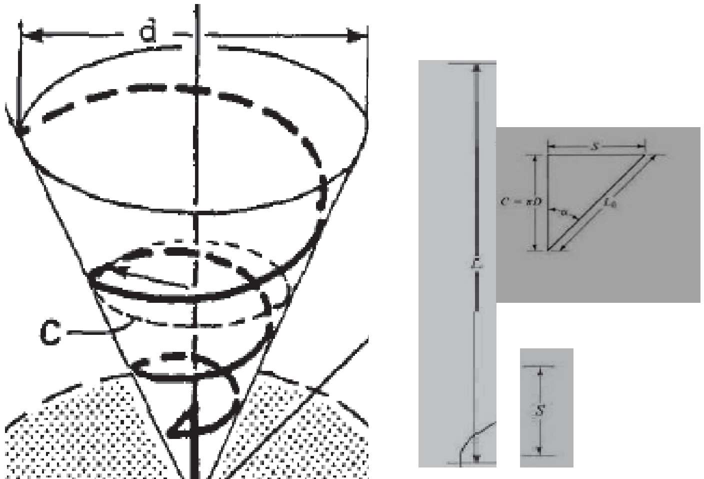


Fig.1: Tapered helical antenna.

In particular, these antennas due to t

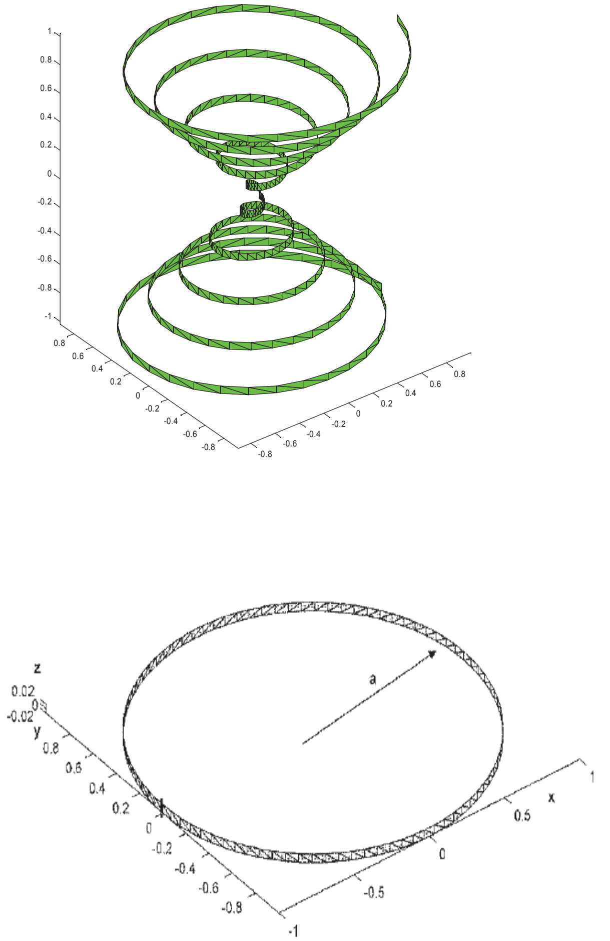


Fig.2: RWG edge element equivalent of tapered helical antenna with continuously varying radius 'a' [4].

THE HELIX DIMENSIONS ARE STATED BELOW [1, 2], where;

d = diameter of cone on which the wire is wound

C = circumference of cone on which the wire is wound

D = diameter of one turn (center to center)

c = circumference of one turn = πD

S = spacing between turns (spacing between turns)

α = pitch angle = $\tan^{-1}$ (S / πD)

L_0 = Length of one turn

N = number of turns

L = total length of antenna/axial length =NS.

The RWG edge elements constitute a subclass of the surface patch modeling technique in which an arbitrary surface is partitioned into triangular piecewise smooth patches. Each pair of triangles attached to any non boundary (i.e. interior edges) edge of the surface constitutes the so called edge elements. Fig.3 below shows two such triangles T_n^+ and T_n^- corresponding to the nth edge of a triangulated surface.

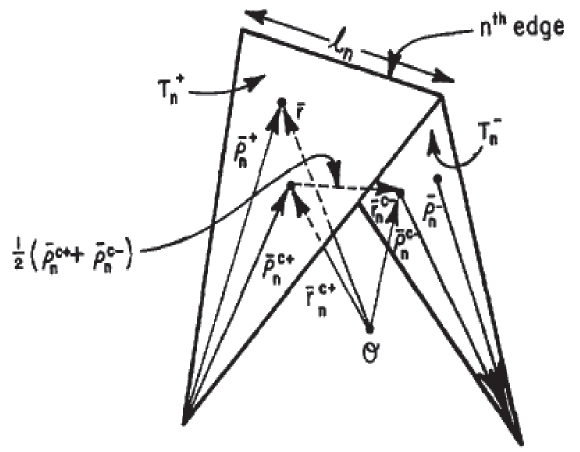


Fig.3: Triangle pair and geometrical parameters associated with interior edge [1].

Points in T_n^+ may be designated either by the position vector $\vec{r}$ defined with respect to point O or by the position vector $\vec{\rho}_n^+$ defined with respect to the free vertex of T_n^+.Similar remarks apply to the position vector $\vec{\rho}_n^-$ except that it is directed towards the free vertex of the triangle T_n^-. The plus or minus designation of the triangles is determined by the choice of a positive current reference direction for the nth edge, the reference of which is assumed to be from T_n^+ to T_n^- [3].

Most importantly, a basis/expansion function is defined for an interior edge of the patch model such that it vanishes everywhere on the arbitrary surface except in the two triangles attached to that edge. The basis function associated with the nth edge is defined as [3-6]:

$$\vec{f}_{\mathrm{n}}(\vec{r}) = \begin{cases} {l_n}/{2A_n^+} \quad \vec{\rho}_n^+ & \vec{r} \ in \ T_n^+ \\ {l_n}/{2A_n^-} \quad \vec{\rho}_n^- & \vec{r} \ in \ T_n^- \\ 0 & otherwise \end{cases}$$

………………………….. (1)

Here ℓ_n is the length of the nth edge, $A_n^+ and A_n^-$ are the areas of triangles T_n^+ and T_n^- respectively. With reference to [1], the subscripts refer to edges while superscripts refer to faces. The basis function $\vec{f}_n$ is used to approximately represent the surface current. Certain properties which make the basis function in (1.0) uniquely suited to its role were enumerated in [3].

In Fig.4 below, the normal component of $\vec{\rho}_n^{\,+} / \vec{\rho}_n^{\,-}$ along edge n is just the height of triangle T_n^+ / T_n^- with edge n as the base and the height expressed as $2A_n^+/$ $\ell_n or 2A_n^- / \ell_n$.This latter factor normalizes $\vec{f}_n$ such that its flux density normal to edge n is unity, ensuring continuity of current normal to the edge; this result together with the two potential expression for the electric field implies that all edges of T_n^+ and T_n^- are free of line charges [3,4].

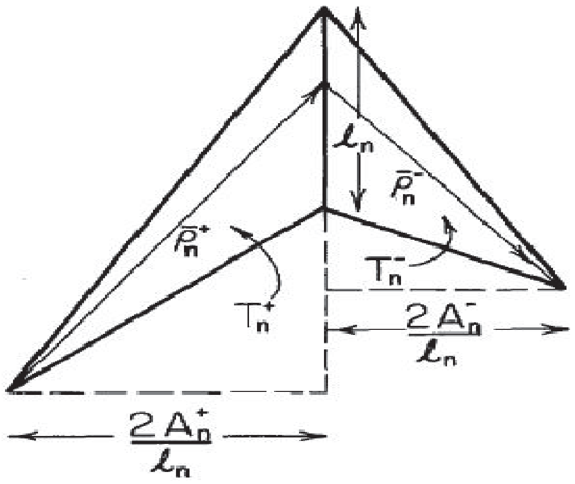


Fig.4: Geometry for construction of component of basis function normal to edge [3].

3.0 Materials and Methods

In order to simulate the radiation phenomena, short MATLAB scripts/surface patch codes [4, 5] based on the integral equation-moment method technique [8,11] was utilized. The surface patch codes were implemented and executed on a MATLAB 7 software package running on a windows 7 operating system environment. The source code can be classified into two types; one tagged with the identity/source file name 'rwg' and the other denoted as 'efield'. The codes bearing the identity 'rwg', deals with the physical structure of the antenna while codes bearing 'efield' deals with the radiation characteristics. Two main subroutines 'impmet'(which executes the impedance matrix computation) and 'point' (concerned with the antennas radiated fields), run as the core of rwg and efield respectively.

These source codes are stored in the MATLAB environment as the scripts rwg.m, efield.m, Impmet.m and point.m. The figure below shows the algorithm for the system; the first step in executing the algorithm involved creating the antenna surface patch model which was shown in fig.2 using surface patch codes with the file source name 'mesh'. Once the antenna mesh is fed into the file 'rwg.1' the code sequence of Fig.5 can be executed.

4.0 System Implementation

Fig. 5 reveals that the first step in the simulation process involves generating the antenna mesh; before the mesh can be generated, the helix strip dimensions must be fed into the mesh generator script. In this section, an attempt was made to provide the approximate dimensions for a 7½ turn conical helix antenna on infinite ground plane with respect to the axial mode of operation.

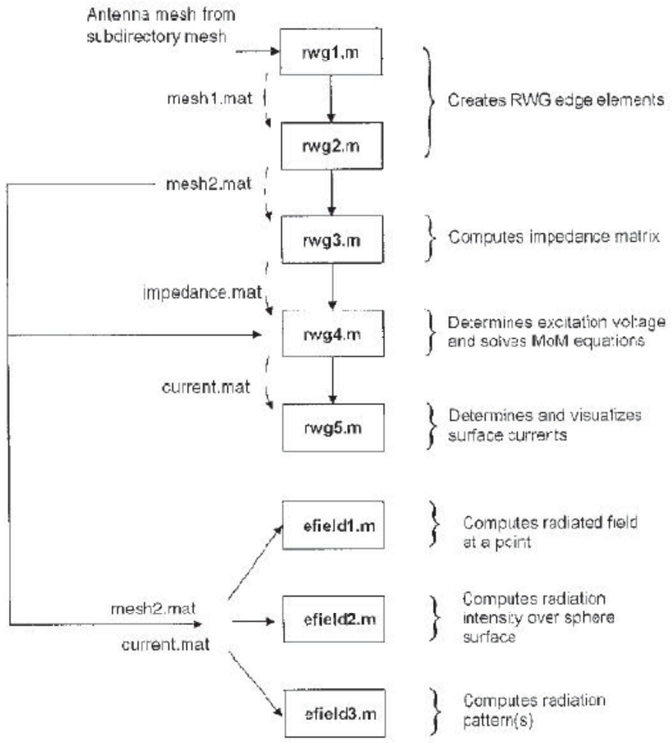

Antenna mesh from subdirectory mesh
rwg1.m
mesh1.mat
rwg2.m
Creates RWG edge elements
mesh2.mat
rwg3.m
Computes impedance matrix
impedance.mat
rwg4.m
Determines excitation voltage and solves MoM equations
current.mat
rwg5.m
Determines and visualizes surface currents
efield1.m
Computes radiated field at a point
mesh2.mat
current.mat
efield2.m
Computes radiation intensity over sphere surface
efield3.m
Computes radiation pattern(s)

Fig.5: Code Sequence [4].

The speed of the signal radiated by the antenna is given as:

$c = f\lambda$ …………………………….. (2)

Where c= 3.0x10^{8}m/s and $f = 2.1GHz$.

Therefore λ=$\frac{3.0\times10^8}{2.1\times10^9} \approx 0.143m$.

The helix dimensions is determined from the range 3/4<C/λ<4/3, where C is the turn circumference. Circular polarization of the main lobe is achieved by setting C ≈ λ and S ≈ λ/4, [1, 2, 4], where S is the helical antenna turn spacing.

In general, C and S are chosen as large fractions of the wavelength for the generation of the axial mode of radiation [1,2]. Running the scripts in fig.5 within the specified circumference to wavelength ratio reveals that an optimal gain value of about 9.2 decibel is obtained at a circumference of 0.92λ.

Thus C can be computed as follows: C= 0.92λ, since the

Value of λ obtained at model frequency = 0.143

Then C= 0.92x0.143 ≈ 0.132m.

Hence, the top/bottom turn radius with respect to fig.2 will be a_{max} ≈ 0.02m; if a_{min} or center turn radius is assumed to be 1/10th of a_{max} then a_{min} ≈0.002m.

In addition, running the scripts of fig.5 with approximate turn spacing values of about λ/4 also revealed that a turn spacing of 0.03m was largely responsible for the optimal gain of 9.2 decibel.

Thus, the turn spacing can be computed as S = $\frac{0.143m}{4.4}$

≈ 0.03m.

Assuming that the helical windings are made from an AWG wire of diameter=1.6mm,

The strip width can be determined from [4] as,

$$a_{eqv} = 0.25\,h \text{ ....(3)} \; where\; h \;\; is\; the\;\; strip\;\; width$$

Thus: h = 4x 0.0008m ≈ 0.003m.

It was observed that slight variations in the strip width produced very minute or no significant change in the gain value and radiation pattern. Now that the strip dimensions have been determined, the Delta gap excitation voltage of 1Volt was impressed on the antenna structure. Observance of the code sequence in Fig.5 reveals that the script rwg4.m is responsible for the determination of the excitation voltage and the computation of the input impedance. In this script, the source is conventionally placed closest to the point (a min, 0, 0) [4, 5]. In the present work, the feed voltage will be placed at the point

(-1, 0, 0).

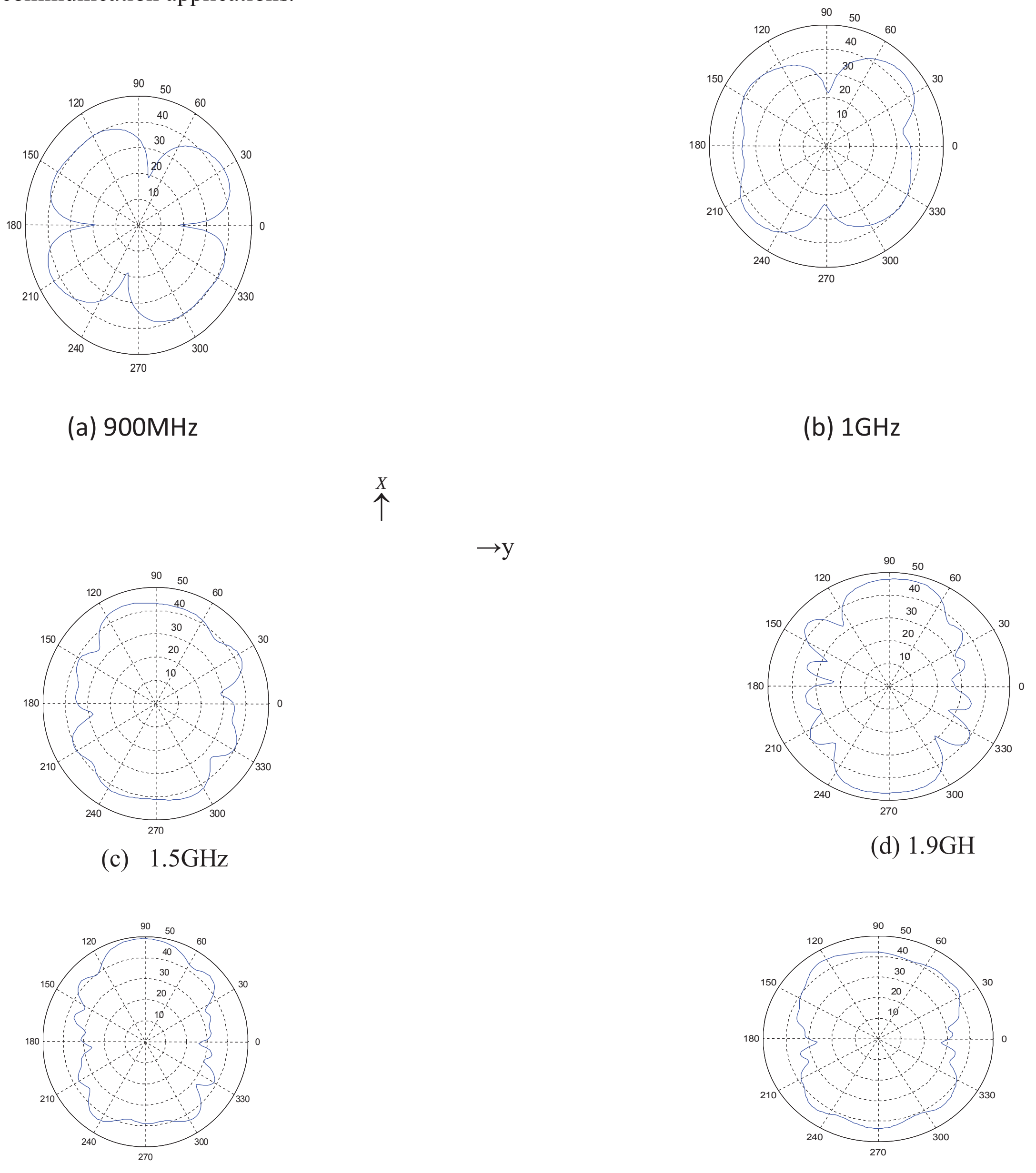
5.0 Testing and Simulation Results

The antenna radiation pattern or directivity plot at a single frequency is obtained by running the script efield3 in Fig.5. The script yields 2D radiation patterns in the xy, xz and yz planes. The radiation pattern at different frequencies within the specified bandwidth is displayed below in Fig.6.

From the 2D directivity patterns displayed below, Fig.6, it can be observed that at 0.9GHz the pattern has its major lobes almost evenly distributed in the four quadrants; this feature almost approximates to an omni-directional mode of radiation. At 1GHz, the axial mode of radiation is not yet well formed; the pattern transformation seems closer to the normal mode of radiation. The 1.5GHz plot shows an axial mode of radiation but the end fire character is not yet optimal. At 1.9GHz, the axial mode of radiation is well formed but a considerable amount of power is spent in the side or minor lobes. The 2.1GHz plot shows an optimal end fire radiation and minimized side lobes relative to 1.9GHz. The 3.0 and 3.5GHz plots depict an axial mode of radiation with reduced end fire directivity relative to the 2.1GHz plot.

6.0 Conclusion

An attempt has been made in the present work to simulate a conical helix antenna over an infinite ground plane in terms of a thin strip model (RWG edge elements); the results obtained (radiation patterns) are encouraging and show that the RWG edge elements surface patch modeling technique can be applied successfully in the design of a conical helix antenna for wireless communication applications.



(a) 900MHz

(b) 1GHz

(c) 1.5GHz

(d) 1.9GH

(e) 2.1GHz

(f) 3GHz

Figure6:Directivity Patterns in the xy-plane at (a) 900MHz, (b) 1GHz, (c) 1.5GHz (d) 1.9GHz, (e) 2.1GHz, and (f) 3GHz.

REFERENCES

1. C. A. Balanis. Antenna Theory: Analysis and Design, 2nd ed. Wiley, New York, 1997.
2. J. D. Kraus, and Ronald J. Marhefka. Antennas for all Applications, (3rd. ed.), McGraw-Hill, New York 2002.
3. S. M. Rao, D. R. Wilton, and A. W. Glisson. Electromagnetic Scattering by Surfaces of Arbitrary Shape. IEEE Trans. Antennas and Propagation, 30 (3): 409-418, 1982.
4. S. Makarov. Antenna and EM Modeling with MATLAB. Wiley, New York, 2002.
5. A. Apte. Simulation of Patch Antennas on Arbitrary Dielectric Subsrates- RWG basis functions. M.Sc Thesis, Worcester Polytechnic Institute, May 2003.
6. S.M.Rao and T.K. Sakar. Numerical Solution of Time Domain Integral Equations for Arbitrarily Shaped Conductor/ Dielectric Composite Bodies. IEEE Trans. Antennas and Propagation, 50 (12): 1831-1837.
7. S.R. Saunders and A.A. Zavala. Antennas and Propagation for Wireless Communication Systems. 2nd ed, Wiley, England, 2007.
8. D.G. Fang. Antenna Theory and Microstrip Antennas. CRC Press, New York, 2010.
9. S. Koziel, and S.Ogurstov. Antenna Design by Simulation-Driven Optimization. Springer, New York, 2014.
10. F.B. Gross (Ed). Frontiers in Antenna: Next Generation Design and Engineering. McGraw Hill, New York, 2011.
11. J. Van Bladel. Electromagnetic Fields. 2nd ed, IEEE Press, 2007.
12. Y.H. Lee. Introduction to Engineering Electromagnetics. Springer, New York, 2013.

Design of the Pig Management Application

Hoseok Jeong[1], **Misuk Kim**[2], **Hyun Yoe**[*]

[1, *] Department of Information and Communication Engineering, Sunchon National University, Suncheon-si, Jeollanam-do, Republic of Korea

[2] Agrifood Convergence ICT Research Center, Sunchon National University, Suncheon-si, Jeollanam-do, Republic of Korea

Abstract - *Korea hog industry has grown with the biggest product scale. However, the current circumstance is that it is lagging behind in market share by imported livestock products as the market has opened as a result of the recent FTA signing, and it is experiencing difficulties due to increased production cost that resulted from the increase in international grain price. In addition, productivity of farms is not being improved due to low level management technology. To solve such problem, this paper proposed IT-based hog breeding management application. The proposed application has been developed at the eye level of farms for conveniently and easily managing the items required by hog farms. In this application, items were determined by focusing on the sow, market pigs and feed feeding directly connected to the income of hog farms, and shipment prediction item was inserted to allow users to pay attention to business management by being aware of income for each month. In addition, feed cost that currently occupies 60~70% of production cost could be reduced through feed feeding management. Based on this, it is expected to improve the productivity of hog farm and increase the competitiveness of domestic livestock products through systematic breeding management.*

Keywords: IT, Application, Breeding, Pig, Management

1 Introduction

Korea hog industry has been continuously growing by becoming a main protein supply source of people[1]. In addition, it has grown into most competitive industry with largest production scale in livestock industry next to rice in agriculture[2].

In spite of this, hog industry of our country is experiencing many difficulties domestically and internationally[3]. First off, the market share is lagging behind due to the incoming of cheap imported livestock products. And 60~70% of business cost is being spent for feed due to the increase in feed price as a result of the increased international grain price[4][5].

In addition, productivity of farms is low due to insufficient level of facilities and breeding management technology compared to that of livestock advanced countries[6].

To overcome such situation and increase the safety and profitability of farms, there is a need to improve productivity and reduce production cost. For the purpose of improving productivity, it is necessary develop hog breeding management program that is easy to use by farms. Hog management programs have been developed domestically by integrating IT but they were very inconvenient to use by actual hog farms as their IT part has been emphasized. Accordingly, they should be developed as the eye level of farms to allow them to conveniently and easily use while managing the items that hog farms actually need.

This paper proposed IT based hog breeding management application according to such situation. Through the proposed application, sow and market pigs information can be seen via smart phone, as well as the market pig information of farms per month through shipment forecast. In addition, it can reduce the loss in feed by setting food intake amount according to each entity. Based on this, it is expected to provide convenience to hog farms and reduce production cost.

2 Related research

2.1 Sow Breeding Management

The first goal of a pig farm is to produce a large number of healthy pigs. To do so requires a systematic management of sows breeding, and this breeding management is directly related to farm productivity[7]. High productivity and nursing ability of sow is made through the systematic management of breeding[8]. In this paper, I assumed that the breeding environment is same, and focus on the breeding date and farrowing date.

2.2 Piglet Breeding Management

Subcutaneous fat in piglets are physiologically less dense and they have weak temperature control capability for the external environment. And do not have immunity against the disease. Also should be keep optimum environment always because piglets is exposed to digestive and respiratory disease.

[*] Corresponding author

Especially when temperature is falling because of wide daily temperature range, piglets are very dangerous. Thus, the weight is light and weak piglets must be managed carefully according to temperature and humidity shown in table 1[9].

Table 1. Date or weight condition

Date or weight	Titration temperature range(℃)	Optimum temperature(℃)	Humidity(%)
Immediately after birth	30 ~ 35	35	60 ~ 70
1 week	25 ~ 30	25	60 ~ 70
Before weaning	20 ~ 25	20 ~ 25	60 ~ 80
Immediately after birth	30 ~ 35	35	60 ~ 70
Weaning	25 ~ 30	25	60 ~ 80
Weaning ~ 45kg	18 ~ 22	21	50 ~ 80
45kg ~	15 ~ 20	18	40 ~ 60

Growth of piglets change according to nursing ability of sow and breeding environment. In particular, after birth piglets must manage to intake of colostrum. Because the creation of immunity antibody is to provide from colostrum. In addition, sanitary management of hog farm is very important. If farm is dirty or the humidity is high, piglets increased the risk about disease.

2.3 Market pig Breeding Management

The immunity system of finishing pigs has been strengthened than that of piglets. However, those should be well raised and should be shipped in accordance with shipping schedule. Therefore, special care is necessary. Because the shipping of finishing pigs has a great effect on the monthly management of farm, it should be systematically managed. The environmental factors that have an effect on the productivity and health of pig include temperature, humidity, ventilation, wind velocity, and stocking density, and so on. And the factors that have the greatest effect on the productivity and health of pig at the turning of seasons are temperature and ventilation among these surroundings. Generally, sufficient ventilation should be possible besides temperature in consideration of a point that the maintaining of temperature inside the pigpen is greatly influenced according to the outdoor temperature and facility environment of each farm. And in case of finishing pigs, appropriate temperature of 17 ~ 20 ° C is set up, and then they should decide whether temperature and ventilation volume is set again and whether the pigpen is heated and should check whether something is wrong with machine on occasion by surely measuring the concentration of ammonia gas and carbon dioxide gas. If a draft comes in the pigpen at the turning of seasons, and pigs are exposed to the condition of extreme temperature variations between day and night, a great effect is produced on the feed intake, and moreover, productivity is decreased. Generally in case of finishing pigs, if temperature is 1°C below the appropriate temperature, one pig takes more feed by 25 g a day. On the contrary, if the remaining feed is not taken, weight gain is decreased by 11g. And a major contributor to the decrease of pig productivity is a respiratory disease under the confined environmental condition. And the respiratory disease of pigs may generally occur when temperature is low or diurnal temperature variation is large. Generally in case of growing pigs, if temperature inside the pigpen is 5°C and below, or diurnal temperature variation is 10°C and above, the incidence of respiratory disease severely increases. Together with this ambient temperature, what is important is sensible temperature that pigs feel. And if the skin of pigs is directly exposed to the cold wind, pigs have more stressed than that of low temperature in the pigpen. Therefore, pigpen management requires careful attention at the turning of seasons[10].

2.4 Korea Smart Phone Subscriber Present Condition

The number of smart phone terminal users domestically is over 36 million people as of August 2013, as shown in Table 1. It is continuously rising as about 500,000 users as of 2009 increased exponentially to 20 million users as of October 2011[11]. In developing and applying livestock raising IT convergence technology using mobile application, it is very effective with many advantages in our country since there are many users. According to such domestic condition, IT convergence technologies are being developed by using smart phone application in various areas.

Table 2. Korea smart phone subscriber present condition (Unit : Person)

Division	2011.12	2012.12	2013.8
SKT	11,085,192	15,978,717	17,756,369
KT	7,653,303	10,250,998	11,012,233
LGU+	3,839,913	6,497,534	7,552,372
Total	22,578,408	32,727,249	36,320,974

3 Application Design

3.1 Application Configuration

The hog breeding management application proposed in this paper has been composed as shown in Figure 1. First off, hog information entered by user is saved in Database to distinguish between sow and market pigs.

In the case of sow, entering crossbreeding date will automatically calculate and display farrowing date. In the case of fertilization failure, it has been developed to display crossbreeding once again. As for farrowing date, it was set as 115 days after crossbreeding upon inquiring Korea Pork Producers Association and calendars currently being used by hog farms.

In the case of market pigs, shipment forecast has been set as 24 weeks after weaning. Market pigs are shipped when they are about 110~115kg and when they pass this range, their price decreases thereby affecting the income of farms. Accordingly, they should be shipping according to the condition of market pigs by conducting feeding management for 24 weeks.

The shipment forecast allows users to know market pigs per month according to the date of shipment. Shipment forecast allows users to manage their business to become aware of their farm fund that is sufficient or lacking in a particular month for advanced preparation by manager. In addition, feed loss can be reduced through feeding amount adjustment menu to adjust the feed amount of sow appropriately.

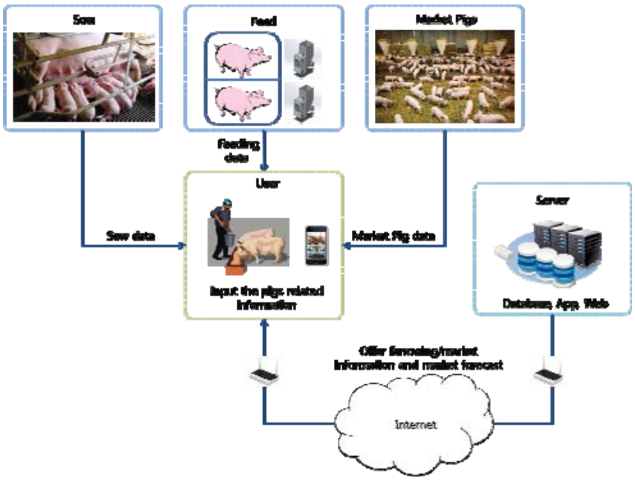


FIgure 1. Configuration of the hog breeding management application

The number of smart phone terminal users domestically is over 36 million people as of August 2013, as shown in Table 1. It is continuously rising as about 500,000 users as of 2009 increased exponentially to 20 million users as of October 2011[11]. In developing and applying livestock raising IT convergence technology using mobile application, it is very effective with many advantages in our country since there are many users. According to such domestic condition, IT convergence technologies are being developed by using smart phone application in various areas.

3.2 Service Process

The proposed hog breeding management application operated through the process shown in Figure 2. First off, when user enters data information on the pigs at pigsty through application, it is stored in database. Stored data is distinguished in server on its type and calculated according to situation to display necessary information.

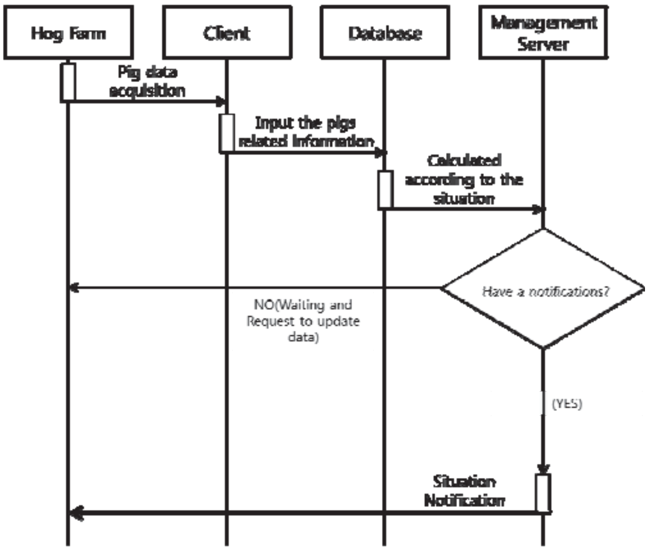

Hog Farm
Client
Database
Management Server
Pig data acquisition
Input the pigs related information
Calculated according to the situation
Have a notifications?
NO(Waiting and Request to update data)
(YES)
Situation Notification

FIgure 2. Process of hog breeding management application

In the case of data on crossbreeding date of sow, farrowing date is calculated and shown and in the case of data on piglet after weaning, shipment forecast date is calculated and shown. When time passes and farrowing date or shipment forecast date approaches, it notifies user to take necessary action.

In the case of data on the feed feeding amount of sow, it allows user to check the date to reduce feed loss by adjusting feeding amount according the intake amount of sow.

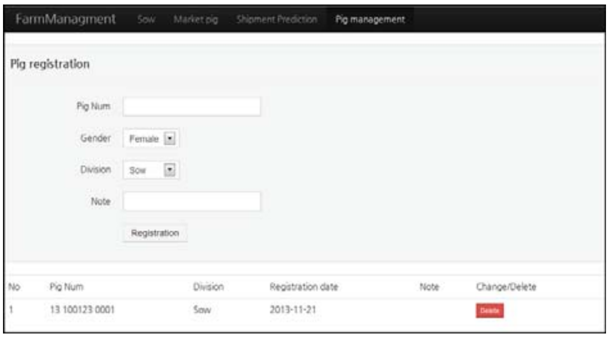
4 Implementation and Result

Specifications of system used to collect pig data in the proposed system are as shown in Table 2.

Table 3. Development environment

Type	Version
Windows 7	Ultimate K 32 bit
Java	JDK 1.7.0.2
Eclipse	Kepler
Android	4.1.2 Jelly Bean
Tomcat	7.0.42
Database	Mysql 5.6
VM ware	5.0.2
Smart Phone	LG optimus G pro

Figure 3 is the hog registration screen of the hog breeding management application that has been developed. The hog registration consists of entity number, gender, hog classification and feeding amount, and user needs to directly enter the initial data. Entered data is stored in Database server and the registered content is shown at bottom of screen that can be revised and deleted.



Pig registration

Pig Num

Gender Female

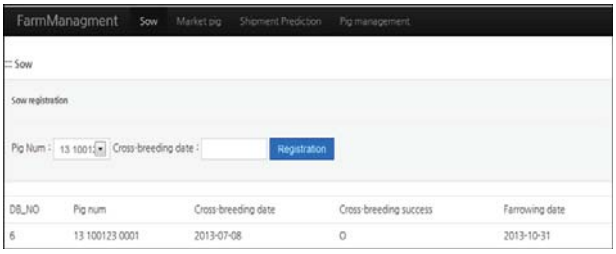
Division Sow

Note

No	Pig Num	Division	Registration date	Note	Change/Delete
1	13 100123 0001	Sow	2013-11-21		Delete

FIgure 3. Hog registration screen of application

Figure 4 is a screen of sow information. The success or failure of crossbreeding can be seen by selecting the registered sow order number, entity number, feeding amount, expected date of crossbreeding and crossbreeding status after crossbreeding, and it notifies expected farrowing date based on the registered expected date of crossbreeding.



:: Sow

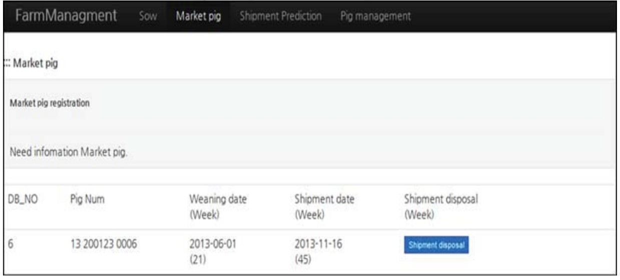
Sow registration

Pig Num : 13 1001 Cross-breeding date :

DB_NO	Pig num	Cross-breeding date	Cross-breeding success	Farrowing date
6	13 100123 0001	2013-07-08	O	2013-10-31

FIgure 4. Sow information screen of application

Figure 5 is a screen of market pig information. Registered market pig's order number, entity number, expected date of weaning, expected date of shipment and feeding amount can be seen, and it notifies expected date of shipment and weekly unit based on registered expected date of weaning. Market pig that has undergone shipment processing can be deleted from the screen using shipment completion button.



:: Market pig

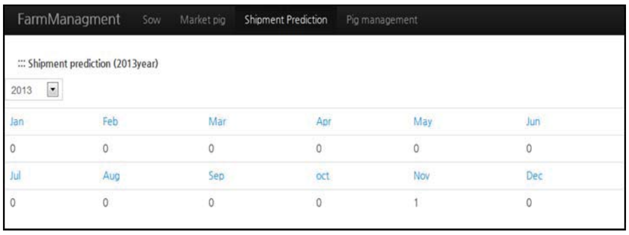
Market pig registration

Need infomation Market pig.

DB_NO	Pig Num	Weaning date (Week)	Shipment date (Week)	Shipment disposal (Week)
6	13 200123 0006	2013-06-01 (21)	2013-11-16 (45)	Shipment disposal

FIgure 5. Market pig information screen of application

Figure 6 is a screen of shipment management. It shows the number of hogs being shipped each month by calculating expected shipment date based on the registered market pig information. Based on this, user can manage the business of farm based on the number of market pigs per month.



::: Shipment prediction (2013year)

2013

Jan	Feb	Mar	Apr	May	Jun
0	0	0	0	0	0
Jul	Aug	Sep	oct	Nov	Dec
0	0	0	0	1	0

FIgure 6. Market pig information screen of application

5 Conclusions

This paper proposed IT based hog breeding management application. The purpose of this paper was to allow user to use breeding management that is currently being written manually with application in a convenient way. In addition, application items were selected based on the fact that the most important area of the proposed application is to increase the productivity and reduce production cost of hog farms.

To verify hog breeding management application, test was conducted at actual hog farm. The rest result showed that the manger of hog farm was able to easily adapt to it because it was much more convenient to use than the existing breeding management program. In addition, it was also very helpful in business management by allowing user to identify the number of hogs being shipped each month upon adding shipment forecast function.

The hog breeding management application can meet the needs of farms, improve the productivity of hog farms and provide convenience, and it is expected to reinforce the national competitiveness of domestic livestock products through systematic breeding management.

5 Acknowledgment

"This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the the C-ITRC(Convergence Information Technology Research Center) (IITP-2015-H8601-15-1007) supervised by the IITP(Institute for Information & communications Technology Promotion)"

6 References

[1] The Agriculture Fisheries & Livestock News, “2011-2012 Korea Livestock Year book”, 47—60, 2011

[2] Sangsoo Lee, “Cost savings through improved feed efficiency to elevate productivity”, Korea Pig&pork, 200—203, 2013

[3] Yudong Kim, “MSY 25 with realize the pig farm case management point”, Korea pork producers association, Vol.1, No. 7, 186—188, 2007

[4] Kwanhyun Yoo, “Design and Implementation of Hog-Raising Population Management System using RFID”, Incheon Natinal Univ. a master´s thesis, 13—24, 2010

[5] Jeongsoo Jeon, “A study on the productivity improvement method of a pig”, Bulletin of the animal biotechnology, Vol. 1, No. 6, 21—25, 2008

[6] Hwasoon Kang, “Analysis of Farm Management Factors affecting the MSY in Korea's Hog Industry”, KonKuk Univ. doctorate thesis, 20—23, 2010

[7] Marrit Van Engen, Kees Scheepens, “Sows is part of the Pig Signals”, 6—39, 2012

[8] Joonghwan Jeon, “Preview pig farm animal welfare certification standards and procedures”, Hyundai pig magazine, Vol. 295, 42—47, 2013

[9] Korea pork producers association, http://www.koreapork.or.kr

[10] Jan Hulsen, Kees Scheepens, “Pig Signals”, 26—93, 2012

[11] Statistics Korea, http://kostat.go.kr

A Novel Approach for Congestion Notification in Ethernet Networks

[1]D. Hema Latha, [2] D. Rama Krishna Reddy, [3]Azmath Mubeen, [4] K.Sudha

[1]Asst. Professor, Dept of Computer Science, Osmania University College for Women, Koti, Hyderabad,Telangana, India
[2]Asst. Professor in Computer Science, Dept of Mathematics, UCS, Osmania University, Hyderabad, Telangana, india
[3]Asst. Professor, Dept of Computer Science, University College for Women, OU, Koti, Hyderabad, Telangana, India
[4]Teacher in Computer Science, Disney Land High School, Hyderabad, Telangana, India

Abstract - *In present situation of data communication Datacenters are utilizing faster Ethernet and unlocking new capabilities with the pushing east-west bandwidth demand for big data analytics, social media, and other new applications. Ethernet uses optical fiber to interconnect fabric to connect servers and networking and storage equipment. After years of 1 G bit/s Ethernet (GbE) connections linking servers, datacenters are utilizing 10GbE and this is enhanced to 40GbE today. In near future real-world deployment of 100GbE is possible.*

Ethernet is replacing the traditional storage networking technologies like Fiber Channel and Infinite band in Datacenters. The main important characteristic of these traditional technologies that make them suitable for datacenter is their less-loss and low-delay operation. Consequently IEEE 802.1standards committee is developing new specification for congestion management for Ethernet networks. The two new approaches for congestion management for Ethernet networks are: Backward Congestion Notification (BCN) and Forward Explicit Congestion Notification (FECN). Each approach has its own advantages and disadvantages. FECN outperforms BCN in fairness and response time while BCN is able to respond to sudden increases in load in less than a round trip time. In this paper, the authors proposed an enhanced version of FECN that takes the best of both proposals and appends BCN if severe congestion arises suddenly. It is shown that E-FECN performs better than the previous proposals.

*Keywords: Congestion Control, Congestion Notification, **FECN, enhanced FECN, BCN, and DCN***

1 Introduction

Congestion control has traditionally been handled at the transport (TCP) layer [1]. TCP New Reno, TCP SACK, and TCP Vegas for wired networks, I-TCP, MTCP and Freeze TCP for wireless networks, TCP-F, ELFN, and ATCP for ad-hoc wireless networks, E-TCP, STCP, Fast TCP, and High Speed TCP for high-speed networks are examples of such proposals [2, 3, 4]. However, the storage traffic in datacenter networks does not use TCP. So, data link level algorithms have been developed which control all types of traffic regardless of the transport (or even network) protocol. Datacenter networks (DCNs) are used for data storage and file transfers. These applications require a high throughput and a low latency. The transmission speeds of links used in DCNs are 1-10 Gbps. To maintain a low latency and to avoid extensive large queuing delay, queue lengths should be kept low. Moreover, packet loss is unacceptable since every packet is critical and will need to be retransmitted resulting in unacceptable delays. Therefore, IEEE 802.1Qau group has been formed in IEEE 802.1 standards committee to develop new congestion control schemes specifically designed for datacenter networks. Backward Congestion Notification (BCN) and Forward Explicit Rate Notification (FECN) are two of the proposals made in IEEE 802.1Qau group. BCN has many variations with optional features such as BCNMAX, BCN00, and BCN with drifting and over-sampling. The main issues of BCN are its unfairness and large oscillations in throughputs. FECN, on the other hand, is fair but has a slower start. In this paper, we propose Enhanced Forward Explicit Rate Notification (E-EFCN) scheme, which is basically FECN mechanism with BCN sent back to the source under severe congestion. We show that E-FECN can achieve perfect fairness like FECN and allows a fast start.

2 Related Work

David Bergasamo and his colleagues at Cisco [5 and 6] has proposed BCN mechanism for Ethernet datacenter networks. BCN is a rate-based closed-loop feedback control mechanism. BCN system model is shown in fig.1. As illustrated in figure 1, at the sources, rate regulators are used to adjust the individual flow rate according to the BCN messages received from switches. The switches are referred as congestion points (CPs) while the sources are the reaction points (RPs). The buffer utilization is supervised by the switches and sends BCN messages back to the source based on the status and variation of the buffer queue. Two thresholds Qeq (equilibrium queue length) and Qsc (severe congestion queue length) are used to trigger BCN messages.

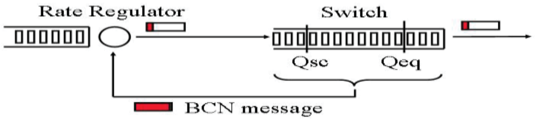


Figure 1: BCN System Model

Several variations of the basic BCN scheme have also been proposed, such as BCNMAX, BCN00, and BCN with drifting and over-sampling. The purpose of BCNMAX is that it prevents the system from severe congestion. When the queue length reaches *2×Qeq*, a BCNMAX message is sent back to the source. The source is expected to reduce the transmission rate substantially. This helps in avoiding severe congestion. Similarly, a BCN00 message from the switch causes the source to reduce to a very low rate. It is observed that, BCN is unfair in the sense that once a source reduces its transmission rate; it remains low. To avoid this problem, BCN with floating or drifting feature is introduced, which allows the sources to enhance their transmission rate randomly [7]. In order to keep the overhead low, it is necessary to keep the queue sampling rate low or at minimum. However, during severe congestion, it is necessary to react quickly, so an "over-sampling feature" was added to allow more frequent sampling during severe congestion. BCN is also known Ethernet Congestion Manager (ECM) [8]. It has been shown that BCN achieves only proportional fairness not max-min fairness. Furthermore, BCN is slow in convergence to fair state and has large oscillations in throughput. FECN was proposed to deal with the fairness and oscillation issues [9]. Figure 2 shows FECN, FECN is a close-loop explicit rate feedback control mechanism. The sources periodically send inquiry or probe messages that pass through the switches and returns back to the sources from the destination. On the forward path, the switches reduce the rate field in the probe and when the probes return to the sources, they contain the exact rate that the flow should follow. Therefore, the sources change their rate according the probes. The switches announce the same rate to all the flows passing through the switch. This ensures that all flows are treated equally and fairly.

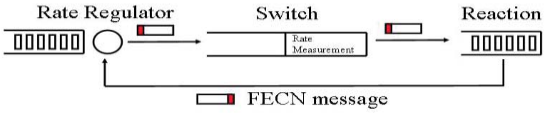


Figure 2: FECN System Model

Cyriel Minkenberg and Mitch Gusat introduced E2CM [10] probing idea of FECN to ECM along with a few other enhancements. For example, E2CM computes per flow loads at the source, determines forward latency with the help of source clock, performs continuous probing, and accelerates rate recovery. Technically, E2CM is much like ECM but with probing technique and also tuning up some parameters such as additive increase gain parameter and multiplicative decrease gain parameter. Figure 3 illustrates FECN and BECN mechanisms in a network.

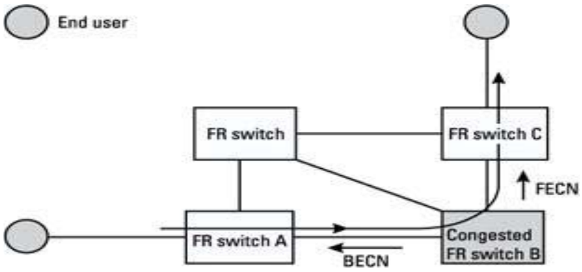


Figure 3. Network with FECN and BECN mechanisms

3 Comparison of BCN and FECN

BCN and FECN are designed to fix congestion problem in datacenter networks. Each has their own advantages and disadvantages. Table.1 shows some of the characteristics of BCN and FECN. For example, BCN messages can reduce the source rate quickly because the message is sent directly from the congestion point. On the other hand, with FECN, the sources need to wait for one round trip time in order for FECN probe message to come back. However FECN can reach the perfect fair state within a few round trip times because all sources get the same feedback. Unlike BCN, FECN does not have large oscillations in source throughput. The overhead of FECN is low and can be predicted because FECN message is sent every 1 ms and the message size is small with a payload of about 20 bytes.

From Table 1, it is seen that FECN outperforms BCN in several ways, but there are a few issues that still need to be considered, e.g., fast start, link disconnection, and the number of rate regulators. FECN was designed for congestion avoidance while BCN was designed for congestion control. In FECN, the sources start at a low rate and move to the equilibrium rates as successive probes return. Thus FECN starts each flow with a rate regulator. In BCN, the sources start at full rate and come down if a BCN is received from a switch. Thus, in some cases, they may not need the rate regulator at all. Secondly, if a link breaks, the FECN probes may not return and the source may continue to send at the current rate. BCN can issue a control message to the source in order to decrease or stop the source transmission. To deal with such situations, FECN has a timeout feature, which requires the sources to return to lower rate if the probes are not received. But, there is some packet loss. Finally, FECN, as originally proposed, requires as many regulators as the number of concurrent flows. This is because each flow starts in a regulated state. E2CM, an enhanced ECM, limits the number of rate regulators to equal to a number of congested flows. A variation of FECN in which the number of rate regulators is equal to the number of congestion point has also been introduced [11]. An internal mapping mechanism is needed to be implemented in the Networks Interface Card (NIC). It turns out that this mapping scheme can be applied to both BCN and FECN.

Table 1: A Comparison of BCN, FECN, and E-FECN

Parameters	**BCN**	**FECN**	**E-FECN**
Fairness	Unfair (better with drift)	Perfect	Perfect
Feedback Control	Backward	Forward	Forward with becon
Overhead	High (Unpredictable)	Low (predictable), 20bytes	Medium
Load Sensor	Queue based	Rate based	Rate + Queue based
Link Disconnection	Support	N/A	Support
Fast Start	Support	N/A	Support
Number of Rate Regulators	Variable (E2CM)	Fix (= number of source flows)	Variable

4 E-FECN : Enhanced Explicit Congestion Notification

Enhancement of FECN is described in this section which allows fast start and thereby reduces the number of rate regulators, and also allows FECN to cope with link disconnection. This enhanced version of FECN is called E-FECN.

E-FECN is shown in Figure 4, in this mechanism, in addition to the normal probing mechanism of FECN, the switches are permitted to send BCN messages under severe congestion. E-FECN allows sources to start sending the data at full rate (Fast Start) without a rate regulator. If this results in congestion at any switch, that switch sends a BCN00 message so that the source reduces to a low initial rate.

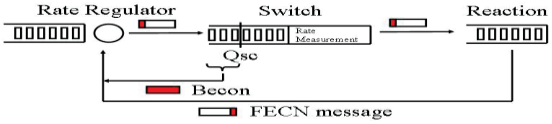


Figure 4: E-FECN System Model

The algorithm of E-FECN is as follows:
1. The process at the sources:
a. All flows start sending at the full rate. Similar to BCN, the sources start at full rate and come down due to the congestion if either a BCN or FECN is received. Thus, in some cases, they may not require rate regulator at all.
b. Some congested flow rates are limited by the rate regulators.
c. Flows are set to the proper rates once the source receives FECN control message, and it behaves the same as that in FECN mechanism.
2. At the switches:
a. Almost all E-FECN operations are same as in FECN.
b. The switches also monitor their queue length.
c. If the queue length exceeds the severe congestion threshold (*Qsc*), a BCN00 message is sent to the source. The source then reduces the rate to *Rmin* (*C/N0*).
d. If a BCN00 message is sent to one source, to maintain the rate consistency, the switch also sets the advertised rate for all sources to *Rmin*. This is the rate that is sent in FECN messages.

5 Performance Evaluation

To evaluate the performance of E-FECN, Network Simulator Version 2 (NS2) is used for the simulations work. One sample simulation result is presented in this paper. Figure 5 shows the network configuration. Networks parameters are listed in Table 2. There are four source nodes (SU1, SU2, SU3, and SU4) and only one destination node (DU). Each source node is linked to an edge switch (SW1, SW2, SW3, and SW4). All edge switches are linked to a single core switch (CS). Link propagation delays are 0.5 □s. Node processing delays are 1 □s. Link speeds are 10 Gbps. At all switches, a drop-tail queue mechanism is used if the buffers overflow. The switch output buffers can hold 100 packets of 1500 bytes each, i.e., the buffer size is 1,500×100 = 150,000 bytes). Ethernet's standard PAUSE mechanism is not used. The traffic generation is at a constant bit rate (CBR) with UDP traffic over Ethernet. One CBR continuous flow is used per source node. The simulation duration is 100 ms. All four flows start at the 5 ms and two out of four ends at 80 ms. Other parameters are shown in Table 2. We used BCNMAX option with BCN, the maximum negative feedback is send back to the source at a threshold of $2 \times Qeq$, because it improves BCN and has been used by their developers in all recent simulations [12]. Together with oversampling technique, BCNMAX results in a faster response to sudden and quick positive changes in queue length. *Rmin* is set to 500 Mbps since FECN recommends *N0 = 20* for small topology and in FECN with slow start, the initial rate is set to *C/N0*, which is 500 Mbps.

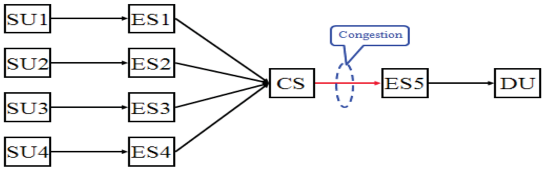


Figure 5: Network configuration

Table 2: BCN, FECN, and E-FECN parameters

Parameters	**BCNMAX**	**FECN**	**E-FECN**
Qeq (equilibrium)	16 packets (24,000 bytes)	16 packets (24,000 bytes)	16 packets (24,000 bytes)
Qsc (severe control)	80 packets (120,000 bytes)	80 packets (120,000 bytes)	80 packets (120,000 bytes)
Sampling Rate	2% (every 75,000 bytes)	N/A	N/A
Over-sampling (Q>Qsc)	10% (every15,000 bytes)	N/A	N/A
W, Ru	W = 2, R_u = 1 Mbps	N/A	N/A
Gi, Gd	G_i = 0.533 , G_d = 0.0002667	N/A	N/A
a, b, c	N/A	a = 1.1, b=1.002, c = 0.1	a = 1.1, b=1.002, c = 0.1
Initial rate (Fast Start)	10 Gbps	10 Gbps (N0=1)	10 Gbps (N0=1)
Tagging frequency	N/A	Every 1ms	Every 1ms
Rmin (E-FECN)	N/A	N/A	500 Mbps
Timeout	N/A	2 ms	2 ms

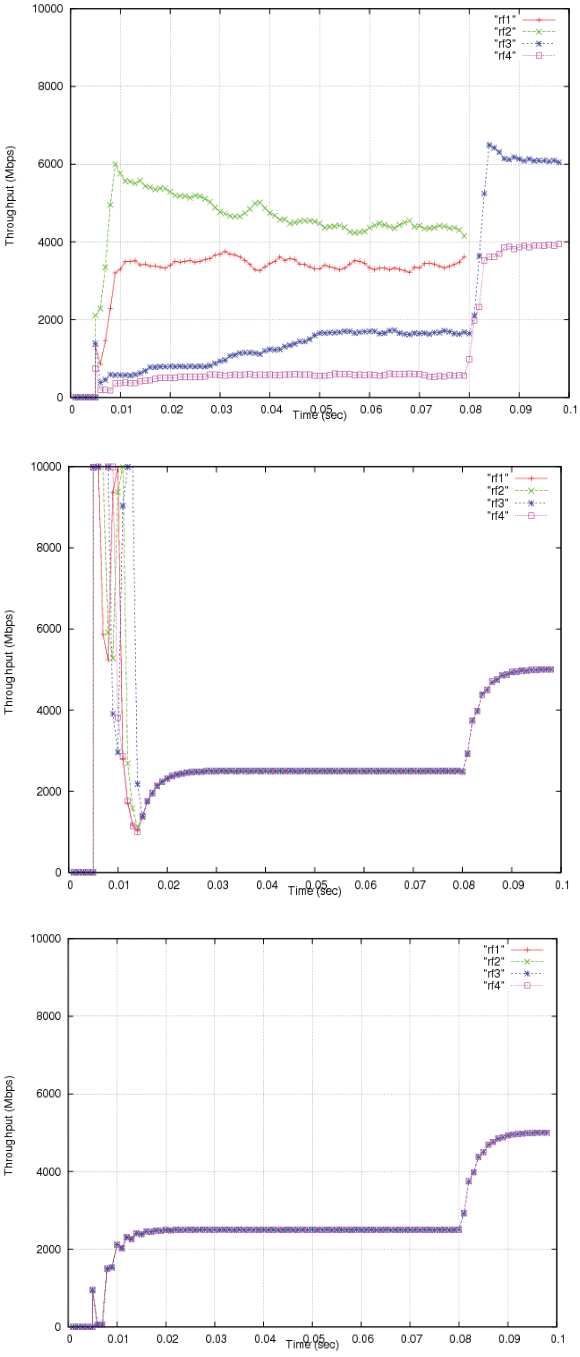


Figure 6: Source rates

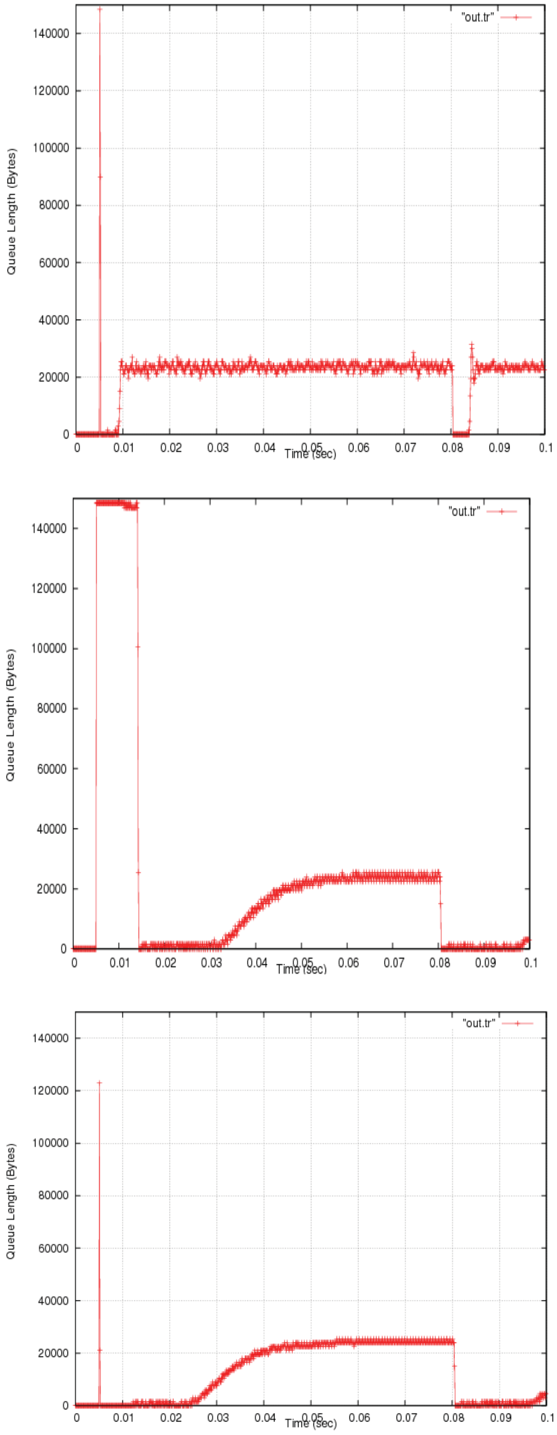


Figure 7: Queue length at the core switch

5.1 Simulation Results

Figure 6 shows source rates for the schemes. Left most graphs are for BCN, middle graph is for FECN and the right most graphs is for E-FECN. Notice that the four sources get very different rates with BCN. This shows the unfairness of BCN. The lowest rate source gets to increase at 80 ms when two of the four flows stop. Even after that the remaining two sources have different rates. With FECN, the four curves (for four sources) are on the top of each other and so it is fair but there are large transients in the beginning before steady state is achieved. With E-FECN, the transients are eliminated and fairness is maintained. Thus, we get both fast convergence and fairness. Note that the convergence time for fair and efficient throughput for FECN is around 20 ms, while for E-FECN; it is only 15 ms. Figure 7 shows queue length at the core switch for the three schemes. Note that all three schemes can stabilize the queue at the desired *Qeq* (24,000 bytes). BCN has a few packet losses at the beginning (spike). FECN has high queue for around 10 ms. E-FECN has no packet losses at all.

6 Conclusions and Future Work

BCN and FECN are two proposed schemes for congestion notification in datacenter Ethernet networks under IEEE 802.1Qau group. There are several variations of BCN and there are issues of parameter selections. BCN with over-sampling, drifting, BCNMAX seems to be the best among BCN variations. Therefore, this variation of BCN is used for comparison with FECN and E-FECN in this paper. Although, we present only one simulation result, it clearly shows the strengths and weaknesses of the three schemes. It is obvious that BCN is unfair; FECN is fair but needs to start at a low rate, which means that each flow needs a rate regulator. In this paper, we proposed E-EFCN, an enhancement to FECN that also uses a backward congestion notification BCN00 message to limit the source rate under severe congestion. With this feature, E-FECN maintains the perfect fairness of FECN and also allows sources to start at high rates. This reduces the number of rate regulators to the same number as in BCN mechanism.

7 References

[1] Yi Lu, Rong Pan, Balaji Prabhakar, Davide Bergamasco, Valentina Alaria, and Andrea Baldini, “Congestion control in networks with no congestion drops,” September 2006.

[2] Joerg Widmer, Robert Denda, and Martin Mauve, “A Survey on TCP-Friendly Congestion Control,” IEEE Network, vol.15, no.3, pp.28-37, May 2001.

[3] Eric He, Pascale Vicat-Blanc, and Michael Welzl, “A Survey of Transport Protocols other than “Standard” TCP,” Global Grid Forum, Data Transport Research Group, April 2005.

[4] Frank.Kelly, "The mathematics of traffic in networks," in the Princeton Companion to Mathematics," Princeton University Press.

[5] Davide Bergamasco, "Datacenter Ethernet Congestion Management: Backward Congestion Notification," IEEE 802.1 Meeting, May 2005.

[6] Davide Bergamasco and Rong Pan, "Backward Congestion Notification Version 2.0," IEEE 802.1 Meeting, September 2005.

[7] Bruce Kwan and Jing Ding, "BCN Calibration Simulation with Global Pause and Drift," IEEE 802.1 Meeting, Davide Bergamasco, "Ethernet Congestion Manager," private communications, March 2007.October 2006.

[8] Davide Bergamasco, "Ethernet Congestion Manager," private communications, March 2007.

[9] Jinjing Jiang, Raj Jain, and Chakchai So-In, "Congestion Management for Ethernet In Datacenter Application Using Forward Explicit Rate Notification," WUSTL technical report, 2007.

[10] Cyriel Minkenberg and Mitch Gusat, "E2CM updates,"

[11] Jinjing Jiang, Raj Jain, and Chakchai So-In, "An Explicit Rate Control Framework for loss free Ethernet operation," Accepted to appear in ICC 2008.

[12] Bruce Kwan and Jing Ding, "BCN Calibration Simulation with Global Pause and Drift," IEEE 802.1 Meeting, October 2006.