SESSION

SENSOR NETWORKS AND APPLICATIONS + RELATED ISSUES

Chair(s)

TBA

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

Performance Analysis of IEEE 802.15.4 For Intra-Vehicle Wireless Networks

Utayba Mohammad
Department of Electrical and Computer Engineering, University of Detroit Mercy, Detroit, MI, USA

Abstract - *In the past decade, cars have evolved to become very complex systems, offering a verity of safety and infotainment options. This evolution was made possible by the development of advanced intra-vehicle communication networks that allowed for implementing highly distributed information processing systems and achieving significant cost reduction in manufacturing. Many efforts have been directed to improve the efficiency of intra-vehicle networks, enhance its scalability, and reduce its cost. Wireless technology presents a potentially viable approach to meet all of these goals. Recently, industries in general, and automotive in particular, have been exploring the deployment of wireless technologies in harsh environments and attempting to identify the underlying challenges to these networks.*

This paper evaluates IEEE 802.15.4 as an intra-vehicle wireless network protocol. It explores implementing in-vehicles space division multiplexing with IEEE 802.15.4, develops an intra-vehicle wireless connectivity diagram with IEEE 802.15.4, and evaluates the performance of IEEE 802.15.4 under extreme interference conditions with coexisting IEEE 802.11 networks.

Keywords*:* Intra-vehicle Network; IEEE 802.15.4; wireless sensor networks; real-time wireless communication

1 Introduction

Multiplexed vehicle communications emerged to support the increased complexity of automotive system, and to enable scalable, upgradable, efficient, and low-cost designs [1][2]. In multiplexed vehicle networks, a single bus is shared among multiple devices and is usually accessed, either by using a Time Division Multiplexing Access (TDMA) technique, or by using bus contention resolution techniques, such as CSMA/CR. Many multiplexing protocols have been proposed in this area. SAE classified those protocols, according to speed, into three classes A, B, and C [3]. Class A is used for interfacing with simple switches or sensors. It represents low speed protocols with a bit rate up to 10 Kbps and a message rate about 100 ms. Class B protocols provide a bit rate between 10 and 125 Kbps and a message rate around 20 ms, allowing for the communication protocol to handle some parts of the engine and transmission controllers. Finally, class C protocols provide bit rates starting from 125 Kbps and up to 1 Mbps, and a message rate of less than 5 ms, which makes these protocols most suitable for real time and critical control system communications. However, several protocols have emerged recently, offering bit rates over 1 Mbps and marking a new class of protocols, “Class D.”

The main two objectives of multiplexed vehicle networks are: reducing the wiring costs and complexity of the network, and distributing th in-vehicle processing load over multiple low-end processors. In this context, a wireless communication system is better than the single bus multiplexing system, since it supports the concept of distributed systems, and yet does not require any wires. Moreover, intra-vehicle wireless networks become even more attractive when plug-and-play sensors are implemented. Currently, the potential wireless protocols for intra-vehicle networks are IEEE 802.15.4, IEEE 802.15.1/Bluetooth, and the IEEE 802.11 family [4][5]. This paper presents the potential use of each of these protocols and evaluates the IEEE 802.15.4 performance for intra-vehicle network applications.

The rest of the paper is organized as follows: IEEE 802.11 family is discussed in section II, and Bluetooth is reviewed in section III. IEEE 802.15.4 is then presented in section IV and evaluated experimentally in section V. Finally, the paper is concluded in section VI.

2 IEEE 802.11

IEEE 802.11 standards were developed to serve wireless LANs. These protocols operate at the ISM 5.9 GHz or 2.4 GHz bands, and offer high bit-rates that range from 11 Mbps in 802.11b to 300 Mbps in the draft 802.11n. Since designed for LANs, these protocols were optimized to deal with big chunks of data, such as images and video files, and address some of the common challenges in wireless LANs, such as hidden and exposed terminal problems. As a result 802.11 protocols are not best suited for intra-vehicle networks, especially when it comes to transmission power and protocol overhead. Therefore, researchers have not considered 802.11 protocols for limited area wireless sensor networks, and rather, they focused more on the Personal Area Networks (PANs) that resemble in many ways the intra-vehicle networks.

3 IEEE 802.15.1/Bluetooth

Bluetooth is one of the most common WPAN protocols. It was first developed by Ericson in 1994 and released in 1998 as a standard by the Bluetooth Special Interest Group (SIG), including Ericson, IBM, Intel, Nokia, and Toshiba [6], [7], [8]. The initial idea was to replace short cable connectivities with wireless devices that are cheap, battery operated, and have power saving modes. However, these features were extended later on to meet the requirements of an ad-hoc network's nature and emphasize short-range communication, as well as, satisfy the low cost constraints.

Since Bluetooth started as a cable replacement, it was reasonable for it to adopt a master-slave, peer-to-peer communication approach. However, the need for networking support extended the master-slave concept to a one-master, multiple-slaves network concept, which is called piconet. The piconet must have one, and only one, master, and can contain up to seven slaves and 255 parking nodes. Bluetooth uses the Frequency Hopping Spread Spectrum (FHSS) technique to provide immunity against noise and interference. Therefore, a frequency-hopping sequence is generated based on the master node address and is used by all piconet members. Each piconet's hopping sequence is unique because their master's address is unique, which explains why one piconet's slaves cannot communicate with another piconet's master directly. This raises the question of how we can propagate data between different piconets. The answer came with the scatternet concept. A scatternet is a network of piconets. A master of one piconet can be a slave in another piconet, and a slave in one piconet can be slave in three other piconets.

Bluetooth radio operates in the 2.4 GHz unlicensed Industrial Scientific Medical (ISM) band and uses a fast Frequency Hopping Spread Spectrum (FHSS) and Gaussian Frequency Shift Keying (GFSK) modulation (generally) to transmit its data. Throughout the transmission, Bluetooth radio uses 79 channels (carrier frequencies), spaced 1 MHz from each other. The time between two frequency hops is called a slot and is equal to 625 µsec. Power-wise; Bluetooth is categorized into three classes. Class 1 has a maximum transmission power of 100 mW and a minimum transmission power of 1 mW, and can reach typically to 100 m. Class 2 ranges from 1mW to 2.5 mW and can reach up to 20m. Finally, Class 3 has a maximum transmission power of 1 mW and can reach up to 10 m [9].

Although supports many of the in-vehicle network requirements, Bluetooth faces some serious problems when it comes to its architecture, especially, the piconet limited scalability, the involved delays in scatternet and operating mode switching, and the master-slave topology that increases the bandwidth demand [10].

4 IEEE 802.15.4

IEEE 802.15.4 standard was developed to meet the different needs of Wireless Personal Area Networks (WPANs), which includes low power consumption, self-organizing, self-healing, and the ability for expansion [9], [10], [11], [12]. Since it was designed for WPANs that satisfy home, industry, and environment applications' requirements, it was necessary to choose its operational frequencies from the free unlicensed frequency range. Therefore, IEEE 802.15.4 has been chosen to operate in three different free unlicensed frequency ranges, with each one having a different number of channels, bandwidth, and data rate. These frequency ranges are:

- 868 – 868.6 MHz, used in Europe, has one channel only and provides an ideal bit rate of 20 Kbps.
- 902 – 928 MHz, used in USA, has ten channels and provides an ideal bit rate of 40 Kbps.
- 2400 – 2483.5 MHz, used in most of the world and has sixteen channels with a maximum bit rate of 256 Kbps.

4.1 Network topology

The network devices in IEEE 802.15.4 are classified according to their actual physical specifications or their logical task in the network. Physically, the network devices are classified as Full-Function Devices (FFDs) or Reduced-Function Devices (RFDs); while logically they are classified as PAN coordinator, routers, and end devices. These two classifications intersect with each other, i.e. a FFD can operate as PAN coordinator, router, or regular end device; on the other hand, a RFD can only operate as an end device. FFDs have higher physical requirements than RFDs, such as more memory, higher computational capabilities and, consequently, larger power consumption requirements (usually powered by a main supply). This level of complexity makes FFDs capable of communicating with other FFDs or RFDs. On the other hand, RFDs have limited resources of memory, computational capabilities, and power (usually powered by batteries), and, hence, can only communicate with FFDs only.

Based on the three WPAN logical components (coordinator, router, and end device), IEEE 802.15.4 supports three Network topologies. The first one is the Star topology, in which all data transfers occur between the PAN coordinator and the other network devices as shown in Fig.1-a. The second topology is the Peer-to-Peer (Mesh Network) topology, which still has one PAN coordinator, but its devices can communicate among each other's without accessing the coordinator as shown in Fig.1-b. Since only FFDs can communicate with other FFDs and RFDs, they are the most commonly used devices in the mesh networks, while RFDs only reside on the leaves of these networks to provide sensing and actuating. The third topology is a special formation of the peer-to-peer topology, and is called the Cluster-Tree topology. The cluster-Tree topology is a trade off option, offering less power consumption and connectivity than the peer-to-peer topology and more, of both, than the star topology. Fig.1-c depicts the cluster tree topology.

4.2 IEEE 802.15.4 operation schemes:

The communication management in IEEE 802.15.4 can operate in one of two modes, the beacon-enabled mode or the nonbeacon-enabled mode. In the beacon-enabled mode, the coordinator sends periodic beacons to synchronize all devices with the PAN superframe. The superframe is bound by two beacons, and divided into two main periods, active and inactive as shown in Fig. 2. The active period comprises two main periods, the Contention Access Period (CAP) and the Contention Free Period (CFP). The CAP is divided into time slots, so that devices willing to transmit shall use a slotted CSMA/CA algorithm to compete for the medium.

On the other hand, the CFP is divided into Guaranteed Time Slots (GTS), where each slot is allocated to a predefined device to insure critical data transmission.

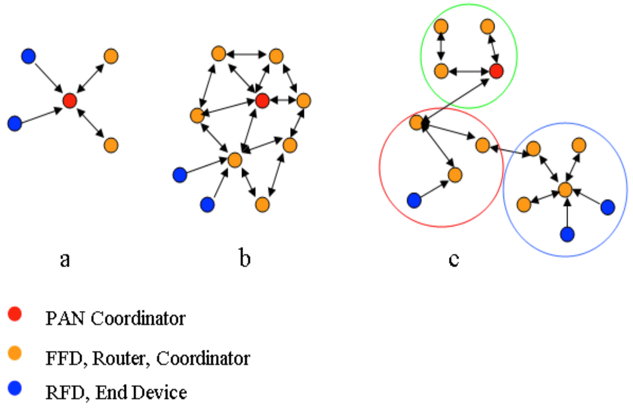


Fig. 1. ZigBee Topologies, a- Star, b- Mesh Network, and c-Cluster Tree

The Nonbeacon-enabled mode is much simpler than the beacon-enabled one. The devices in this mode compete for the medium with CSMA/CA algorithm. Hence, delays in the physical layer are less than those in the beacon-enabled mode, but there is no guaranteed delivery mechanism.

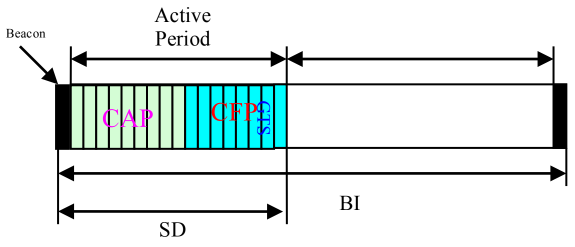


Fig. 2. IEEE 802.25.4 Superframe

4.3 IEEE 802.15.4 MAC layer:

The IEEE 802.15.4 MAC layer plays a major role in driving and determining the efficiency of the communication system. It setups the superframe period division, allocates GTSs, and orchestrates the association process to form up the whole network. The superframe structure in the beacon-enabled IEEE 802.15.4 has a significant importance in optimizing the communication among the different devices. The MAC layer is responsible of structuring the superframe using several user-defined constants, such as the Beacon Order (BO) and the Superframe Order (SO). As described previously, the superframe is bounded between two beacons, therefore, the period of complete superframe is called the Beacon Interval (BI), while the active part of the superframe is called the Superframe Duration (SD). The Beacon Interval and Superframe Duration are defined by Superframe Order and Beacon Order respectively according to Eq. (1) and Eq.(2), respectively.

$$BI = B * 2^{BO}, \text{ where, } 0 \leq BO \leq 15 \qquad (1)$$

$$SD= B * 2^{SO}, \text{ where, } 0 \leq SO \leq BO \qquad (2)$$

Where B is a MAC layer constant called aBaseSuperframeDuration.

In case BO=BI=15, the system functions in a nonbeacon-enabled fashion; while BO=BI=0, will lead to a minimum beacon interval and fully active superframe, resulting in maximum utilization of the superframe period.

With its flexible and configurable topology, IEEE 802.15.4 has a good potential to serve as an intra-vehicle communication protocol. The rest of the paper evaluates the IEEE 802.15.4 protocol for intra-vehicle applications, and tries to find what type of communication class it can support in comparison to the vehicle wired multiplexed communication protocols.

5 IEEE 802.15.4 evaluation

To assess IEEE 802.15.4's potential for automotive applications a set of experiments were conducted. In [14], ZigBee's performance was evaluated thoroughly in a General Motors 2005 Cadillac STS. Although that study covered a wide range of measurements, it was very particular to the case study. In this section the physical layer performance of IEEE 802.15.4 in vehicles is evaluated with a more general approach. Three experiments are conducted: the first, determines the major factors that affect the wireless signal strength at different parts of a vehicle; the second, develops an approximate wireless connectivity diagram inside vehicles; and the third evaluates IEEE 802.15.4 coexistence with WiFi. Throughout the experiments MICAZ nodes running the TinyOs embedded operating system were used to determine the Received Signal Strength and Packet Error Rate (PER).

5.1 Main factors in intra-vehicle wireless signal propagation

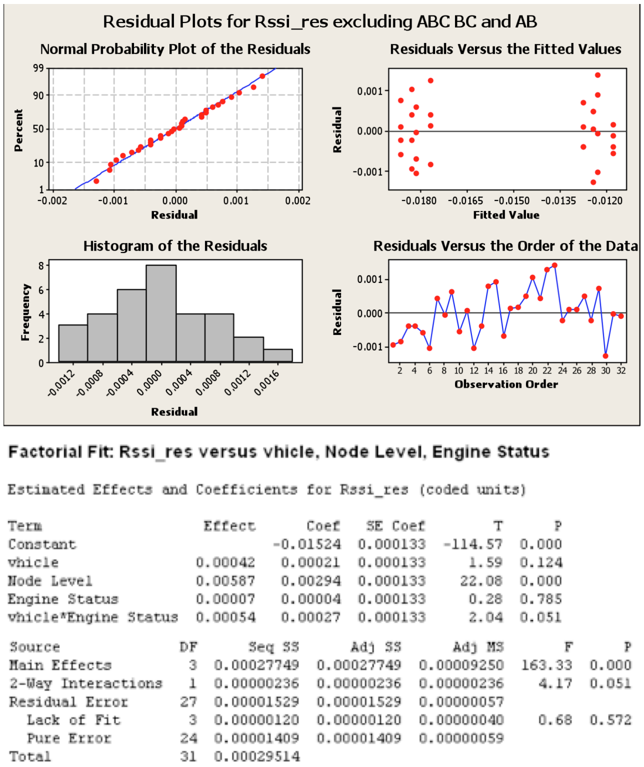
This experiment was designed with three factors and four replicates. the factors are:

- Vehicle Type (factor A): Two vehicles are used as test plants, a Chevrolet Prizm 1999 and a Mazda LX 1996.
- Node level (factor B): Both nodes are installed at the same level (height), and then one node is installed at the ground level and the other at the engine level.
- Engine Status (factor C): Engine is OFF, and Engine is ON running idle.

A 2^3 factorial analysis with 32 measurements was performed to study the effect of the aforementioned factors, and their interactions, on the Received Signal Strength (RSS). However, due the dependencies between the mean and the variance of the RSS a linear 1/Y variance stabilization transform was applied to the data sample and the insignificant 2-way and 3-way factor interactions were discarded. Fig. 3 shows the residual plots of the transformed data and the associated 2^3 factorial analysis. Fig. 3 further shows that the residuals do approximate a normal distribution and have a constant variance, which satisfies the assumptions of a factorial design. As a result the reported P-values can be used from the analysis to conclude that we are almost certain and 94.9% confident, that the node level and the “vehicle type * engine status” interaction, respectively, affect wireless signal propagation in the engine compartment. This conclusion indicates that a connectivity diagram needs to be developed before planning to deploy wireless nodes in a vehicle, and that this diagram is unique to each vehicle based on the vehicle’s internal component distribution and its electronic and electrical wiring.

5.2 Wireless connectivity diagram in vehicles

This section attempts to draw a rough diagram of wireless connectivity inside vehicles. MICAZ nodes were used in this experiment, with one transmitter, one receiver, and a transmission power of -25 dBm. The transmission power was chosen as the minimum possible value in order to guarantee the least interference between adjacent vehicles.



```
Factorial Fit: Rssi_res versus vhicle, Node Level, Engine Status

Estimated Effects and Coefficients for Rssi_res (coded units)

Term                       Effect      Coef   SE Coef        T      P
Constant                           -0.01524  0.000133  -114.57  0.000
vhicle                    0.00042   0.00021  0.000133     1.59  0.124
Node Level                0.00587   0.00294  0.000133    22.08  0.000
Engine Status             0.00007   0.00004  0.000133     0.28  0.785
vhicle*Engine Status      0.00054   0.00027  0.000133     2.04  0.051

Source                  DF      Seq SS      Adj SS      Adj MS       F      P
Main Effects             3  0.00027749  0.00027749  0.00009250  163.33  0.000
2-Way Interactions       1  0.00000236  0.00000236  0.00000236    4.17  0.051
Residual Error          27  0.00001529  0.00001529  0.00000057
  Lack of Fit            3  0.00000120  0.00000120  0.00000040    0.68  0.572
  Pure Error            24  0.00001409  0.00001409  0.00000059
Total                   31  0.00029514
```

Fig. 3. The 2^3 factorial experiment analysis results after after applying a linear transformation Y* = 1/Y and aliasing the insignificant terms

The vehicle under test was virtually divided into three sections: the engine compartment, the cabin, and the trunk. The wireless connectivity in each of these sections and among themselves was evaluated and a general connectivity diagram was deduced. The following sub-sections show the node placement and the corresponding received power levels, while the conclusion of these measurements is presented at the end of this experiment.

5.2.1 Engine compartment test

The transmitter and receiver are placed in three different configurations: at the same level under the hood, at the same level on the ground, and at the two ends of the engine compartment diagonal. Fig. 4 shows the placement configurations for the engine compartment test, and Table I elaborates on the received signal power at different points in the engine compartment.

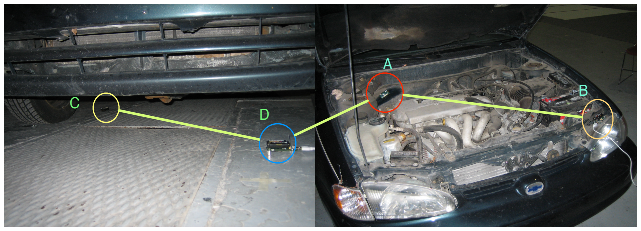


Fig. 4. Nodes’ placements for the engine compartment test

TABLE I. RECEIVED SIGNAL POWER IN THE ENGINE COMPARTMENT

Transmitter	Receiver	Received Power [dBm]
A	B	-73
C	D	-72
A	D	-91

5.2.2 The Vehicle Cabin Test

The main communication in the vehicle cabin is usually between the seat modules, door modules, and the dashboard cluster. To emulate such communications, the transmitter was placed in the Glove Box and the receiver was moved around the front seat pocket, back seat pocket, and under the front seat and for both sides of the vehicle. The received power is recorded as shown in Table II.

TABLE II. RECEIVED SIGNAL POWER IN THE VEHICLE CABIN

Transmitter	Receiver	Received Power [dBm]
Glove Box	Front Left door pocket	-75
Glove Box	Front Right door pocket	-61
Glove Box	Back Left door pocket	-75
Glove Box	Back Right Door Pocket	-78
Glove Box	Under the Front Left Seat	-75
Glove Box	Under the Front Right Seat	-62

5.2.3 Trunk test:

The communication in the trunk section is mainly concerned with the tail and the center light signaling. Therefore, our measurements were taken in three locations in the trunk as shown in Fig. 5. Table III shows the received signal power at different locations in the trunk.

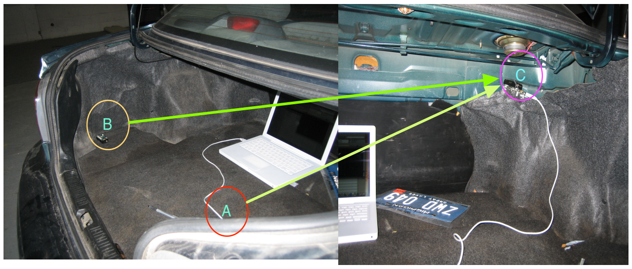


Fig. 5. Communication testing in the vehicle trunk

TABLE III. RECEIVED SIGNAL POWER IN THE VEHICLE TRUNK

Transmitter	Receiver	Received Power [dBm]
B	A	-65
C	A	-54

5.2.4 Inter-section Connectivity test:

After the connectivity within the individual section was evaluated, the inter-section connectivity was studied. Table IV shows the received power level between the different sections when the vehicle is empty. "No connectivity" indicates that the received power falls below the receiver sensitivity, which is -95 dBm for MICAZ (CC2420).

5.2.5 Conclusions on intra-vehicle wireless connectivity

In an empty vehicle and with a -25 dBm transmission power, the wireless connectivity provided by IEEE 802.15.4 compliant transceivers could be divided into seven main sub-networks: the headlights, engine compartment, dashboard, left door module, right door module, seat controls, and the trunk. The connectivity within each of these sub-networks is preserved. However the connectivity among the different areas is not guaranteed. More specifically, the engine compartment nodes are connected with the dashboard and headlight modules; however, the last two are not connected. Also, the dashboard is connected to both door modules and, sometimes, with the seat control; but it fails to connect with the trunk, as well as the seat modules when the vehicle is loaded with passengers. Finally, the seat controls are connected with the trunk module all the time. Fig. 6 depicts a rough wireless connectivity diagram in a vehicle when using IEEE 802.15.4 transceivers.

TABLE IV. RECEIVED SIGNAL POWER DURING THE INTER-SECTION COMMUNICATION TEST IN THE VEHICLE

Transmitter Location	Receiver Location	Received Power [dBm]
Left Tail Light in the Trunk	Front Head Lights (Both Left & Right)	No Connectivity
Rear Left and Right Side – ground level	Front Left & Right Side – ground level	-84
Glove Box	Center of the Engine Compartment	-83
Glove Box	Front Headlight	No Connectivity
Left Door Pocket	Center of the Engine Compartment	-88
Glove Box	Left Tail Light in the trunk	-87
Back Seat	Trunk	-72

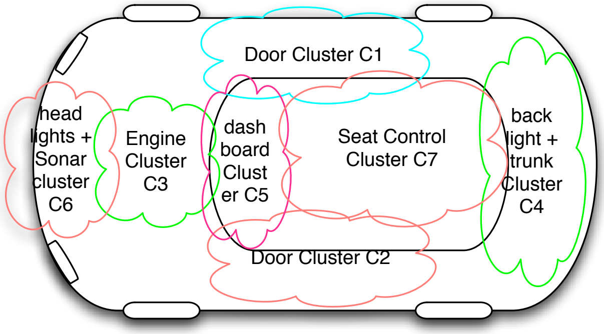


Fig. 6. Wireless connectivity Diagram in a vehicle when using IEEE 802.15.4 compliant transceivers

5.3 Coexistence with Wifi

Nowadays, It is very common to come across Wifi-enabled devices, E.g. cell phones, wifi hotspots, wifi-based remote controls, security cameras ...etc. This poses the questions: how well will ZigBee/IEEE 802.15.4 tolerate the co-existence of Wifi-enabled device? And would it still be able to serve under extreme conditions of interference and spectrum contention?

To answer these questions two MICAz nodes were used as a transmitter and a receiver. The transmitter was set up to send bulk messages of 200 packets each every 5 seconds with transmission power of -25 dBm, in compliance with the previous connectivity test. The receiver was set up to report the number of received packets from each message to a monitoring computer. The MICAz nodes were alternated to operate at IEEE 802.15.4's channel 26 and 22, while two computers were used to provide an active Wifi file transfer over channel 11. The experimenter occupyied the passenger seat and used a notebook to transmit a data file to nearby WiFi base station. At the same time the IEEE 802.15.4 transmitter was positioned in the glove box and the receiver was placed in the close door pocket. This setup was meant to imitate a real life scenario of a user using WiFi inside a vehicle while sensory data is exchanged wirelessly via IEEE 802.15.4. The experimental setup is illustrated in Fig. 7.

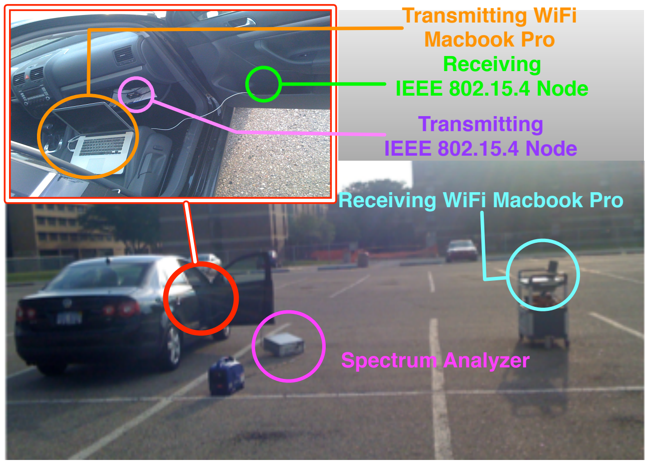

Transmitting WiFi Macbook Pro
Receiving IEEE 802.15.4 Node
Transmitting IEEE 802.15.4 Node
Receiving WiFi Macbook Pro
Spectrum Analyzer

Fig. 7. The setup for testing IEEE 802.15.4 coexistence with WiFi

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

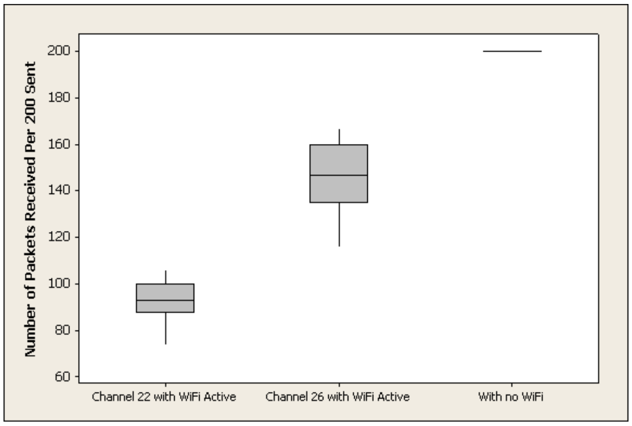


Fig. 8. Box plots for the IEEE 802.15.4's interference with channel 11 of WiFi

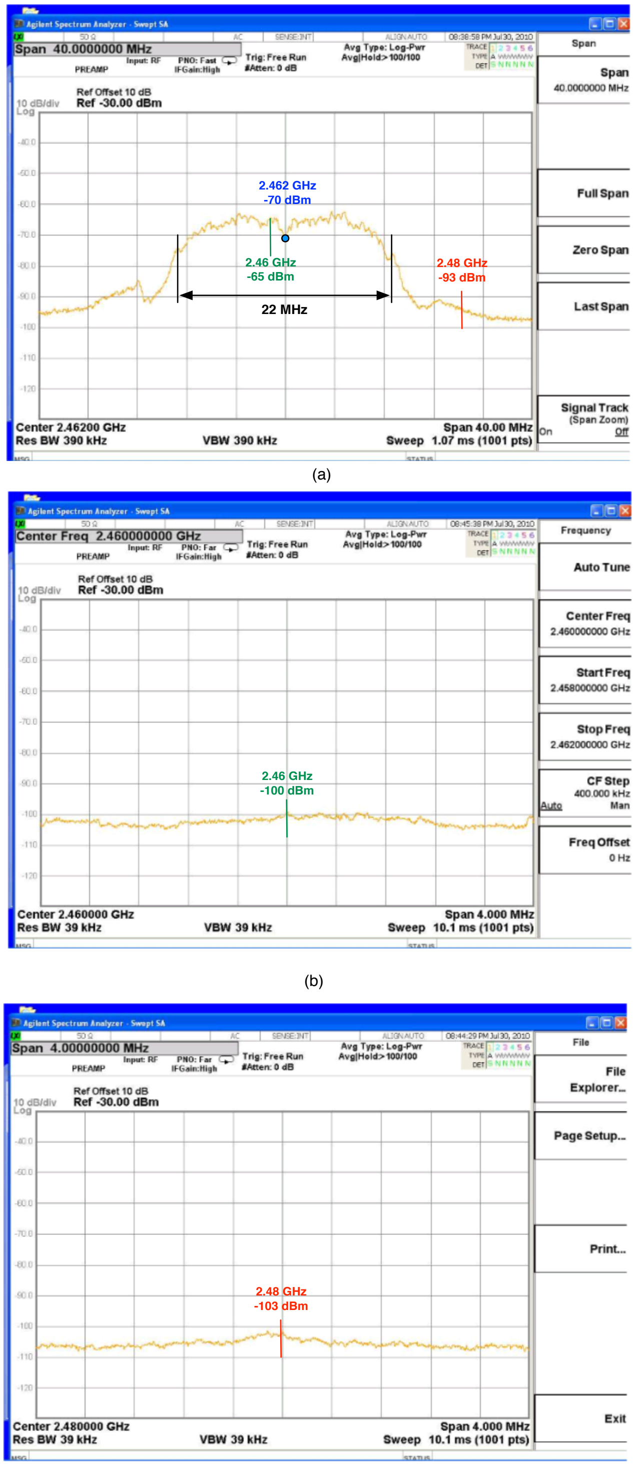
In IEEE 802.15.4, channel 26 and 22 are centered at 2.48 GHz and 2.46 GHz respectively, with a channel bandwidth of 2 MHz. On the other hand, channel 11 of WiFi is centered at 2.462 GHz and has a bandwidth of 22 MHz. In this setup, overlapped channels (22 of IEEE 802.15.4 and 11 of WiFi) were used to measure the maximum effect of interference. On the other hand, channel 26 of IEEE 802.15.4 and 11 of WiFi were used as two adjacent non-overlapping channels in order to measure the minimum interference effect on IEEE 802.15.4 performance. Fig. 8 shows the box plots, over 15 measurements, for the number of received packets during active and inactive channel 11 WiFi and for both IEEE 802.15.4's channels 22 and 26. We indicate here that we did not distinguish between channels 22 and 26 when WiFi is inactive because 100% of the messages were received in both cases and for all measurements.

Although channel 26 of IEEE 802.15.4 is, theoretically, spaced 7 MHz from the upper band of WiFi's channel 11 (2.473 GHz), Fig. 8 reveals a significant drop in number of received packets when WiFi is active vs. when WiFi is not active. The box plots show more resilience to WiFi when using channel 26 vs channel 11, but suggest that they are both not practically usable for control application purposes.

These results are further explained by using a spectrum analyzer. It is found that channel 11 of WiFi still overlaps with Channel 26 of IEEE 802.15.4, but at a much lower level than is the case with channel 22. Fig. 9 shows snapshots of the spectrum during the experiment for channel 11 of WiFi, during active and inactive transfers, and for channel 26 and 22 of IEEE 802.15.4 during inactive WiFi transfer.

6 Conclusion

The experimental results show that IEEE 802.15.4 is capable of operating in an intra-vehicle environment, and that it exhibits good immunity against vehicular electromagnetic noise when the engine is on and when it is off. The propagation characteristics throughout the vehicle are shown to be affected by the obstructing metal parts in the vehicle. At the lowest transmission power of -25 dBm the wireless signal is decently contained within the vehicle vicinity, and therefore a vehicle wireless connectivity diagram of 7 connected sections is derived at this power. Further experimental work indicated that full connectivity could be achieved by increasing the transmission power but with risking the result of inter-vehicle interference.



(c)

Fig. 9. (a) channel 11 spectrum, (b) channel 22 spectrum, (c) channel 26 spectrum.

In contrast to the general assumption that IEEE 802.15.4 and WiFi can coexist without degradation in the communication quality, our expiremnt shows that using the -25 dBm transmission power of IEEE 802.15.4 in the close vicinty of active WiFi transmitter can degrade the communication quality significantly. The worst case takes place IEEE802.15.4 and WiFi are operating at in overlapping channles with less effect being exerted whe operating in adjacent channles. Hence, using IEEE 802.15.4 for intra-vehicle communication requires coordinating its active channels with the active WiFi channels all the time.

7 References

[1] Mark Thompson, "Replacing wirings with multiplexed in-vehicle communication system Papers: Thick and thin of car cabling", Motorola, IEEE Spectrum, v33, n 2, February 1996, pp.42-45

[2] Pretson, N. C. G. N. (University of Liverpool, Engl), J. Lucas, "Multiprocessor Implementation Of The Logic Function Of A Multiplexed Wiring System For Automotives", IEEE Proceeding, Part E: Computers and Digital Techniques, v 129, n 6, November 1982, pp. 223-228.

[3] Christopher A. Lupini, "Vehicle Multiplex Communication - Serial Data Networking Applied to Vehicular Engineering", SAE, April 2004.

[4] A. Willig, K. Matheus, A. Wolisz, "Wireless Technology in Industrial Networks", Proceedings of the IEEE, June 2005, Volume (93), Issue (6), pp. 1130- 1151.

[5] T. Nolte, H. Hanssonlo, L. Bello, "Automotive communication past current and future", Emerging Technologies and Factory Automation, 2005. ETFA 2005. 10th IEEE Conference on, Volume (1), pp. 985- 992.

[6] C. Bisdikian, "An overview of the Bluetooth wireless technology", IEEE Communications Magazine, v 39, n 12, December, 2001, p 86-94.

[7] Specification of the Bluetooth System, Covered Core Package version: 2.0 + EDR Current Master TOC issued: 4 November 2004 volume (0), Available: http://www.bluetooth.com/Bluetooth/Learn/Technology/Specifications/.

[8] M. Hayoz, " The Bluetooth Wireless Technology An Overview", Master Seminar in Telecommunications, University of Fribourg, Switzerland, 2003.

[9] M. Petrova, J. Riihijarvi, P. Mahonen, S. Labella, "Performance study of IEEE 802.15.4 using measurements and simulations", Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE, 3-6 April 2006, volume(1), pp. 487- 492.

[10] Mohammad, U., Al-Holou, N., "Development of Wireless Protocols for Automotive Applications", Worldcomp ICWN'07, Las Vegas, pp.

[11] IEEE Std 802.15.4™-2003, Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs), 2003.

[12] Jin-Shyan Lee, "An Experiment on Performance Study of IEEE 802.15.4 Wireless Networks", Emerging Technologies and Factory Automation, 2005. ETFA 2005. 10th IEEE Conference on Volume 2, Issue , 19-22 Sept. 2005 Page(s): 451 – 458.

[13] Electronics Design, Strategy News. Shreharsha Rao, "Estimation ZigBee transmission range in the ISM band", Texas Instrument, EDN Europe, 01 July 2007. [online] available: http://www.edn-europe.com/estimatingzigbeetransmissionrangeintheismband+article+1608+Europe.html

[14] Hsin-Mu Tsai; Tonguz, O.K.; Saraydar, C.; Talty, T.; Ames, M.; Macdonald, A.; "Zigbee-based intra-car wireless sensor networks: a case study".Wireless Communications, IEEE, Issue 6, December 2007, pp. 67-77.

[15] Ns2 Network Simulator website: http://www.isi.edu/nsnam/ns/.

Mohammad, U., Al-Holou, N., Balas, C.,"Performance Evaluation of IEEE 802.15.4/ZigBee Protocol for Automotive Applications", SAE Congress 2008, In-Vehicle Networks and Software, SP-2197, pp. 69-74

Time Synchronization Protocol using Lighting Control for Wireless Sensor Network

Shoichi Nagamitsu[1], Hiroto Aida[2], Ryoga Okunishi[1], Yo Motoya[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— *Wireless sensor networks use a large number of collaborating sensors with a built-in wireless device to enable the collection of information in real space. Their wireless sensor nodes require time-synchronization in order to achieve a high degree of sensing accuracy, which is imposed upon them by wireless sensor network applications. A number of researchers have proposed time-synchronization protocols, since these protocols require packet transmissions, they consume excessive energy and impose additional traffic load on the network. Therefore, it is important to implement a time-synchronization technique that does not require packet transmission or reception. This paper proposes time synchronization protocol based lighting control, which is a time synchronization technique that uses light-controllable lighting fixtures and wireless sensor nodes equipped with an illuminance sensor without packet transmission or reception. This paper also examines synchronization errors with the proposed method and verifies time-synchronization wireless sensor nodes.*

Keywords: Time Synchronization, Lighting Control, Wireless Sensor Network, Energy Conservation

1. Introduction

Wireless sensor networks use a large number of collaborating sensors with a built-in wireless device to enable the collection of information in real space. Compared with wired networks, such wireless networks significantly reduce the cost of implementation. In addition, such networks enable detailed observation of events that were impossible to observe in the past. Wireless sensor networks require their network components to time-synchronize in order to maintain the time integrity of data and to implement power-saving protocols. Their wireless sensor nodes also require time-synchronization in order to achieve a high degree of sensing accuracy, which is imposed upon them by wireless sensor network applications. A number of researchers have proposed time-synchronization protocols, including approaches using satellite-based GPS and Network Time Protocol [9] to time-synchronize nodes on the Internet. However, these protocols are not necessarily optimal in wireless sensor networks with a diversity of node characteristics and strict power constraints. Time synchronization protocols that take advantage of small propagation delays in wireless sensor networks have been proposed. They include Timing-sync Protocol for Sensor Networks (TSN) [2], Reference Broadcast Synchronization (RBS) [1], and Flooding Time Synchronization Protocol (FTSP) [5]. However, since these protocols require packet transmissions, they consume excessive energy and impose additional traffic load on the network.

The authors of this paper are currently developing an Intelligent Lighting System, i.e., a lighting control system that takes power-saving into account [4],[8]. The Intelligent Lighting System consists of lighting fixtures whose light levels are controllable, wireless sensor nodes, a sink node, a control PC, and a power meter [7]. This system enables lighting fixtures to autonomously adjust their light levels and provides each location within an office the level of lighting that the office worker at that particular location desires. By introducing this Intelligent Lighting System, light conditions in an office will improve and power consumption will reduce. The Intelligent Lighting System uses wireless sensor nodes equipped with an illuminance sensor to obtain illuminance at fixed intervals. The Intelligent Lighting System uses the collected illuminance data for control and to estimate geo-locations and disturbances. However, since wireless sensor nodes are not time-synchronized, different wireless sensor nodes may exhibit different times. Thus, it is also necessary to time-synchronize wireless sensor nodes in the Intelligent Lighting System.

In addition, challenges in using wireless sensor nodes in an office include the cost of replacing batteries in these nodes and instability in the wireless communication environment due to a person or an obstacle in the wireless communication pathways. In particular, when multiple packets are lost consecutively due to an increased level of network traffic caused by an unstable wireless communication environment, existing time-synchronization approaches may not only fail to time-synchronize wireless sensor nodes but also increase the number of packet transmissions and create an even larger network traffic load. Therefore, it is important to implement a time-synchronization technique that does not require packet transmission or reception. This paper proposes Time synchronization Protocol based Lighting Control (TPLC), which is a time synchronization technique that does not require packet transmission or reception. TPLC uses light-controllable lighting fixtures and wireless sensor nodes

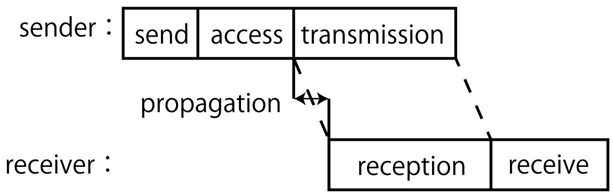


Figure 1: The sender and the reciever packet of FTSP.

equipped with an illuminance sensor. This paper also examines synchronization errors with the proposed TPLC and verifies TPLC time-synchronization wireless sensor nodes.

2. Time Synchronization Protocol for Wireless Sensor Networks

2.1 An importance of time syncronization protocol

Wireless sensor networks require time-synchronization to maintain the time integrity of data and to implement power-saving protocols. Due to this, there exists a large body of research on time synchronization in wireless sensor networks. Examples of such research include TPSN, RBS and FTSP. These are time-synchronization protocols specifically designed for wireless sensor networks that also achieve a high degree of accuracy in time-synchronization with relatively simple mechanisms.

2.2 Flooding Time Synchronization Protocol

FTSP is a technique that extracts all time synchronization errors occurring in a single hop and minimizes such time synchronization errors to achieve a high degree of time synchronization accuracy [5]. In order to minimize time synchronization errors in a single hop, i.e., propagation errors, and errors due to computation, the MAC layer at both the sender and the receiver uses a timestamp. Using a timestamp with both the sender (during transmission) and the receiver (during reception) enables time synchronization errors to be minimized between the sender's and the receiver's timestamps and to be ignored at the time of send, access and receive as shown in Figure 1. In addition, since FTSP only requires packet transmission in one direction, FTSP may time-synchronize an entire network by broadcasting a packet to the entire network.

By implementing FTSP on Mica2 [3], it is confirmed that FTSP achieves time synchronization with an average time synchronization error of 1.4 us and with a max time synchronization error of 4.2 us. FTSP does not, however, consider the instability in a wireless communication environment such as packet loss. In addition, FTSP requires packet transmission and reception, resulting in a possible increase in power consumption at wireless sensor nodes.

3. TPLC: Time Synchronization Protocol based Lighting Control

3.1 Overview

This paper proposes TPLC, a time-synchronization technique based on lighting control that does not require the transmission or reception of packets.

With TPLC, lighting fixtures autonomously adjust their luminance to cause changes in the illuminance level measured by illuminance sensors. TPLC then uses the changes in the measured illuminance to determine when to initiate time synchronization and achieves time synchronization. When adjusting the luminance of lighting fixtures, the luminance adjustment needs to be within a range that is not noticeable by office workers to maintain a comfortable work environment. Existing research indicates that, if the change in illuminance due to luminance adjustment is within 7% of the current illuminance level, one may not detect such luminance adjustments [11]. TPLC performs time synchronization using this small change in illuminance. Therefore, in order to design an algorithm for TPLC, it is necessary to examine how the illuminance changes in time when the luminance is adjusted.

3.2 The measurement of illuminance on a wireless sensor node

We examine changes in illuminance measured by an illuminance sensor on a wireless sensor node for a given illuminance change within approximately 7% of the current illuminance. In these experiments, a MOTE MICAz made by Company C serves as a sensor node [6]. MDA088, a general-purpose external sensor board, is installed onto the MOTE MICAz, and a lead-type NaPiCA illuminance sensor [10] is embedded in the external sensor board to measure illuminance. In these experiments, an illuminance sensor is configured to measure illuminance every 100 ms.

Experiments were conducted in a lab at Doshisha University. The lab used in the experiments simulates an actual office environment and used 15 Panasonic white fluorescent lamps (FHP45EN) and one wireless sensor node. Figure 2 shows a bird's-eye view of the experimental set up. White color partitions were placed by the windows to prevent outside light from entering the lab. The vertical distance between a fluorescent lamp and a wireless sensor node was 1.9 m when the wireless sensor node was placed vertically under the fluorescent lamp. A wireless sensor node was placed vertically under fluorescent lamp 7, and only fluorescent lamp 7 was light-controlled.

A NaPiCa illuminance sensor has a low resolution. Therefore, in these experiments we used a highly precise ANA-F11 illuminance sensor and adjusted the illuminance measurements obtained through the NaPiCa illuminance sensor using equation 1.

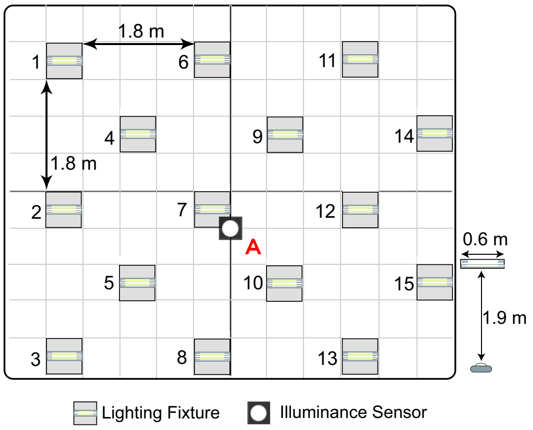


Figure 2: The experiment enviroment for measuring illuminance.

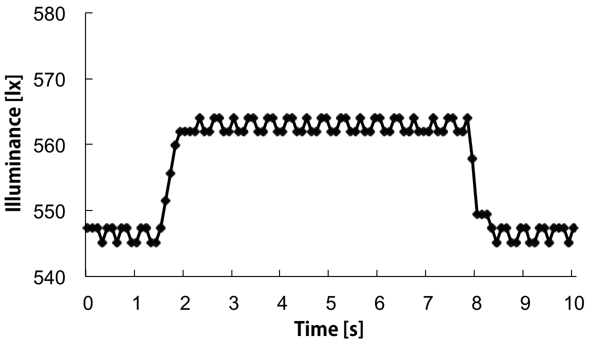


Figure 3: Result in 100 ms intervals to measure illuminance.

$$ANA = 2.096 * NAPICA + 17 \tag{1}$$

ANA : Illuminance given ANA-F11[lx]

$NAPICA$: Illuminance given NaPiCa illuminance sensor[lx]

With the initial condition where illuminance on a desk surface measured by ANA-F11 is 500 lx, we increased illuminance by 15 lx (i.e., 3% of 500 lx), kept illuminance at the increased level for six seconds, and then restored the original illuminance, and obtained how the measured illuminance changed. Figure 3 uses 100 ms intervals to measure illuminance and shows how illuminance changed at a given time.

Figure 3 shows that a NaPiCa illuminance sensor successfully detects a change in lighting luminance even for a small illuminance change within 7% of the current illuminance. Based on these findings, this paper proposes a time synchronization algorithm for TPLC.

3.3 Algorithms

This section explains the time synchronization algorithm for TPLC. Illuminance data does not contain time information such as a timestamp in a packet. Due to this, it is necessary to implicitly convey time information to an illuminance sensor so that time synchronization is initiated based on illuminance measurements obtained by the illuminance sensor. TPLC initiates time synchronization by varying the luminance of a lighting fixture to cause fluctuations in illuminance and having an illuminance sensor detect the illuminance fluctuations. However, as seen in Figure 3, even when the lighting stays at the same luminance level, errors occur in the illuminance measurements, and fluctuations occur in the measured illuminance. When varying illuminance, changes in illuminance need to be larger than errors yet still small enough to not cause discomfort for office workers. Therefore, in performing time synchronization, TPLC applies a small luminance change and initiates time synchronization only when the increase in the measured illuminance caused by the luminance change is at the predetermined ratio.

The following describes a set flow to perform to initiate time synchronization. Assume that all lighting fixtures are initially on and that all wireless sensor nodes are placed within a distance to measure illuminance and are powered on at the same time within a single illuminance measurement time interval. First, periodically measures the current illuminance at every illuminance measurement interval; Calculate the change in illuminance from the measured current illuminance and the previously measured illuminance using the equation; Initiate time synchronization when the illuminance change is equal to the predetermined value. The following set flow constitutes one step.

(1) An illuminance sensor measures the current illuminance.
(2) A lighting fixture increases its luminance by x%, where x < 7%.
(3) If an illuminace sensor detects a change in illuminance, calculate the amount of illuminance change.
(4) If the change in illuminance is equal to the predetermined value, store the measured illuminance value obtained in (1) and the illuminance value after the illuminance change.
(5) If the measured illuminance values before and after the illuminance change in the current step and in the previous step are equal within a margin of error, time-synchronize wireless sensor nodes.
(6) A lighting fixture returns its luminance to what it was before the luminance change. Perform the above operation (1).

This is the time synchronization algorithm for TPLC. Flow (4) is necessary because of the following reasons. Without flow (4), TPLC recognizes an illuminance increase of larger than the predetermined value and may falsely initiate time synchronization. Referring to the measured illuminance value after time synchronization in the previous step prevents false initialization of time synchronization.

4. Evaluation

4.1 Overview

We implemented TPLC on a wireless sensor node and examined synchronization errors. Experiments were conducted under the same conditions as those seen in Figure 2 and used two wireless sensor nodes. In order to examine synchronization errors in TPLC, wireless sensor nodes were placed in three different patterns. These patterns were: (A) Both wireless sensor nodes were placed directly under one lighting fixture. (B) One wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other. (C) Both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixtures.

First, in placement that both wireless sensor nodes were placed directly under one lighting fixture, only one lighting fixture was used. Thus, it was a placement that did not require wireless sensor nodes to consider the distance between the lighting fixture and wireless sensor nodes, nor the luminance changes at adjacent lighting fixtures. Figure 4 shows this placement.

Second, in placement that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other, one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other. Figure 5 shows this placement.

Third, in placement that both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixture, wireless sensor nodes were placed at the midpoint in the horizontal direction between two adjacent lighting fixtures. Figure 6 shows this placement.

Experiments were conducted with the aforementioned placement patterns to examine synchronization errors. In the experiments, time synchronization was performed through increasing and decreasing the illuminance of a lighting fixture (or lighting fixtures) by 15 lx every two seconds. For all three placement patterns, luminance was first increased and then decreased, and this increase/decrease in luminance was repeated a total of 1,000 times. The experiments used the FTSP's global time as the correct time. TPLC's synchronization error is defined as the time synchronization error of the global time obtained through comparing global times between sensor nodes when a wireless sensor node initiates time synchronization. Since the average synchronization error of FTSP was within 1 ms, the experiments did not consider synchronization errors that occurred in FTSP.

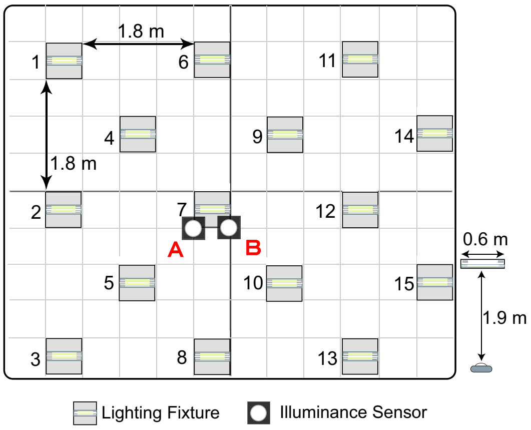


Figure 4: The experiment environment that both wireless sensor nodes were placed directly under one lighting fixture.

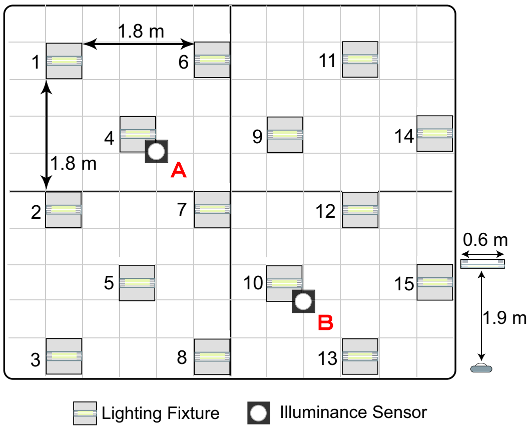


Figure 5: The experiment environment that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other.

4.2 Evaluation in TPLC when both wireless sensor nodes were placed directly under one lighting fixture

In the following, we examine synchronization errors in TPLC for wireless sensor node placement pattern that both wireless sensor nodes were placed directly under one lighting fixture. In placement that oth wireless sensor nodes were placed directly under one lighting fixture, wireless sensor nodes A and B were placed directly under lighting fixture 7 and only lighting fixture 7 was light controlled.

Figure 7 shows a distribution of synchronization errors. The horizontal axis shows the synchronization error [ms], and the vertical axis shows the probability of the error occurring. Table 1 summarizes the synchronization errors.

Synchronization errors depend on the length of the time intervals measuring illuminance. Time synchronization with TPLC becomes more precise as the length of the illuminance measurement time intervals decreases. However, the

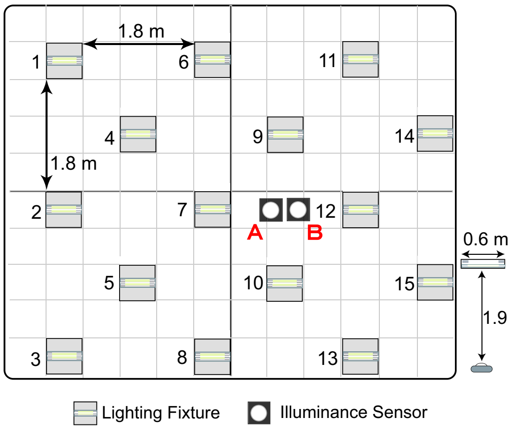


Figure 6: The experiment environment that both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixtures.

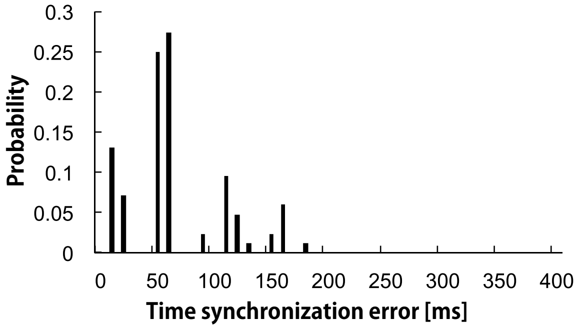


Figure 7: The distribution of synchronization errors in TPLC when both wireless sensor nodes were placed directly under one lighting fixture.

shorter the illuminance measurement time intervals are, the more power wireless sensor nodes consume. It is therefore necessary to design an algorithm that does not depend on the length of illuminance measurement time intervals.

4.3 Evaluation in TPLC when one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other

In the following, we examine synchronization errors in TPLC for wireless sensor node placement pattern that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other. Wireless sensor node A was placed directly under lighting fixture 4, and wireless sensor node B was placed directly under lighting fixture 10. For this experiment, only lighting fixtures 4 and 10 were light controlled. Figure 8 shows a distribution of synchronization errors. Table 2 summarizes the synchronization errors.

Table 1: The values of synchronization errors in TPLC when both wireless sensor nodes were placed directly under one lighting fixture.

synchronization error	value [ms]
average	62
maximum	172
minimum	4

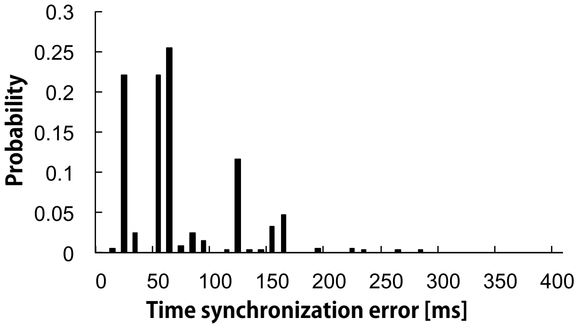


Figure 8: The distribution of synchronization errors in TPLC when one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other.

The average synchronization error for the placement that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other is similar to that seen in placement that both wireless sensor nodes were placed directly under one lighting fixture. Although the maximum error for the placement that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other is larger than that both wireless sensor nodes were placed directly under one lighting fixture, the average error for the placement that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other is similar to that seen in placement that both wireless sensor nodes were placed directly under one lighting fixture, and the error distribution for the placement that one wireless sensor node

Table 2: The values of synchronization errors in TPLC when one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other.

synchronization error	value [ms]
average	63
maximum	273
minimum	0

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

Table 3: The values of synchronization errors in TPLC when both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixture.

synchronization error	value [ms]
average	88
maximum	277
minimum	1

was placed directly under each of two lighting fixtures that were located at a distance from each other is almost identical to in placement pattern that both wireless sensor nodes were placed directly under one lighting fixture.

The spatial relation between the two lighting fixtures in the experiments was such that they did not affect each other. Thus, each of the two illuminance sensors independently initiated time synchronization based on the illuminance from its respective lighting fixture. The lighting fixtures used in placement that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other differed from the lighting fixture used in placement that oth wireless sensor nodes were placed directly under one lighting fixture. To perform light control on lighting fixtures, a light control signal was sent to each and every lighting fixture. The signal reached different lighting fixtures with different propagation delays.

In the experiments, the propagation delay of the light control signal could have introduced synchronization errors. However, the propagation delay of the light control signal for the laboratory used in the experiments was significantly smaller than the length of illuminance measurement time intervals, and thus, it did not impact synchronization errors. This explains how the experimental results for synchronization errors for the placement that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other are similar to those for placement pattern that both wireless sensor nodes were placed directly under one lighting fixture.

4.4 Evaluation in TPLC when both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixtures

In the following, we examine synchronization errors in TPLC for wireless sensor node placement pattern than both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixture. Wireless sensor nodes A and B were placed at the midpoint in the horizontal direction between lighting fixtures 7 and 12. For these experiments, only lighting fixtures 7 and 12 were light controlled. Figure 9 shows a distribution of the synchronization errors. Figure 10 shows a cumulative error distribution for 3 placement patterns. Table 3 summarizes the synchronization errors.

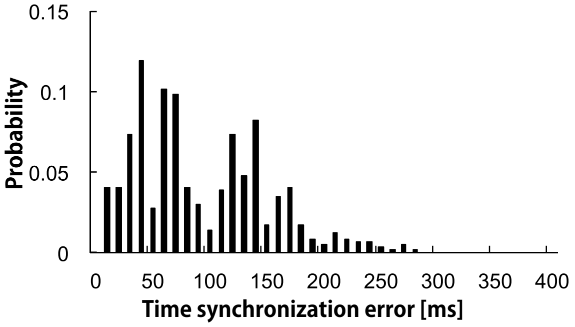


Figure 9: The distribution of synchronization errors in TPLC when both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixture.

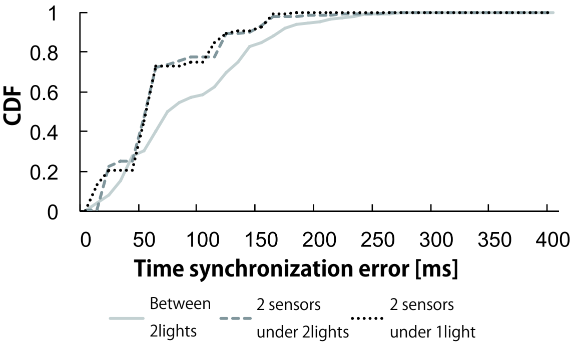


Figure 10: The cumulative error distributions for TPLC.

Figure 10 shows that the experiments for the placement than both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixture resulted in a cumulative error distribution that differs from those seen in placement that both wireless sensor nodes were placed directly under one lighting fixture and in placement that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other. With the placement than both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixture, there are less synchronization errors within 100 ms, and the average synchronization error is larger.

In the experiments, illuminance sensors were placed at the midpoint between two adjacent lighting fixtures that were close to each other. A luminance change of one lighting fixture affected the illuminance sensor placed directly under the other lighting fixture. As a luminance change at either lighting fixture affects the illuminance measured by illuminance sensors, it is possible that an increase in illuminance in the current step becomes different from that in the previous step, depending on when luminance changes occur at lighting fixtures. When a change in illuminance differs in each step, it significantly affects synchronization errors. For example, assume that 500 lx was referred to in the previous step. In the current step, whether time synchronization is initiated or not is determined when the measured illuminance approaches the referred value of 500 lx. Different lengths of time required for illuminance to reach 500 lx after the initial change in luminance clearly result in different times to initiate time synchronization. When each sensor node experiences different changes in measured illuminance in every step, it necessarily results in large synchronization errors. Due to this, we sure that synchronization errors are larger with the placement than both wireless sensor nodes were placed at the midpoint between two adjacent lighting fixture than with the placement that both wireless sensor nodes were placed directly under one lighting fixture and the placement that one wireless sensor node was placed directly under each of two lighting fixtures that were located at a distance from each other.

5. Conclusion and Future Improvements

This paper proposes Time synchronization Protocol based Lighting Control (TPLC), a time synchronization technique that does not require packet transmission or reception. TPLC varies luminance of a lighting fixture (or lighting fixtures) to cause changes in measured illuminance and uses the resulting illuminance changes to initiate time synchronization. By implementing TPLC on MOTE MICAz, we examined synchronization errors. When two wireless sensor nodes were placed directly under a single lighting fixture, TPLC achieved time synchronization with an average error of 62 ms, a maximum error of 172 ms and a minimum error of 4 ms. Experiments were also conducted for wireless sensor node placements where a wireless sensor node was placed directly under two lighting fixtures and where wireless sensor nodes were placed at the midpoint between two adjacent lighting fixtures. Experimental results for these placements show that TPLC achieved time synchronization, although it resulted in larger synchronization errors when compared to the placement where wireless sensor nodes were placed directly under a single lighting fixture.

Future research includes improving time synchronization to be more precise. Time synchronization becomes more precise as the length of illuminance measurement time intervals decreases. However, the shorter the illuminance measurement time intervals are, the more power wireless sensor nodes consume. One may explore an algorithm with varying illuminance measurement time intervals to search for the time when a luminance change completes. Such an algorithm may use shorter illuminance measurement time intervals near the time to initiate time synchronization, but longer intervals otherwise.

References

[1] Elson, J., Girod, L. and Estrin, D., “Fine-grained Network Time Synchronization using Reference Broadcasts,” Proceedings of the ACM Symposium on Networked Embedded Systems (SenSys), 2003.

[2] Ganeriwal, S., Kumar, R. and Srivastava, M. B., “Timing-sync Protocol for Sensor Networks,” Proceedings of the ACM Symposium on Networked Embedded Systems (SenSys), 2003.

[3] Hill, J.L., and Culler, D.E., “Mica: a wireless platform for deeply embedded networks,” IEEE micro, Vol.22, pp.12-24, 2002.

[4] Inoue, S., Mitsubishi Estate Company Ltd.,
“Towards the of city of the future,”
Avaliable: http://www.jetro.org/documents/green_innov/
Shigeru_Inoue_Presentation.pdf

[5] Maróti, M., Kusy, B., Simon, Gyula. and Lédeczi, Á., “The flooding time synchronization protocol,” Proceedings of the 2nd ACM Conference on Embedded Networked Sensor Systems, Baltimore, MD, USA, Nov., pp.39-49, 2004.

[6] MEMSIC, Wireless module, MICAz, Avaliable:
http://www.memsic.com/

[7] Miki, M., “An Intelligent Lighting System and the Consortium for Smart Office Environment,” Journal of Japanese Society for Artificial Intelligence, Vol.85 No.5, pp.346-351, 2001.

[8] Miki, M., Hiroyasu, T., and Imazato, K., “Proposal for an intelligent lighting system, and verification of control method effectiveness,” Proc IEEE CIS, pp.520-525, 2004.

[9] Mills, D.L., “Internet time synchronization : The network time protocol,” IEEE Trans. Commun., Vol.39, pp.1482-1493, 1991.

[10] Panasonic, Lights sensors: NaPiCa (AMS), Avaliable:
http://pewa.panasonic.com/components/built-in-sensors/light-sensors/napica/

[11] Shikakura, T., Morikawa, H., and Nakamura, Y., “Research on the Perception of Lighting Fluctuation in a Luminous Offices Environment,” Journal of the Illuminationg Engineering Institute of Japan, Vol.85 No.5, pp.346-351, 2001.

Lifetime extension algorithms for the connected sensor cover

Daisuke Matsumoto and Akihiro Fujiwara
Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan

Abstract— *In the sensor network, a set of connected sensors that covers discrete targets is called the connected sensor cover (CSC). Construction of the CSC reduces network energy and communication costs. However, connectivity of the CSC is lost in case that a sensor is down because of low battery. In this case, removing of the sensor with low battery from CSC and addition of another sensor enable extension of lifetime of the CSC.*

In the present paper, we first propose a centralized algorithm for the lifetime extension using an artificial bee colony optimization technique. We also propose a distributed algorithm for the lifetime extension of the CSC. The experimental results show the effectiveness of the two proposed algorithms.

Keywords: sensor network, connected sensor cover, bee colony optimization

1. Introduction

In the sensor network, a set of sensors is called *connected* if any two sensors in the set can communicate using the other sensors, and the sensor network can gather data of using the connected sensors. In addition, a set of sensors is called *cover* if a target region is completely covered by union of sensing area of sensors in the set. Thus, the connected sensor cover (CSC) is defined as a set of connected sensors that covers a target region. Although construction of the CSC with the minimum number of sensors is NP-hard [4], the CSC with a small number of sensors is desirable because construction of the small CSC can reduce network energy and communication costs. Therefore, a number of algorithms [2], [3], [4], [6], [7] are proposed for constructing the CSC with a small number of sensors.

On the other hand, each sensor generally works with non-rechargeable battery, and initial voltages of batteries may be different between sensors in the network. Then, connectivity of the CSC may be lost in case that a sensor is down because of low battery. In this case, we can maintain the CSC by removing the sensor with low battery from the CSC and adding another sensor to the CSC. We assume that lifetime of the CSC is an interval such that connectivity and coverage of the CSC are satisfied by removal and addition of sensors, and consider extension of lifetime of the CSC. Since difference of remaining battery life is not considered in the above known algorithms [2], [3], [4], [6], [7], the constructed CSCs may have short lifetime in the known algorithms.

In the present paper, we propose centralized and distributed algorithms for extending lifetime of the CSC, which covers discrete targets, for sensors that have different remaining batteries life. We first propose a centralized algorithm for the lifetime extension using an artificial bee colony optimization technique. We next proposed a distributed algorithm for the lifetime extension. In the distributed algorithm, each sensor independently determines one of three modes, which are active, relay and sleep, using informations of adjacent sensors only.

We implement our proposed algorithm and a basic centralized algorithm for extending lifetime of the CSC in simulation environment, and compare the number of sensors in the obtained CSC. The results show that our proposed centralized algorithm obtains a CSC with a smaller number of sensors, and our proposed distributed algorithms also obtains reasonable size of the CSC.

2. Preliminaries

2.1 System model

In this paper, the sensor network $G = (V, E)$ is defined by a set of sensors $V = \{s_1, s_2, \ldots, s_n\}$, where n is the number of sensors, and the set of links E that is a set of communication links between sensors. The set of discrete targets is represented by $T = \{t_1, t_2, \ldots, t_m\}$, where m is the number of discrete targets. Each sensor s_i has an unique identification number ID_i. The sensors are deployed on two-dimensional plane R, and each sensor knows the geographical location of itself. In addition, we also assume that each sensor s_i has different initial battery charge b_i. If $b_i \leq 0$, the sensor becomes inoperable by the shortage of battery.

Figure 1 is the sensor model used in this paper. Each sensor s_i has communicating area C_i and can communicate with other sensors in the communication area. In case that sensors s_i and s_j can communicate each other, a communication link $e_{ij} \in E$ exists between the sensors s_i and s_j on a graph G, and the sensors are called adjacent. We assume that direct communication of messages is possible for only between adjacent sensors without collision. In addition, two

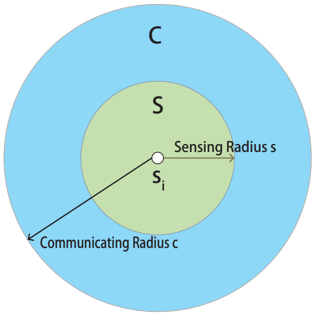


Fig. 1: Sensor model

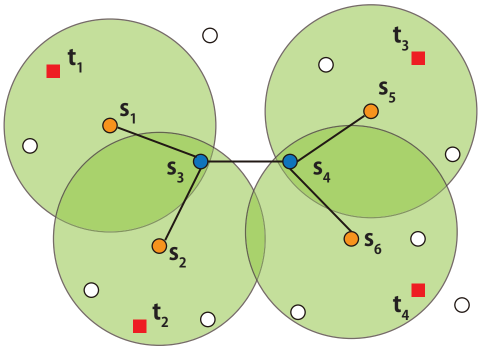


Fig. 2: Example of CSCDT

sensors are called connected if there is a path of adjacent sensors between two sensors in G.

We also assume that each sensor s_i has a circular sensing area S_i, and can sense a target in the sensing area. We call that a sensor s_i covers a target t if t is in sensing area of s_i.

In addition, each sensor has three modes, which are called active, relay, and sleep. Available functions and power consumptions are different in each mode as follows.

Active mode:

- Sensing and communication functions are available.
- Energy consumption is large.

Relay mode:

- Communication function is available.
- Energy consumption is slightly less than the active mode.

Sleep mode:

- Sensing and communication functions are disabled.
- Energy consumption is small.

Let c_a, c_r and c_s be energy consumptions of active, relay and sleep modes, respectively. Then, the following condition holds.

$$c_a > c_r \gg c_s \tag{1}$$

2.2 Connected sensor cover for discrete targets

Since each sensor equips a limited battery, coverage with a fewer number of sensors should be desired for discrete targets. However, connectivity of sensors is needed to communicate data of targets in the sensor network. As a set of sensors which satisfies these conditions, the connected sensor cover is defined as follows.

Definition 1 (Connected sensor cover for discrete targets): Given a sensor network $G = (V, E)$ and a set of discrete targets $T = \{t_1, t_2, \ldots, t_m\}$, the subset of sensors which satisfies the following condition $M = \{s_{i_1}, s_{i_2}, \ldots, s_{i_n}\}$ $(M \subseteq V)$ is called the connected sensor cover for the discrete targets (CSCDT).

1) M is connected.
2) Any target in T is in at least one of sensing areas of sensors in M. □

Targets covering problems using line segments are known as NP-hard, and CSCDT with the minimum number of sensors is also known as NP-hard by reducing from the problem [3]. Therefore, a smaller number of sensors should be desired for CSCDT.

In this paper, we assume that the number of sensors n are enough large to construct CSCDT. In other words, CSCDT, which covers a set of input discrete targets, always exits for an input set of sensors.

Figure 2 shows an example of CSCDT. In this example, the set of sensors $\{s_1, s_2, \cdots, s_6\}$ is CSCDT, which covers a set of discrete targets $\{t_1, t_2, t_3, t_4\}$ with an union of sensing areas of sensors $\{s_1, s_2, s_5, s_6\}$. The set of sensors maintains connectivity of sensors by including a set of sensors $\{s_3, s_4\}$.

2.3 Network lifetime

The network lifetime is an interval such that a set of sensors can monitor all discrete targets and collect informations continuously. We first define energy consumption in the sensor network.

Definition 2 (Energy consumption in the sensor network): In the sensor network, each sensor executes a given procedure in each step, and then, consumes battery charge defined by each mode. In other words, each sensor s_i reduces c_a, c_r, or c_s from the battery charge b_i according to the modes. □

Using Definition 2, the network lifetime is defined as follows.

Definition 3 (Network Lifetime): Let t_s be the step such that the set of sensors M starts to satisfy the condition of CSCDT. By executing algorithms, sensors in M consume batteries and are changed dynamically. We also assume that t_e is the first step such that M does not satisfy the condition of CSCDT. Then, the network lifetime is defined as $t_e - t_s$, which is the difference between t_e and t_s. □

3. Centralized Algorithm

3.1 A sub-procedure for deleting a sensor

We first define unnecessary sensors for CSCDT, and next propose a procedure, which deletes the unnecessary sensor, as a basic operation in the proposed algorithm.

Definition 4 (An unnecessary sensor for CSCDT): Let S_i be sensing area of sensor s_i in the sensor network $G = (V, E)$, and also let $M \subseteq V$ be CSCDT for the sensor network. We also assume that T is a set of discrete targets. Then, a sensor $s_d \in M$ is an unnecessary sensor for M if the sensor satisfies the following two conditions.

1) $M - \{s_d\}$ is connected.
2) Any target in T is in at least one of sensing areas of sensors in $M - \{s_d\}$. □

Using the above definition, we proposed a procedure, which is called *Simple deletion*, for deleting a sensor from CSCDT.

Simple deletion

Step 1:Check whether a sensor is an unnecessary sensor or not.

Step 2:Delete the sensor in case that the sensor is unnecessary. □

The procedure, *Simple deletion*, ensures connectivity and coverage of the sensor network.

Figure 3 shows an example for unnecessary sensors and the procedure. Figure 3 (a) shows an input CSCDT, and the set of sensors $\{s_1, s_2, s_3, s_4\}$ is in CSCDT. An union of sensing areas of the set of sensors $\{s_1, s_3\}$ covers the set of discrete targets $\{t_1, t_2, t_3, t_4\}$, and the sensors are connected. In this case, sensor s_4 is deleted if *Simple deletion* is executed for s_4, and we obtain CSCDT in Figure 3 (b).

3.2 A centralized algorithm using Simple deletion

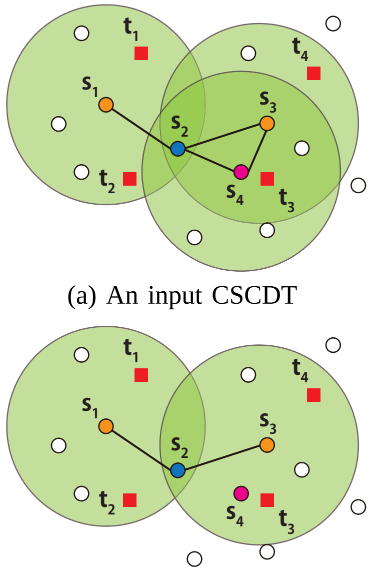
We now introduce a basic centralized algorithm which constructs CSCDT. We can obtain the minimal CSCDT by a repetition of *Simple deletion*. (If no sensor can be deleted by application of *Simple deletion*, the CSCDT is the minimal.)

For example, after removing sensor s_4 for CSCDT in Figure 3 (b), no sensor can be removed by *Simple deletion*. Then, CSCDT in Figure 3 (b) is the minimal CSCDT.

Since battery charge is not considered in the above algorithm, the algorithm cannot achieve lifetime extension of the sensor network. We next propose a modified optimization algorithm that consider the battery charge.

3.3 A centralized algorithm using bee colony optimization

We now propose our modified algorithm that consider the battery charge. The algorithm computes CSCDT using the bee colony optimization [8]. The Bee colony optimization is an optimization technique based on the following habit of honey bees. The bee gathers honey outside region, and shares information of gathered honey with other bees when the bee arrives at comb. Then, each bee decides a next way of exploration using exchanged informations. As a result, the honey bees can collect high-quality honey.



(a) An input CSCDT

(b) An output CSCDT after *Simple deletion*

Fig. 3: An example of *Simple deletion*

The basic idea of the algorithm is based on the centralized algorithm using *Simple deletion*, and the bee colony optimization is used for reducing the number of sensors. In the following algorithm, b_i denotes the initial battery charge of sensor s_i $(1 \leq i \leq n)$. We also assume m bees, B_j $(1 \leq j \leq m)$, are used to optimize CSCDT.

A centralized algorithm using bee colony optimization

Step 1:Each bee B_j computes $CSCDT_j$ for an input set of sensors using a centralized algorithm using *Simple deletion*. Since *Simple deletion* is applied in random order for each bee, different bees probabilistically obtain different CSCDTs.

Step 2:The following 3 sub-steps are repeated by a given number of trials, and output a CSCDT with the minimum number of sensors among all bees.

(2-1) Each bee B_j computes the number of sensors N_j in $CSCDT_j$, and also computes the maximum difference between battery charge of sensors in $CSCDT_j$ as follows.

$$V_j = \max\{|b_x - b_y| \mid 1 \leq x, y \leq n\}$$

Then, each bee B_j computes evaluation value O_j using N_j, V_j, and two weight parameters, α and β, as follows.

$$O_i = N_i^{\alpha} \times V_j^{\beta}$$

(2-2) Each bee B_j decides whether to become a recruiter or a follower according to the following probability F_j, where $O_{\min} = \min\{O_k \mid 1 \leq k \leq n\}$ and u is the number of repetition in Step 2.

$$F_j = e^{-\frac{O_j - O_{\min}}{u}}$$

(2-3) Each bee B_j maintains $CSCDT_j$ in case of the recruiter. On the other hand, in case of the follower, B_j selects one of solutions computed by recruiters according to following probability R_i, where R is a set of recruiters.

$$R_j = 1 - \frac{O_j}{\sum_{O_k \in R} O_k}$$

□

4. Distributed Algorithm

In this section, we propose a distributed algorithm that extends the network lifetime. We assume that first CSCDT is constructed using a known distributed algorithm [5]. After construction of the CSCDT, our proposed distributed algorithm is started for extending network lifetime. In the proposed algorithm, each sensor executes the algorithm, and decides one of modes and an interval of sleep.

The algorithm consists of the following 2 phases. The second phase is executed on a sensor that is switched to the sleep mode in the next step.

Phase 1On each sensor, decide a next mode from one of active, relay, and sleep modes.

Phase 2On each sensor, decide an interval of the sleep in case that the next mode is a sleep mode.

In our proposed algorithm, we assume that each sensor s_i knows a set of adjacent sensors S_{N_i}, which is a set of sensors in the communication area of s_i, and T_i, which is a set of discrete targets in sensing area of s_i. In addition, we also assume that sensor s_i also knows the following information for $s_k \in S_{N_i}$.

$Mode_k$$Mode_k$ is a mode of sensor s_k, and then, $Mode_k \in \{$ active, relay, sleep $\}$.

S_{N_k}: S_{N_k} is a set of adjacent sensors, which are in the communication area of s_k. (s_k is not included in S_{N_k}.)

T_k: T_k is a set of discrete targets in sensing area of s_k.

In the following description, SA_i denotes a set of sensors in S_{N_i} in the active mode. We also assume that $T_{N_i} = T_i \cup \bigcup_{s_k \in S_{N_i}} T_k$.

4.1 Decision of mode

In this section, we explain an algorithm for the first phase that decides a next mode of each sensor. Each sensor execute one of the following (A), (B), and (C), according to the current mode.

(A) Active

If sensor s_i is in the active mode, and satisfies the following three conditions, the next mode of s_i is sleep.

1) T_i is included in an union of sensing areas of sensors in SA_i.
2) S_{N_i} is connected.
3) No sensor $s_k \in S_{N_i}$, whose ID_k is smaller than ID_i, changes the mode in the current step.

Next, if sensor s_i does not satisfy the above conditions, but satisfies the following two conditions, the next mode of s_i is relay.

1) T_i is included in an union of sensing areas of sensors in SA_i.
2) No sensor $s_k \in S_{N_i}$, whose ID_k is smaller than ID_i, changes the mode in the current step.

In the other cases, the next mode of s_i is still active.

(B) Relay

If sensor s_i is in the relay mode, and satisfies the following three conditions, the next mode of s_i is sleep.

1) T_i is included in an union of sensing areas of sensors in SA_i.
2) S_{N_i} is connected.
3) No sensor $s_k \in S_{N_i}$, whose ID_k is smaller than ID_i, changes the mode in the current step.

Next, if sensor s_i does not satisfy the above conditions, but satisfies the following condition, the next mode of s_i is active.

1) T_i is not included in an union of sensing areas of sensors in SA_i.

In the other cases, the next mode of s_i is still relay.

(C) Sleep

A sensor s_i in the sleep mode maintains a variable SI_i, which denotes counter for sleep. (The variable SI_i is set in Phase 2.) If $SI_i > 0$, the next mode of s_i is still sleep, and set $SI_i = SI_i - 1$. On the other hand, if $SI_i = 0$, the next mode of s_i is changed to the mode previously decided in Phase 2. (The decision of the mode is described later.) □

Figure 4 shows an example of the above algorithm. Orange and yellow circles denote sensors in the active mode, blue and pink circles denote sensors in the relay mode, and yellow and pink sensors denote sensors that decide a mode in the next step. We assume that ID of the yellow sensor is smaller than ID of the pink sensor.

In this example, we assume that yellow and pink sensors try to decide the mode at the same time. Since the yellow sensor satisfies the conditions for changing to the relay mode, the next mode of the yellow sensor is the relay mode. On the other hand, ID of the pink sensor is larger than ID of the yellow sensor, and the pink sensor does not satisfy the conditions for changing to the sleep mode.

4.2 Decision of sleep interval

The second phase, which decide an interval of the sleep, is executed on only a sensor that decides the next mode is

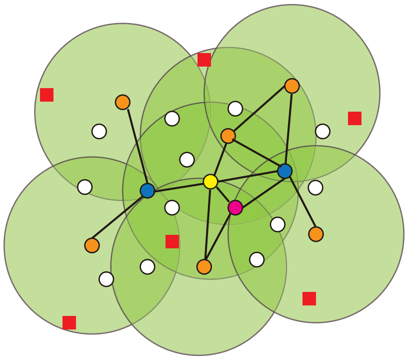


Fig. 4: Example of Mode Changing

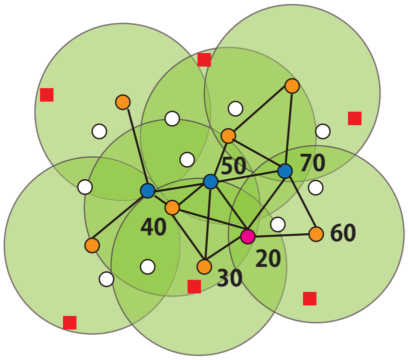


Fig. 5: Initial State

sleep. In this section, we explain a method that decide a sleep interval SI_i on each sensor s_i using informations on s_i and S_{N_i}.

We now introduce a procedure for deciding the SI_i on sensor s_i. In the description, SS_i denotes a set of adjacent sensors of the adjacent sensor of s_i, which is given as follows.

$$SS_i = \left(\bigcup_{s_k \in S_{N_i}} S_{N_k} \right) - \left(\bigcup_{s_h \in S_i} S_h \right) - \{s_i\}$$

In addition, we define the following value L_i, which denotes lifetime for each sensor s_i.

$$L_i = \frac{b_i}{c_i}$$

In the above equation, b_i denotes a battery charge of sensor s_i, and c_i denotes energy consumption, which is defined in Section II. (The energy consumption is varied according to the mode.)

A procedure for deciding sleep interval

Step 1:Set $TS_i = S_{N_i}$, and compute a sensor $s_{min} \in TS_i$ such that $L_{min} = \min\{L_h \mid s_h \in TS_i\}$

Step 2:If s_{min} satisfies the following three conditions, s_{min} is removed from TS_i, and then, repeat Step 2.

(c1) T_i is included in an union of sensing areas of sensors in TS_i.

(c2) TS_i is connected.

(c3) Any sensor in SS_i is connected to a sensor in TS_i.

On the other hand, if one of the above conditions is not satisfied, SI_i is set to $L_{min} - 1$. In addition, set $WM_i = active$ in case that non-satisfied condition is (c1), otherwise, set $WM_i = relay$. □

The sensor s_i in sleep ends the mode when $SI_i = 0$, and switch to the one of two modes, active or relay. The mode is decided according the a value of WM_i. The sensor switches to the active mode in case of $WM_i = active$, otherwise, the sensor becomes the relay mode.

Figure 5 shows an example of the above procedure. White circles denote sensors in the sleep mode, orange circles denote sensors in the active mode, and blue and pink circles denote sensors in the relay mode. In this case, the pink sensor s_i is trying to switch to the sleep mode. We assume that the remaining lifetime of the pink sensor is 20, and the remaining lifetime of adjacent sensors of it is 40, 30, 50, 70, 60 as described in the figure.

In this example, SI_i is set to 39 because the condition (c1) is not satisfied if an adjacent sensor, whose lifetime is 40, is removed. In addition, s_i switches to the active mode when $SI_i = 0$ because $WM_i = active$.

5. Experimental Results

Our three algorithms, which are a centralized algorithm using *Simple deletion*, a centralized algorithm using bee colony optimization, and a distributed algorithm, are implemented in a simulation environment using a software library LEDA [1], and we compare network lifetimes between the algorithms.

5.1 Simulation model

In our simulation model, input sensors and discrete targets are randomly located in 100×100 square area. The number of input sensors is 200, and the number of discrete targets is 20. The number of trials of algorithms using bee colony optimization is 100, and initial battery charge of sensors are random values between 12500 and 25000. In addition, battery charges consumed on each sensor in one step are 38, 33 and 0 for active, relay and sleep modes, respectively.

We assume that the sensing and communication radiuses are the same, and the radius of each sensor is a value between 20 and 40.

5.2 Simulation results

We now show results for centralized and distributed algorithm in the simulation environment. Figure 6 shows network lifetimes by the proposed algorithms. The vertical

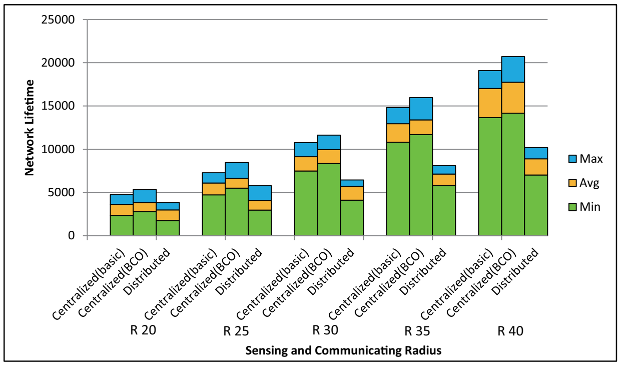


Fig. 6: Network lifetime of proposed algorithms

axis denotes the network lifetime, and the horizontal axis denotes sensing and the communicating radiuses.

The result between two centralized algorithms implies that our algorithms using bee colony optimization achieves better network lifetime than the algorithm using *Simple deletion*. On the other hand, network lifetime obtained by the distributed algorithm 0.6 times shorter than the algorithm using bee colony optimization. Although the centralized algorithms can reduce sensors using information of constructed CSCDT, the distributed algorithm only uses informations in adjacent sensors. In addition, sensors in sleep mode may switch its mode at an unpredictable step in the distributed algorithm, and the unpredictable behavior shorten the network lifetime of the distributed algorithm.

6. Conclusions

In this paper, we proposed centralized and distributed algorithms that construct CSCDT to extend the network lifetime. We showed both of the proposed algorithms can extend network lifetime, and distributed algorithm achieves a litter shorter network lifetime compared with the centralized algorithm.

In the future work, we consider to improve our proposed algorithms for further extension of the network lifetime.

References

[1] Leda - algorithmic solutions software gmbh. `http://www.algorithmic-solutions.com/leda/`.

[2] S. Begum, N. Tera, and S. Sultana. Energy-efficient target coverage in wireless sensor networks based on modified ant colony. *International Journal of Ad hoc, Sensor & Ubiquitous Computing*, 1(4), 2010.

[3] I. Cardei and M. Cardei. Energy-efficient connected-coverage in wireless sensor networks. *International Journal of Sensor Networks*, 3(3), 2008.

[4] H. Gupta, Z. Zhou, S. Das, and Q. Gu. Connected sensor cover:self-organization of sensor networks for efficient query execution. *ACM/IEEE Transactions on Networking*, 14(1), 2006.

[5] K. Kawachi and A. Fujiwara. The fault tolerant connected sensor cover algorithm for discrete targets. *International Conference on Networking and Computing*, 2011.

[6] M. Lu, J. Wu, M. Cardei, and M. Li. Energy-efficient connected coverage of discrete targets in wireless sensor networks. *International Journal of Ad Hoc and Ubiquitous Computing*, 2009.

[7] K. Nakamoto and A. Fujiwara. Distributed algorithms for 2-connected sensor cover in sensor network. In *Proceedings of the International Conference on Wireless Networks*, 2010.

[8] D. Teodorovic, T. Davidovic, and M. Selmic. Bee colony optimization: The applications survey. *ACM Transactions on Computational Logic*, 2011.

Performance of Hierarchical Polling-based MAC scheme for Wireless Body Sensor Network

S. Motoyama
Faculty of Campo Limpo Paulista - FACCAMP
Master Program in Computer Science
Campo Limpo Paulista, S. Paulo, Brazil

Abstract - *The MAC scheme called hierarchical polling-based access scheme for Wireless Body Sensor Network (WBSN), proposed in [24], is analyzed in this paper. The MAC scheme, proposed in [24], is structured in hierarchy to collect data from sensor nodes inserted in human body. In first level of hierarchy the sensor nodes are divided into groups and sensor nodes of each group communicate with a sink node which collects data by using polling technique. In second level, the sink nodes communicate with a master node which collects data by also using polling technique. The performance analysis carried out in [24] considered that the sensor nodes from first level are provided with only single buffer to store data and the sink nodes have infinite size buffers. The master node used exhaustive polling technique. In this paper, the sensor nodes and sink nodes are provided with infinite size buffers and both first and second levels use exhaustive polling technique. The study is carried out using mathematical models known in the literature and two cases are compared.*

Keywords: wireless, body sensor network, MAC, polling, mathematical modeling.

1 Introduction

Wireless Body Sensor Network (WBSN) is composed of tiny electronic devices called sensors which are attached to the human body for remote monitoring of vital signs. A sensor with processing and communication capabilities is denoted sensor node.

Since the sensor nodes of a WBSN can be placed under the human skin of difficult accesses and due to the small size of the nodes and the limited energy storage capacity of battery, the sensor nodes must mainly save energy.

One of the tasks performed by sensor node that most spends energy is the communication. The sensor nodes must communicate externally with some device (sink node) for the transmission of collected data. Since many sensor nodes can be placed at human body, if more than a sensor node begins to transmit packets simultaneously, collisions will occur, and packets must be retransmitted. The packet retransmission can be an energy consuming process. Thus, an efficient medium access control (MAC) mechanism for collision reduction or elimination is fundamental for good operation of a WBSN. Furthermore, the use of sink nodes as centralized nodes for data collection from sensor nodes is more convenient because it simplifies the communication protocol, and it is appropriate for the collision reduction.

A MAC scheme based on polling technique and using sink nodes in a hierarchical structure was proposed in [24]. The performance analysis in that paper was carried out using single buffer at the sensors. In this paper, the performance is carried out using infinite size buffers at sensors and exhaustive service polling.

This paper is divided into five sections. In the second section, the related works to this paper are presented. The hierarchical polling based MAC scheme proposed in [24] is described in section three. In the fourth section, the mathematical modeling and performance analyses of proposed MAC scheme are carried out. Finally, the main conclusions are presented in section five.

2 Related Works

Many MAC schemes proposed in the literature for the WBSN are based on the standard 802.15.4 with beacon-enabled star configuration which provides very low energy consumption [1]. However, since the scheme is not designed for WBSN applications some drawbacks have been pointed out [2], and recently many schemes of MAC protocols specifically for WBSN have been proposed [2-16]. Some proposals are based on the variations of standard 802.5.4 [5], [8] and [11], and others are based on TDMA access technique [3], [4], [7], [10], [14], [15] and [16]. Each of the proposals explores some special features based on medical needs. For instance, in [3-4] to deal with the light and heavy loads in normal and urgent situations, a context aware MAC is proposed. To guarantee QoS of a WBSN, a MAC protocol based on random access technique is proposed in [12]. In the proposal presented in [10], the heart beating is used for the purpose of clock synchronization. In [6], the beacon used for wake-up sensor nodes is used for battery charging, increasing the network life time.

Recently, the standard 802.15.6 has been proposed for the wireless body area network [17]. This standard has three modes of operation: beacon mode with beacon period superframe boundaries, non-beacon mode with superframe boundaries and non-beacon mode without superframe boundaries [17]. The beacon mode is designed for medical and non-medical applications and has been the object of main standardization.

The non-beacon mode without superframe boundaries has been less explored. In [20] and [21] MAC schemes using this mode were proposed. Both proposals are based on polling access scheme that avoids the need for periodical synchronization. In [20] a flexible a scheme that exchanges the normal polling operation mode to the urgent polling operation mode in case of emergency needs is proposed. In [21] the polling scheme using realistic sensor node models for WBSN are investigated by simulation. In [23] a MAC scheme using hierarchical topology based on TDMA technique was proposed for WBSN. The MAC scheme proposed in [24] has also a hierarchical topology but instead the use of TDMA is based on polling technique.

The main objective of this paper is to study the MAC scheme proposed in [24] by using different buffer size at the sensors; instead using single buffer size, in this case an infinite buffer size is analyzed and the two cases are compared.

3 MAC scheme based on hierarchical structure

The MAC scheme proposed in [24] is shown in Fig. 1. Two or more sink nodes are placed in a belt at different locations, so that a group of sensor nodes at back can communicate with the sink node located at back and a group of sensor nodes in front can communicate with sink node placed at front. To collect the data from sink nodes it is provided another node denoted master node. To collect data, the sink nodes as well master node use the polling technique. This structure is denoted hierarchical polling-based access scheme.

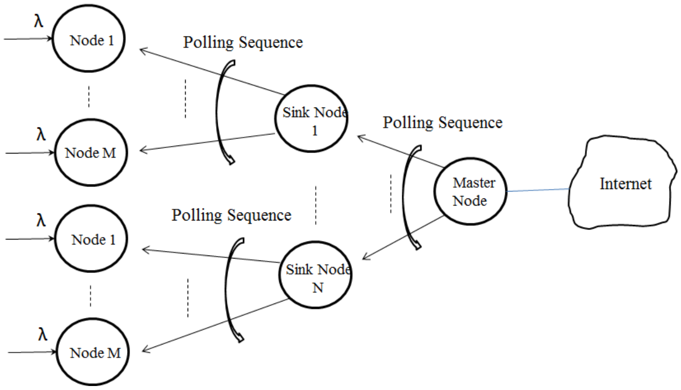


Figure 1. Hierarchical polling-based structure.

The communication protocol for hierarchical polling-based access scheme proposed in this paper can be simplified using the fact that the sensors are located in close proximity to the sink node. The sink node broadcasts a packet carrying the sensor node number to be investigated, i.e., it is sending an authorization to a sensor node to transmit the packets. This authorization packet has in its header enough bits for bit and frame synchronizations of a sensor node. If a sensor node has packets to transmit, it recognizes its sensor node number and starts to transmit. After the transmission, the sensor node waits for acknowledgment in case of the need for retransmission. If a sensor node doesn't have packets to transmit, it can keep the transceiver in an off state and only switches to an on state in the case of packet transmission. The sink node recognizes that a sensor node is in off state after the transmission of an authorization packet and waits for a while. If the data packet from the polled sensor node doesn't arrive, the sink node infers that the node doesn't have packets to transmit and goes to other sensor node in sequence to poll. Thus, in this proposed protocol, the sink node does almost all of the communication functions, leaving the sensor node only the packet transmission function. This same communication protocol can be used in second level, that is, when the master node polls sink nodes to get the data. For WBSN, just two hierarchical levels may be enough.

A sensor node can save energy by keeping the transceiver in an off state when is not transmitting packets. Another way to save sensor node energy is to implement functions at the node in which the sensors send only relevant information to the event observer. For instance, a sensor monitoring body temperature sends only measurements which are above a certain value. The other criterion could be to transmit just the packets that are outside of a certain range.

4 Performance Evaluation

The analysis for infinite buffer case is based on same assumptions used in [24], that is, Poisson distribution of rate λ packets/sec for output of each sensor node and a deterministic packet length distribution with average *E{X}* bits long for all nodes are adopted. The channel capacity from sensor nodes to the sink node (or vice versa) or sink nodes to the master node (or vice versa) is *R* bits/sec. The number of sensor nodes in each group will be considered M and the number of sink nodes is N.

The walk time, *w*, between two consecutive sensor nodes in the polling is constant and same for all nodes. The propagation time of a sensor node to the sink node is same for all nodes and is included in walk time.

For the performance analysis of hierarchical polling, it can be considered that each group of sensor nodes and a sink node together is independent of each other, so that each group can be analyzed independently.

4.1 First Level – exhaustive service case

The following expressions can be written for infinite buffer case. The average cycle time is given by

$$T_{c1} = \frac{Mw}{(1-S_1)}, \qquad (1)$$

where S_1 is given by

$$S_1 = \frac{M\lambda E\{X\}}{R} \qquad (2)$$

The queuing time in a buffer in the first level is given by [22]

$$E\{W_1\} = \frac{Mw(1-S_1/M)}{2(1-S_1)} + \frac{S_1E\{X\}}{2R(1-S_1)} \qquad (3)$$

for deterministic packet length.

The packet transfer time for the first level is given by

$$E\{T_1\} = \frac{E\{X\}}{R} + E\{W_1\}\,. \qquad (4)$$

For illustration of the above equations, the same numerical values of [24] will be used, that is, the packet length of $E\{X\}$ = 900 bits, the number of sensors of M = 10, 20 and 30, the channel capacity from nodes to sink node or vice-versa of R = 20 kbps and the authorization packet length of 10% of data packet $E\{X\}$. Furthermore, the walk time is assumed to be 6.5 msec.

Figure 2 shows the packet transfer time for various values of M. For load up to 0.8 the transfer times are low keeping below 600 milliseconds for any value of M. For loads greater than 0.8, the transfer times have very large values.

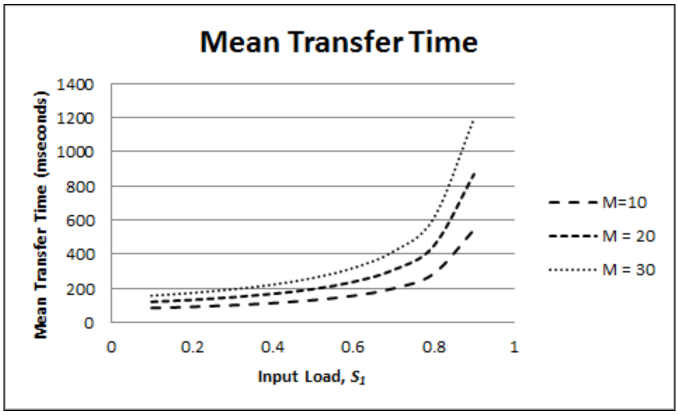


Figure 2. Mean transfer time for first level with infinite buffer.

The mean cycle time is shown in Fig. 3 for various values of M. For load less than 0.7 the polling system is very stable, keeping the cycle time less than 700 ms for any value of M. However, for load greater than 0.8 the system is becoming unstable with very large cycle time.

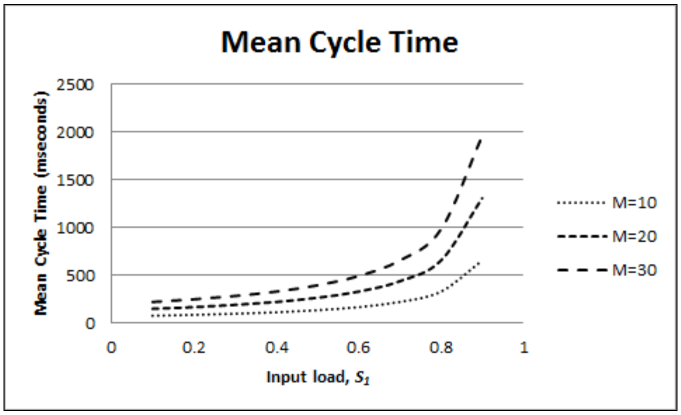


Figure 3. Mean cycle time for first level with infinite buffer.

4.2 Second Level – exhaustive service case

Since there is no loss in first level, the performance model for second level will be as shown in Fig. 4.

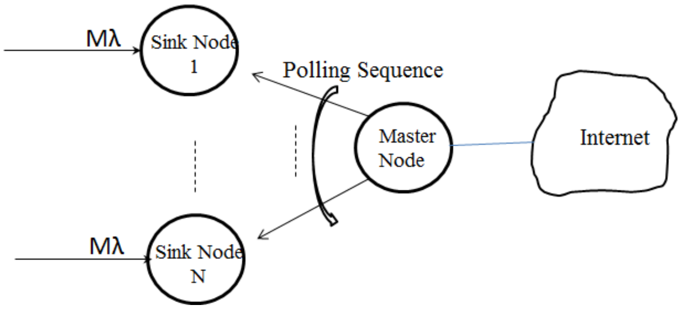


Figure 4. Second level performance model for infinite buffer case.

The mean cycle time for this case is given by

$$T_{c2} = \frac{Nw}{(1-S_2)}, \qquad (5)$$

where S_2 is given by

$$S_2 = NME\{X\}/R = S_1N, \qquad (6)$$

and S_1 is the total load of a group of first level as defined in Eq. 2.

The stability condition is given by

$$S_2 < 1 \Rightarrow NM\lambda < R/E\{X\}. \qquad (7)$$

The queuing time in a buffer in the second level is given by

$$E\{W_2\} = \frac{Nw(1-S_2/N)}{2(1-S_2)} + \frac{S_2E\{X\}}{2R(1-S_2)}, \qquad (8)$$

for deterministic packet length.

The packet transfer time for the second level is given by

$$E\{T_2\} = \frac{E\{X\}}{R} + E\{W_2\}\,. \qquad (9)$$

Figures 5 and 6 show the mean transfer time and mean cycle time, respectively, for N =2 and various values of M. Since the load, S_2, and the queuing time, $E\{W_2\}$, as defined in Eqs. 8 and 9, are only in function of N, the curves in Figs. 14 and 15 are same for M = 10, 20 and 30. However, the meaning of values in Y axes is different for each input load. For instance, for an input load of S_1 = 0.4, the value of transfer time for M = 10 is 154,5 milliseconds. This same value is found for M = 20 or 30, but the packet arrival rate, λ, is half or one third of M = 10, obeying the expression λ = S_1R/ME{X}. The same meaning can be given to the mean cycle time of Fig. 6.

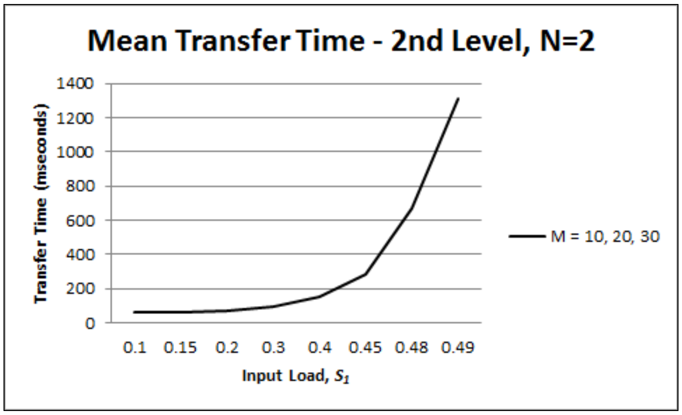


Figure 5. Second level mean transfer time versus first level input load of a group.

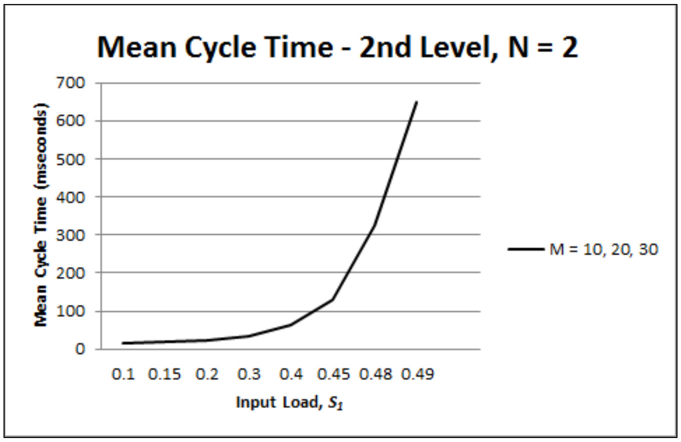


Figure 6. Second level mean cycle time versus first level input load of a group.

Figure 7 shows the total transfer time, including the first and second levels. It can be noticed that an input load up to 0.4 the polling system is stable, with total transfer time less than 400 msec, regardless the value of M. However, for input load greater than 0.4 the transfer time increases very fast, becoming unstable. Comparing Fig. 7 to Fig. 2, it can be observed that although the first level could accept a high load, indeed, this load is limited to the transfer times of the second level.

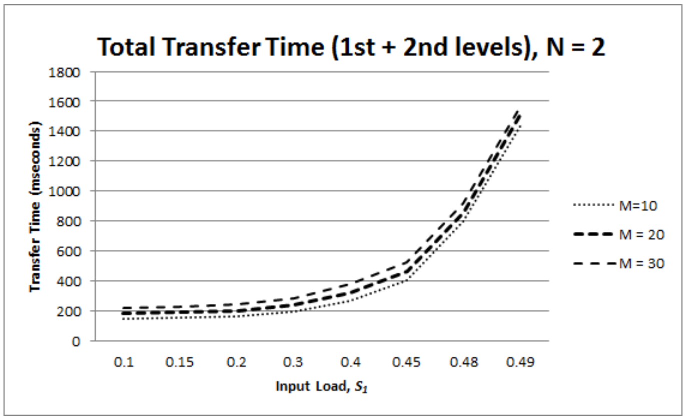


Figure 7. Total mean transfer time including first and second levels in function of input load of a group.

4.3 Single and exhaustive service cases comparison

Figures 8, 9 and 10 are the mean transfer time for first level, the loss probability for first level and total mean transfer for first and second level, respectively, considering a single buffer obtained in [24].

The analysis shows that for the first level using single buffer the transfer times can be kept very small as can be seen in Fig. 8. However, packet loss for high load (above 0.4) is prohibitive and must be avoided, as seen in Fig. 9. For the infinite buffer case in the first level using exhaustive polling technique, the transfer time can be kept below 600 msec for load as high as 0.8, however, the load is limited in function of the performance of second level, as seen in Fig. 7. The analysis for second level using exhaustive service is dependent of number of sink nodes and also the number of sensor nodes of first level. Considering only two sink nodes, which we consider appropriate for WBSN, for single buffer case considering any value of M (10, 20 and 30) and input load S_1 of up to 0.5, the transfer times are less than 350 ms for all cases, as seen in Fig. 10. However, for values above 0.5 the operation is becoming unstable and the transfer times are very larger. The total transfer times considering the first and second levels for load up to 0.5 are less than 500 ms for all cases, as seen in Fig.11. For infinite buffer case the total transfer time including first and second levels is very sensitive to the number of sensor nodes of first level. The load up to 0.4 the transfer times are less than 400 msec and the polling system is stable, as seen in Fig. 7. However, for load above 0.4 the system is becoming unstable and a small increment in the load a larger transfer times are obtained, as seen in Fig.7. It can be concluded that, it is possible in the hierarchical polling based MAC scheme operate with small size buffer but limited to low load. For larger buffer, although there is no loss of packets, the transfer times are larger and depend on highly of number of sensors of first level.

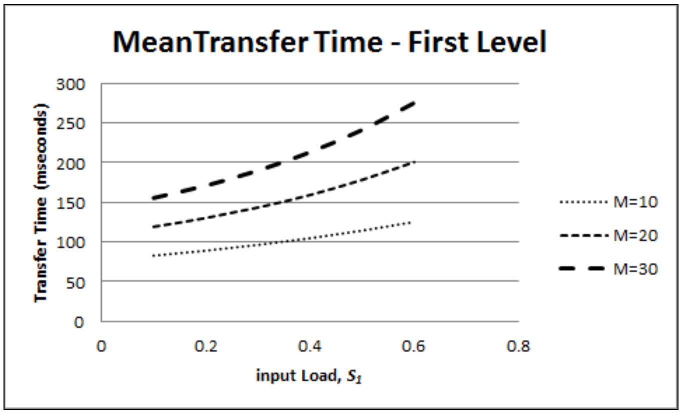


Figure 8. Mean transfer time of first level in function of input load.

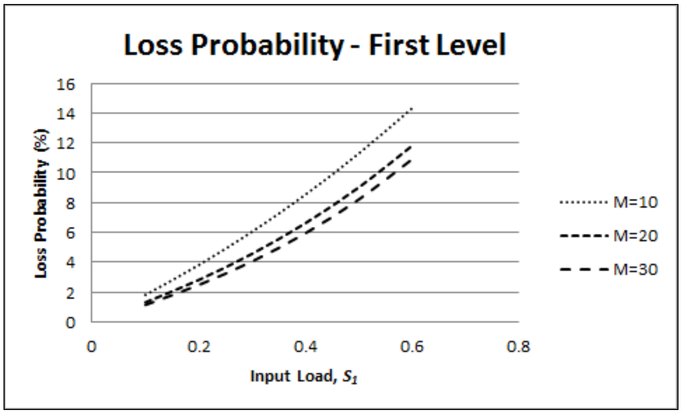


Figure 9. Loss probability of first level in function of input load.

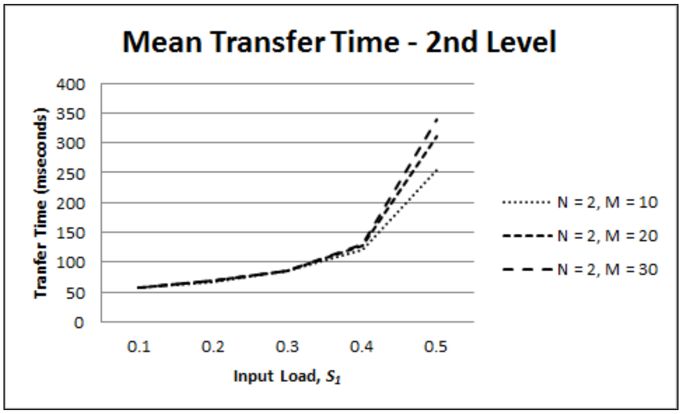


Figure 10. Mean transfer time of second level in function of first level input load of a group.

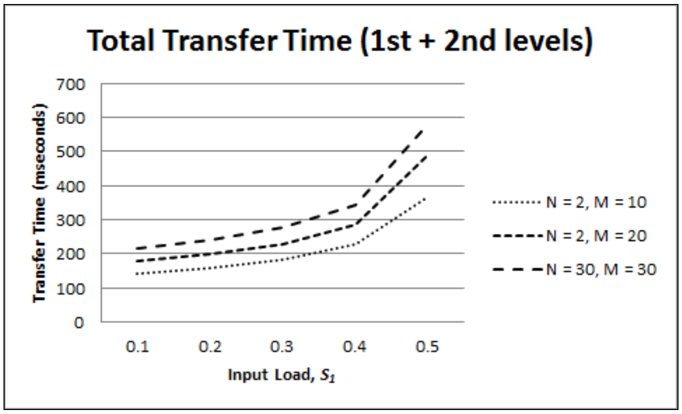


Figure 11. Total meam transfer time in function of first level input load of a group.

5 Conclusions

The performance analysis of hierarchical polling-based access scheme for Wireless Body Sensor Network (WBSN) was carried out in this paper. The mathematical modeling of the proposed scheme was done using infinite size buffers at each sensor node in the first level of hierarchy and also infinite size buffers for sink nodes in second level.

The analysis showed that for the first level using single buffer the transfer times can be kept very small. However, packet loss for high load is prohibitive and must be avoided. For the infinite buffer case in the first level using exhaustive polling technique, the transfer time are higher and the load is limited in function of the performance of second level. The analysis for second level using exhaustive service is dependent of number of sink nodes and also the number of sensor nodes of first level. The total transfer times considering the first and second levels for single buffer can be kept low. For infinite buffer case the total transfer time including first and second levels is very sensitive to the number of sensor nodes of first level. The results show that the hierarchical polling scheme can be convenient for WBSN applications. It must be pointed out that used link capacity is not high, mainly in the case of communication between sink nodes and master node which was only 20 kbps. In this segment a higher transmission capacity can be provided so that a better network performance can be expected.

6 Acknowledgment

This work was supported by São Paulo Research Foundation (FAPESP) under grant No 2011/12463-0.

7 References

[1] B. Latré, B. Braem I. Moerman, C. Blondia and P. Demeester, “A survey on wireless body area networks,” Wireless Networks, Volume 17 Issue 1, January, 2011, Kluwer Academic Publishers Hingham, MA, USA.

[2] B. Otal, L. Alonso and C. Verikoukis, “Towards energy saving wireless body sensor networks,” in Health Care Systems” Proceedings of IEEE International Conference on Communications (ICC 2010), Second International Workshop on Medical Applications Networking (MAN 2010), Capetown, South Africa, 2010.

[3] Z. Yan and B. Liu, “A context aware MAC protocol for medical wireless body area network,” in Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International, pp. 2133-2138.

[4] B. Liu, Z. Yan and C. W. Chen, “CA-MAC: A Hybrid context-aware MAC protocol for wireless body area networks,”
in 13th IEEE International Conference on e-Health Networking Applications and Services (Healthcom), 2011, pp. 213-216.

[5] X. Zhu, S. Han, P. Huang, A.K. Mok and D. Chen, “MBStar: a real-time communication protocol for wireless body area networks,” in 23rd Euromicro Conference on Real-Time Systems (ECRTS), 2011, pp. 57-66.

[6] D. Layerle and A. Kwasinski, “A power efficient pulsed MAC protocol for body area networks,” in IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2011, pp. 2244-2248.

[7] Y. Tselishchev, “Designing a medium access control protocol for body area networks,” in IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011.

[8] L.M. Borges, F. J. Velez and A.S. Lebres, “Performance evaluation of the schedule channel polling MAC protocol applied to health monitoring in the context of IEEE 802.15.4,” in 11th European Wireless Conference - Sustainable Wireless Technologies (European Wireless), 2011, pp. 94-101.

[9] S. Kutty and J.A. Laxminarayan, “Towards energy efficient protocols for wireless body area networks,” in International Conference on Industrial and Information Systems (ICIIS), 2010.

[10] L. Huaming and T. Jindong, “Heartbeat-driven medium-access control for body sensor networks,” IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 1, January 2010, pp. 44-51.

[11] K. A. Ali, J.H Sarker and H.T Mouftah, “Urgency-based MAC protocol for wireless sensor body area networks,” in IEEE International Conference on Communications Workshops (ICC), 2010.

[12] A. A. Khaled, H. S. Jahangir and T. H. Mouftah, “QoS-based MAC protocol for medical wireless body area sensor networks,”
in IEEE Symposium on Computers and Communications (ISCC), 2010, pp. 216-221.

[13] X. Zhang,H. Jiang, X. Chen, L. Zhang, Z. Wang, “An energy efficient implementation of on-demand MAC protocol,” in Medical Wireless Body Sensor Networks” IEEE International Symposium on Circuits and Systems, 2009. ISCAS 2009, pp. 3094-3097.

[14] S. Marinkovic, C. Spagnol and E. Popovici, “Energy-efficient TDMA-based MAC protocol for wireless body area networks,” in Third International Conference on Sensor Technologies and Applications, 2009. SENSORCOMM '09, pp. 604-609.

[15] G. Fang, E. Dutkiewicz, “BodyMAC: energy efficient TDMA-based MAC protocol for wireless body area networks,” in 9th International Symposium on Communications and Information Technology, 2009. ISCIT 2009, pp. 1455-1459.

[16] S. S. Oliveira and S. Motoyama, “Applications oriented medium access control protocols for wireless sensor networks,” IEEE Latin America Transactions, v. 7, Issue 5, 2009, pp. 586-593.

[17] K. S. Kwak, S. Ullah and N. Ullah, “An overview of IEEE 802.15.6 standard,” in ISABEL, 2010, Rome, Italy.

[18] O. Omeni, A. Wong, A.J. Burdett and C. Toumazou, “Energy efficient medium access protocol for wireless medical body area sensor networks,” IEEE Transactions on Biomedical Circuits and Systems, Volume: 2 , Issue: 4, 2008, pp. 254-259.

[19] R. Gravina, A. GuerrierI and A. Fortino, “Development of body sensor networks applications using SPINE,” in IEEE International Conference on Systems, Man and Cybernetics. Singapore, 2008.

[20] S. Motoyama, “Flexible polling-based scheduling with QoS capability for Wireless Body Sensor Network” in Local Computer Networks Workshops (LCN Workshops), 2012 IEEE 37th Conference, pp. 745-752.

[21] T.A. Pazeto, L. F. Refatti and S. Motoyama, “Polling-based Medium Access Control Scheme for Wireless Body Sensor Network” in International Conference on Wireless Networks - ICWN-12, 2012, Las Vegas. pp. 87-93.

[22] H. Takagi, “Analysis of Polling Systems”, The MIT Press Cambridge, Massachusetts London, England, 1986.

[23] S. Marinkovic, C. Spagnol and E. Popovici, “Energy-Efficient TDMA-based MAC Protocol for Wireless Body Area Networks” in Third International Conference on Sensor Technologies and Applications, SensorComm 2009, pp. 604-609.

[24] S. Motoyama, “Hierarchical Polling-based MAC scheme for Wireless Body Sensor Network” in International Conference on Wireless Networks - ICWN-13, 2013, Las Vegas. pp. 103-109.

A Software Update Method using Clustering WSNs

Hyeyeong Jeong and Byoungchul Ahn
Dept. of Computer Engineering, Yeungnam University, Gyungsan, Korea

Abstract *: Wireless Sensor Networks(WSNs) are applied to many monitoring applications. Present sensor nodes can perform many functions at the same time and contain complex software. During the lifetime of sensor nodes, they are required to reprogram their software because of their new functions or software bug fixes. Since typically sensor nodes are inaccessible physically or it is very difficult to upgrade their software by one by one, it is necessary to upgrade software of sensor nodes in WSNs remotely. This paper presents an energy efficient method by selecting an optimal relay node. The Cluster Head Relay (CHR) method is compared with SPIN and RANDOM method. Three methods are simulated using NS-2 with the same environmental parameters. Simulation results show that CHR shows that the update time is reduced up to 17% and the number of relay nodes is reduced up to 19% compared with other two methods.*

Keywords: firmware upgrade, clustering, SPIN, relay node, WSN

1 Introduction

WSNs are becoming one of the most recent Ubiquitous computing technologies. WSN applications are rapidly spreading since recent developments of semiconductor technology gives opportunity to implement many functions in a sensor node. These sensor nodes can be installed to observe inaccessible environment, traffic monitoring and disaster spread places.

After deploying sensor nodes, their battery is not replaceable or rechargeable since they are installed on areas where are difficult to access. Therefore, efficient energy consumption methods are important in WSN applications. Most energy is consumed by radio transmission modules in sensor nodes. Energy consumption of radio modules can be saved by shortening of the transmission distance. In order to reduce the transmission distance, relay nodes have to be chosen optimally. In this paper, we present a new method to update firmware in WSNs, by selecting optimal relay nodes to increase energy efficiency.

In Chapter 2, related work for firmware update is given. Chapter 3 represents the network model and assumptions of proposed method and Chapter 4 describes, the proposed method to select relay nodes, CHR algorithm, is described. In Chapter 5, simulation results of the proposed work are discussed. Conclusion is discussed in chapter 6.

2 Related Work

Stephen *et al.* suggest a model to upgrade software in WSNs[8][9]. This model presents the theoretical approaches to upgrade software. They have not presented any 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).

Infuse implements selective retransmission Go-Back-N scheme using TDMA protocol’s implicit acknowledgement and back pressure mechanisms to provide reliable transfer. In order to save energy, TDMA scheme’s listening method is used. Main feature of Infuse method is continuously sends data to the next node from a predecessor node[1].

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 upgrade many nodes on WSNs. If these methods are used for re-programing sensor nodes of WSNs, each node must be updated first and retransmit the software upgrade data to another node one by one. Therefore it is necessary to develop an efficient upgrade method for sensor nodes with fast upgrade time and small data retransmission.

The direction to upgrade control software is the opposite direction of normal data transfer[5][6]. It is necessary to study for large data transfer from one node to many nodes efficiently. There are some researches about software upgrades for sensor nodes. But they are focused on system management, not an upgrade itself [7]. In this paper, an energy efficient software upgrade method is described by comparing the other methods.

3 Proposed Upgrade Model

All sensor nodes of WSNs are assumed to use the same hardware configuration such as the same memory size and the same processor and so on. It means that all sensor nodes use the same software version. And a distance between two nodes is the same and the location of nodes is fixed.

In this paper, 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 network

③ All nodes use same firmware

④ All nodes are fixed on position

3.1 Energy Consumption Model

In this section, energy consumption model to update software is described. Total energy consumption of nodes can be represented as the sum of energy used for transmitting and receiving control data, software transfer and retransfer data in case of error.

Figure 1 shows data transmission pattern when nodes are placed by one line. In Figure1, "*r*" represents node's transmission radius and "*l*" defines the distance from the base node to the last node. The black node is a base node and gray nodes are relay nodes to update software. White nodes update software only without relaying software update. In Figure 1, the minimum number relay nodes, *N*, from the base node to the last node can be calculated by Equation (1).

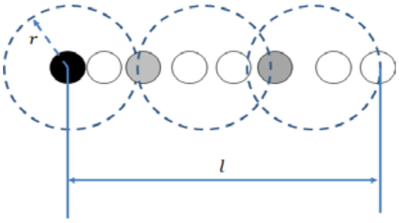


Figure 1. One dimensional placement of nodes for software update

$$N = \frac{l}{r} \tag{1}$$

where, $r = radio\ radius$

$l = the\ distance\ from\ the\ fist\ node\ to\ the\ last\ node.$

Nodes are placed in 2-dimensional space and they are fixed in place. The average number of receiving nodes, (N_a) from transmitting node can be calculated by Equation (2).

$$N_a = \frac{\pi \bullet r^2}{A} \bullet N_t \tag{2}$$

where, $r = Radius\ of\ the\ wireless\ nodes$

$A = The\ area\ of\ the\ sensor\ field$

$N_t = Total\ number\ of\ nodes.$

Energy consumption for transmission and reception can be expressed as energy consumption (E_s) of a transmitting node and energy consumption (E_r) of a receiving node. Equation (3) defines the energy consumption of a receiving node. And Equation (4) expresses energy consumption of a transmitting node.

$$E_S = (D_f \bullet e_s + D_{cr} \bullet e_r) \tag{3}$$

$$E_r = (D_f \bullet e_r + D_{cr} \bullet e_s) \tag{4}$$

where, $D_f = number\ of\ packets\ for\ firmware$

$e_s = transmitting\ energy\ per\ unit\ data$

$D_{cr} = the\ amount\ of\ Control\ data\ for\ error\ recovery$

$e_r = receiving\ energy\ per\ unit\ data.$

To select a relay node in wireless sensor networks, the energy level of the relay node should be checked since software update time and energy level of each node might be different. Equation (5) calculates the energy consumption (*E*) of all nodes in wireless sensor networks.

$$E = (E_r \bullet N_a + E_s) \bullet N \tag{5}$$

where, $N = number\ of\ relay\ node.$

3.2 Firmware Update Time

Time to update software can be expressed by the sum of data transmission time, error recovery time and reprogramming time at a node. Required time (T_{data}) for data transfer is calculated by Equation (6) and overall error recovery time (T_{err}) is equal to Equation (7)

$$T_{data} = \left(\frac{P_d + P_h}{bit\_rate} + P_o \right) \bullet c_f \tag{6}$$

where, $P_d = Data\ packet\ size$

$P_h = Packet\ header\ size$

$P_o = Wireless\ channel\ access\ time$

$C_f = the\ number\ of\ packets\ for\ software\ updates.$

$$T_{err} = \left(\frac{P_d + P_h}{bit\_rate} + P_o \right) \bullet c_f \bullet P_{err} + \left(\frac{P_c + P_h}{bit\_rate} + P_o \right) \bullet C_f \bullet P_{err} \tag{7}$$

where, $P_c = Data\ packet\ size$

$P_{err} = Transmissi on\ error\ rate.$

$$T_{step} = T_{data} + T_{err} + T_{up} \tag{8}$$

where, $T_{up} = Update\ time\ after\ receiving\ data.$

Update time (T_{step}) of node in single hop is calculated by Equation (8). Network software update time (T) is expressed as double of all nodes software update time. Network software update time is shown in Equation (9)

$$T = T_{step} \tag{9}$$

where, $l = distance\ \ from\ base\ node\ to\ last\ node.$

4 Relay Node Selection

4.1 Relay Node Selection

The updating time and energy consumption are affected by the selection method of a relay node. Three selection methods, which are SPIN, random selection and clustering, are considered to update software.

(1) SPIN is a routing method that utilizes metadata to reduce energy loss and data redundancy to twice [10]. Operation steps of SPIN are shown in Figure 2. In Figure 2 (a), if a source node has data to transmit, it broadcasts ADV message to inform other nodes. In Figure 2(b) a node to receive data responds with REQ message. In Figure 2(c), data is transmitted from a source node, black node, to a requested node, B.

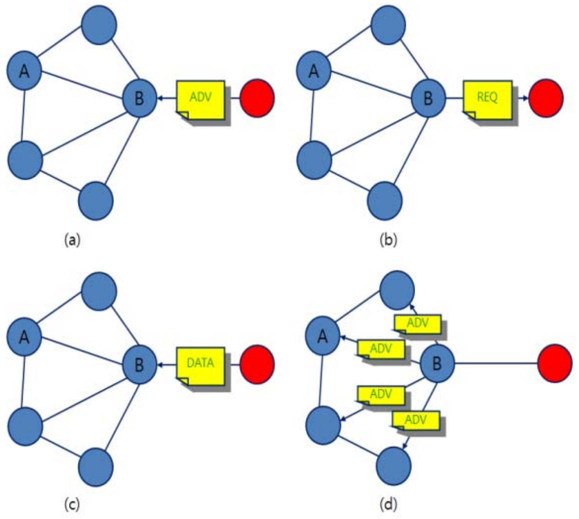


Figure 2. SPIN Protocol

(2) Random: A relay node is selected randomly by the relay node of the previous stage.

(3) Cluster Head Relay(CHR). At this time sensor networks are grouped as clusters. Nodes in a cluster send their collected data to their cluster head. Cluster heads collect data from nodes, and send them to a base station. Configuration of cluster heads follows LEACH's method [11]. Among the received data, the base station searches cluster head ID and selects a cluster head as relay node. If communication between the base station and a cluster head fails, the base station selects a relay node among member nodes. However, cluster heads must have enough energy and strong signal to reduce communication error or fail during updating software.

4.2 Cluster Head Relay Node (CHR) Algorithm

In LEACH method, gathered data are sent to the base station by cluster heads. When data is transmitted to the base station, CH information is added. In Figure 3, cluster head nodes 5, 7, 9, 12 and 17 in the clusters send gathered data to the base station.

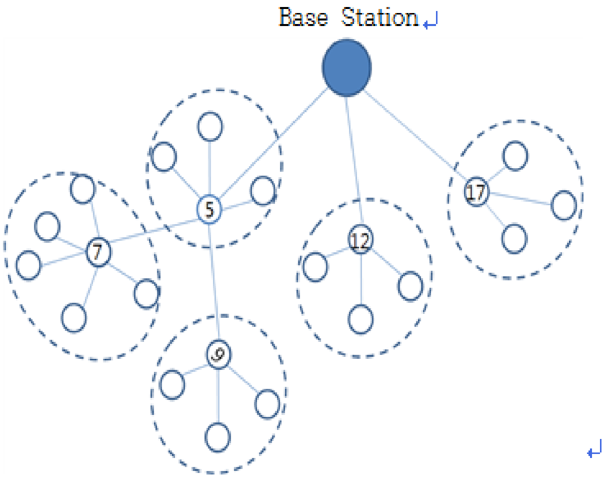


Figure 3. Cluster heads as a relay node

The base station searches node IDs of cluster heads among received data and selects these nodes as relay nodes. Before updating firmware, nodes to be updated acquire event data and send them to neighbor nodes. CHR algorithm is shown as Figure 4. After receiving "Relay-Start" message, software update procedure is started.

```
Data send 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.
  }
 }
Go to 1.1

Data receive node
Broadcast status information periodically
If (receive “data-start”) {
  If (receive firmware ver. > stored firmware ver.) {
    Receive data and store it to memory
    Request missing or lost packet
    Reprogramming it-self.
    Send “reprogram-don” to data send node.
  }
}
If( receive “relay-start”)
  Go to 1.1
Else
  Go to 2.1
```

Figure 4. Proposed CHR algorithm

Relay nodes broadcast “Data-Start” message and new software version to their neighbor nodes, and then inform software updating transmission is started. After software upgrading is finished, updated nodes send “Reprogram-Done” message to their relay nodes.

5 Simulation and analysis of results

5.1 Simulation Environment and Evaluation

As a simulation tool, NS-2 is used. There are one base station and 99 sensor nodes in the network and they are distributed uniformly. CSMA/CA 802.11 and 802.15.4 are used as the wireless protocol. Table 1 shows other simulation parameters.

Table 1. Simulation Parameters

Parameters	Value
MAC Protocol	CSMA/CA Back-off Window 2~26 Slot Time=0.384ms IFS=1.664ms
RF Transmission Range	60m
Wireless Bandwidth	250Kbps
Number of node	100
Transmit / receive energy consumption	75.9mW(TX)/62.7mW(RX)
The distance between nodes	20m~50m
The size of the packet data	64byte ~ 256 byte
Data Header Size	8 byte
Protocol overhead (IP+MAC)	16yte(8 byte +8 byte)
Firmware size	128K byte
The data transfer rate	Wireless Bandwidth 70%

To compare the performance of software update for three methods in Section 4, simulations are taken to analyze four factors below.

① Energy Consumption
② All node update time
③ Number of relay node
④ Loss and error data

5.2 Simulation Results

Distance of nodes is measured at 20*m*, 40*m* and 50*m*. Energy consumption, update time, the number of relay nodes and data loss and error are measured. When the distance between nodes is at 40 meter, simulation results of power dissipation are very close to expected theoretical results. In this paper, the node-to-node distance of 40*m* is used for simulations.

Figure 5 shows the total energy consumption to update software for all nodes. The total energy consumption of wireless sensor networks is calculated by Equation (5). When the packet size is 256 bytes, the energy consumption of nodes is equal to 101 *joules*. When the packet sizes are 64 bytes and 92bytes, energy consumption is increased for three methods, because smaller packet sizes generate more collisions. When data packet size changed from 92 bytes to 224 bytes, energy consumption of the CHR method decreases from 115 *Joules* to 105 *Joules*.

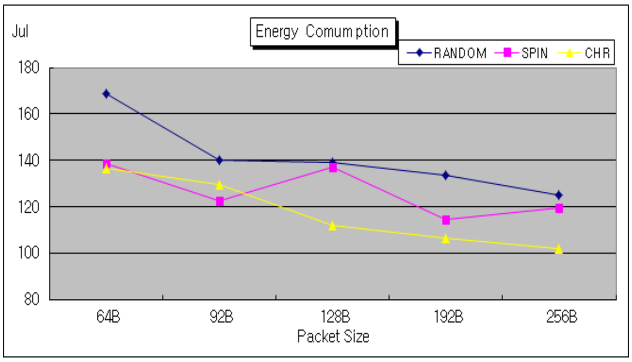


Figure 5. Energy Consumption

Figure 6 shows the software update time of all nodes. The total software update time is calculated by Equation (9). When the packet size is 256 bytes, the update time is 120 seconds. When the packet size is 64 bytes, the software update time takes from 190 seconds to 240seconds. Although the packet size is between 96 byte and 224 byte, the software update time is between 105 and 135 seconds.

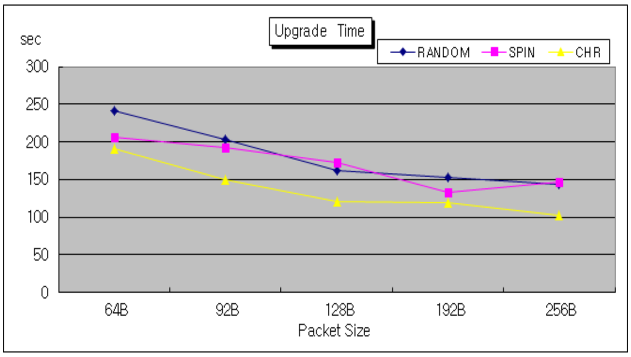


Figure 6. Software update time

Figure 7 shows the number of relay nodes for three methods. CHR shows the best performance of the three methods. When the packet size is 256 bytes, the number of relay nodes is the lowest number as 35 nodes.

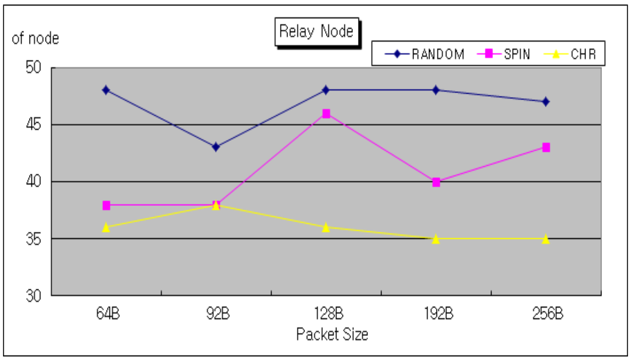


Figure 7. Number of relay nodes

Figure 8 shows the data error rate of the software update. When the packet size is 256 bytes, the data loss rate is 1.05 bytes. During the packet size is taken as 64 bytes, the rate of data loss is high because of data collisions. When the packet size is between 128 and 256 bytes, data error rate is lowest.

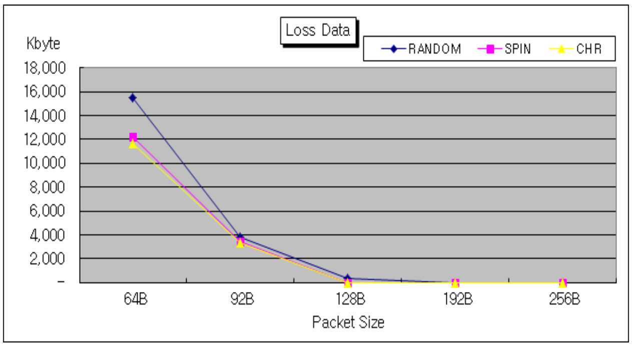


Figure 8. Data error rate

From the simulation results, CHR method shows better performance than SPIN and RANDOM. To select relay nodes, the energy status of relay nodes should be considered. The advantage of CHR method shows less energy consumption than other methods.

Also one disadvantage of SPIN is that data transfer is not guaranteed. For example, when a destination node needs to update its software, it could be located since it is far from a source node. The relay node does not send “REQ message” to the source node and does not broadcast to its neighbor nodes. As a result, the destination node cannot update its software. Therefore, the destination node needs to find another relay node and dissipates more energy. RANDOM method selects relay nodes randomly and update time and energy consumption can be increased by data collisions.

6 Conclusion

In this paper, three software update methods are compared using NS2 simulation tool. When node-to-node distance is at 40 meters and the packet size is between 128 bytes and 256 bytes, it shows best result. CHR method decreases the number of relay nodes to 12%~15% and its data error rate is 0.5%-1.5% compared with other methods. Energy consumption is reduced to 0.07~19% and update time is reduced to 0.22~0.25% compared with other methods. When the packet size is changed from 64bytes to 256 bytes, the packet size of 160 bytes performs better performance compared with other packet sizes, which are 64 or 128 bytes respectively. Please address any questions of this paper to Byoungchul Ahn by Email (b.ahn@yu.ac.kr).

7 Acknowledgement

This work (Grants No. C0149241) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2013.

8 References

[1] S. Kulkarnia, 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] 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.

[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] Honggang Wang , Dongming Peng , Wei Wang , Hamid Sharif , Hsiao-Hwa 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] 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.

[8] S. Brown and C. Sreenan, “A New Nodel for Updating Softeare in Wireless Sensor Networks,” IEEE Network Nov/Dec. pp.42-47, 2006.

[9] 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

[10] Y. Kwon, Ph.D. Dissertation, Yeungnam University, 2011

[11] Kemal Akkaya, Mohamed Younis, "A survey on routing protocols for wireless sensor networks," Ad Hoc Networks 325–349, 2005.

[12] 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

[13] 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.

[14] 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

[15] Zeenat Rehena1, Krishanu Kumar2, Sarbani Roy2, Nandini Mukherjee2: SPIN Implementation in TinyOS Environment using nesC. In : IEEE, 2010

[16] Zeenat Rehena, Sarbani Roy, Nandini Mukherjee: A Modified SPIN for Wireless Sensor Networks. In: IEEE, 2011

[17] 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.

[18] Wei Dong1;2, Yunhao Liu1;3, Chao Wang4, Xue Liu5, Chun Chen2, Jiajun Bu2, “Link Quality Aware Code Dissemination in Wireless Sensor Networks”, in IEEE International Conference, 2011.

[19] Yeungmoon Kwon, Byoungchul Ahn, “A Firmware Upgrade Model for Wireless Sensor Networks,” in the Korean Institute of Informations Scientists and Engineers, 2011(in Korean).

[20] Hossein Sharifi Noghabi 1, Arash Ghazi askar 1, Arash Boustani2, Arash Moghani1, Motahareh Bahrami Zanjani, “IMPLEMENTING A GREEDY CHAIN ROUTING TECHNIQUE WITH SPREAD SPECTRUM ON GRIDBASED WSNS,” International Journal of Wireless & Mobile Networks (IJWMN) Vol. 4, No. 4, 2012.

Packet Loss Resilience Lighting Control for Wireless Sensor Networks

Yushi Uchimura[1], **Hiroto Aida**[2], **Ryoga Okunishi**[1], **Yo Motoya**[1], **and Mitsunori Miki**[2]
[1]Graduate School of Science and Engineering, Doshisha University
[2]Department of Science and Engineering, Doshisha University,
1-3 Tataramiyakodani, Kyotanabe-shi, Kyoto, 610-0394 Japan

Abstract— *The use of wireless sensor networks in offices for the purpose of conserving power has been attracting attention. At the same time, we are conducting research and development of an Intelligent Lighting System that targets offices and realize the illuminance required by the workers. By using a wireless sensor network in the Intelligent Lighting System, we are able to improve the ease of installing illuminance meters. However, in wireless network sensors that use many low power wireless signals, the potential for packet loss due to radio wave conditions exists. In this study, multiple lighting control algorithms that enable convergence on the required illuminance, even if packet loss occurs, are proposed. Moreover, the proposed algorithm was evaluated under multiple environments in which packet loss was generated. By using the Estimation Algorithm, illuminance convergence equivalent to cases involving no packet loss was realized even in environments with a high packet loss rate.*

Keywords: Wireless Sensor Network, Lighting Control, Office Environment

1. Introduction

The use of wireless sensor networks where multiple sensors work together to collect information is expected in a variety of fields. The range of application of wireless sensor networks is extensive and covers a broad array of areas, such as outdoor environment monitoring, and control in buildings and factories [1], [2]. In particular, in recent years the introduction of wireless sensor networks in office buildings has increased due to the increase in awareness regarding security and energy conservation. In contrast, measures to improve the intellectual productivity, creativity, and comfort of employees in offices are attracting attention [3], [4]. According to research by Boyce and his team, it has become clear that the intellectual productivity and creativity of employees in an office are improved by improving the lighting environment [5]. Moreover, power consumed by lighting in an office accounts for a significant percentage of the overall power consumption, and thus methods to reduce power consumption by controlling the brightness of lights are proposed [6], [7], [8]. In the midst of this, we are researching and developing the Intelligent Lighting System [9]. The Intelligent Lighting System uses illuminance meters to realize the illuminance level required by individual office employees and thereby improve the lighting environment of offices. Moreover, with the Intelligent Lighting System, energy savings can be improved by minimizing the brightness of unnecessary lights.

A drawback of wired illuminance sensors is the need for wiring work. In contrast, the advantages of wireless sensor networks include an improved ease of installing the illuminance meters and the ability to flexibly support changes in an office layout. However, with wireless sensor networks that use a lot of low power wireless signals, the potential for severing communications or data loss due to radio wave conditions exists. Offices in particular contain a varied layout of office equipment and obstacles, and often times, the status of radio waves can unexpectedly worsen dramatically due to the movement of office employees and the use of electronic devices. The Intelligent Lighting System changes the brightness of lighting and implements feedback control based on illuminance information transmitted from the illuminance meters. However, in some cases, the feedback control becomes unstable due to severed communication or data loss, and as a result, convergence to the illuminance required by the office employees cannot be achieved.

In this study, Periodic Algorithm that uses an illuminance value prior to a packet loss occurrence, Wait Algorithm that waits until a packet arrives, and Estimation Algorithm that estimates the illuminance value are proposed as lighting control algorithms that enable convergence on the required illuminance even if packet loss occurs. These three algorithms differ in control timing and the illuminance used in lighting control. Construction of a system that is resistant to packet loss is possible by changing the control timing and the illuminance used in the lighting control.

Our contributions in this paper are as follows.

- Presenting a "Packet Loss Resilience Scheme for Lighting Control" that does not place a load on the network or wireless sensor nodes when packet loss is generated.
- Using 15 light controllable ceiling lights and wireless sensor nodes to build the "Packet Loss Resilience Scheme for Lighting Control" in real space.
- Evaluated the stability and speed of illuminance convergence for the "Packet Loss Resilience Scheme for

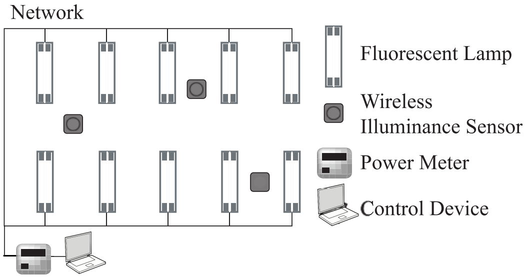


Figure 1: The construciton of an Inteligent Lighting System.

Lighting Control" in multiple environments that generate packet loss.

This paper is configured as follows. First, in the following section, a lighting control algorithm for the Intelligent Lighting System is presented. In Section III, a lighting control algorithm that takes into consideration packet loss and for which a proposal is given in this paper is presented, and the evaluation thereof is given in section IV. In Section V, we discuss related research, and the conclusion is presented in Section VI.

2. Intelligent Lighting System

2.1 Overview of an Intelligent Lighting System

An Inteligent Lighting System achieve each user's required illuminance, and aim to improve the level of intellectual productivity. Using the Intelligent Lighting System, energy consumption will be improved by minimizing the brightness of unnecessary lights.

The Inteligent Lighting System, as indicated in Fig. 1, is composed of dimmable lights, illuminance sensors, control PC, and power meter, with each elements connected via a network. The illuminance sensor is provided for each user, and the illuminance in front of the user is measured by the sensor. In the control device installed for each lightings, illuminance information is collected from each illuminance sensor together with information about the power consumption from the power meter. Based on these datas, the control device will adjust the illuminance according to an optimization technique to achieve the illuminance requested by each user while also striving to keep the power consumption to a minimum.

2.2 Lighting Control Algorithm

The Intelligent Lighting System used an algorithm based on the general purpose optimization method called Hill Clibming. Furthermore, with this algorithm, the factor of influence on lighting for each illuminance sensor was estimated and to efficiently bring about changes in luminance, depending on circumstances.

Control Algorithm using Stochastic Hill Climbing

The Intelligent Lighting System uses Stochastic Hill Climbing method (SHC) for lighting control purposes. The Intelligent Lighting System aims to adjust the illuminance to equal or greater than the target illuminance for the location where the sensors are installed, and autonomously finds the lighting intensity to minimize the amount of electrical power used for lighting. This illuminance is formulated as an objective function. The objective function is shown in Equation 1.

$$f = P + w \sum_{j=1}^{n} g_i \tag{1}$$

$$P = \sum_{i=1}^{m} Cd_i \tag{2}$$

$$g_i = \begin{cases} 0 & (It_i - Ic_i) \leq 0 \\ (It_i - Ic_i)^2 & (It_i - Ic_i) > 0 \end{cases} \tag{3}$$

n : Number of illuminance sensors
m : Number of lightings, w : Weight
P : Amount of consumed electrical power
Ic : Current illuminance, It : Target illuminance
Cd : Current luminance

Making the brightness of the lamps the design variable, we aim to minimize the f in Equation 1. f is formed from the amount of consumed electrical power P(Equation 2) and g_i(Equation 3), which represents the difference in the illuminance between the current illuminance Ic and the target illuminance It. g_i is added only if the current illuminance lower than the target illuminance. Also, g_i is multiplied by a weight w, and the value of this weight w determines whether priority is given to minimizing, either the constraint conditions on the target illuminance, or the amount of consumed power.

Control Algorithm using Regression Analysis

In SHC, the distance of the lightings and the illuminance sensor is not excluded. Therefore, we have developed an Adaptive Neighborhood Algorithm using Regression Coefficient (ANA/RC)[10], which was developed by adapting the Stochastic Hill Climbing method (SHC) specifically for lighting control purposes. It enables the control system to learn the influence of each lighting on each illuminance sensor by regression analysis. By changing the luminous intensity in response,it enables a quick transition to the optimum intensity. However, the level of influence tends to be inaccurate at the start because the amount of change in light intensity and illumination is small. Therefore, the control system use SHC at the start. Then, the control system shifts to ANA/RC after a specified period of time.

3. Packet Loss Resilience Scheme for Lighting Control

We propose a lighting control algorithm that takes into consideration packet loss. The Periodic Algorithm that uses the illuminance value prior to the generation of a packet loss implements control using the last acquired illuminance when packet loss is generated as the present illuminance. The Wait Algorithm that waits until the packet arrives places control on standby when packet loss occurs until the next packet arrives. The Estimation Algorithm estimates the illuminance value when a packet loss is generated and compensates for the lost illuminance.

3.1 Periodic Algorithm

With the Periodic Algorithm that uses the illuminance value prior to the generation of packet loss, when packet loss is generated, calculation of the objective function is performed using the last acquired illuminance as the present illuminance. This algorithm is a technique that is used in existing Intelligent Lighting Systems. Next, the Periodic Algorithm flow is presented. First, the occurrence of packet loss is confirmed when illuminance is acquired. Namely, the presence of packet loss is confirmed by determining whether or not the illuminance data has been updated. Normally, if transmission and reception of an illuminance value has been implemented, the obtained illuminance value is used in the calculation of the objective function as is. If packet loss has occurred, calculation of the objective function is performed using the previously obtained illuminance value as the current illuminance value. Therefore, the generation of the next luminance and the determination of the objective function value are performed in the same manner as when packet loss has not occurred. With Periodic Algorithm, if packet losses are continuously generated, the same illuminance value is continued in the system. As a result, there is the possibility that a difference could occur between the actual illuminance value and the illuminance value in the system.

3.2 Wait Algorithm

With the Wait Algorithm that waits until a packet arrives, when packet loss occurs, control is placed on standby until the next packet arrives. The Wait Algorithm flow is presented next. First, the presence of packet loss occurrence is confirmed when illuminance is acquired. Namely, the presence of packet loss is confirmed by determining whether or not the illuminance data has been updated. Normally, if transmission and reception of an illuminance value has been implemented, the obtained illuminance value is used as is in the calculation of the objective function. If packet loss has occurred, the present luminance is maintained as is, and lighting control is placed on standby. After illuminance information is received, the system shifts to calculating the objective function, and lighting control is implemented. If packet losses are continuously generated, the same illuminance value is continued in the system, but no difference is generated between the actual illuminance value and the illuminance value in the system. With Wait Algorithm, control is not implemented during the occurrence of packet loss, and thus there is the possibility of an occurrence of delay in illuminance convergence.

3.3 Estimation Algorithm

With the Estimation Algorithm, the illuminance value is estimated when packet loss is generated, and compensation is implemented for the lost illuminance. Next, the Estimation Algorithm flow is presented. First, the presence of packet loss occurrence is confirmed when illuminance is acquired. Namely, the presence of packet loss is confirmed by determining whether or not the illuminance data has been updated. Normally, if transmission and reception of an illuminance value has been implemented, the obtained illuminance value is used as is in the calculation of the objective function. If packet loss has occurred, the illuminance is estimated, and the objective function is calculated.

$$I_e = \sum_{i=1}^{m} L_i \times R_i \tag{4}$$

$$x = I_e \times \frac{I}{I'_e} \tag{5}$$

I_e : Estimated illuminance value
m : Number of lightings
L : Luminance, R : Level of influence
x : Corrected illuminance value
I : Last acquired illuminance value
I'_e : Estimated illuminance value when the last illuminance was acquired

The estimated illuminance value is determined from Equation 4. The estimated value for illuminance is obtained from a summation of the luminance of each light multiplied by the level of influence. Because this Periodic Algorithm uses the level of influence, it can only be applied during ANA/RC. Moreover, when compared to the influence obtained by turning on the lights one at a time, the influence dynamically obtained with SHC is inaccurate, and as the lights become further separated from the illuminance meter, the influence tends to become more inaccurate [11]. This is thought to be due to lighting being more easily influenced by the changes in the luminance of other lights as lights become further separated from the illuminance meter.

Therefore, the calculation results of the illuminance value estimation are corrected by Equation 5. The correction is implemented using a ratio of the last acquired illuminance value and the estimated illuminance value when the last illuminance was acquired. The corrected illuminance value is used to calculate the objective function. For example, if

Figure 2: Experimental Situation.

the current estimated value for illuminance is 900 lx, the illuminance when the last illuminance was acquired is 400 lx, and the estimated value for illuminance when the last illuminance was acquired is 800 lx, then the illuminance compensation is performed as 450 lx.

4. Evaluation

4.1 Evaluation Overview

In order to evaluate the effectiveness of Periodic Algorithm, Wait Algorithm, and Estimation Algorithm, illuminance convergence was verified in environments in which packet loss occurs. Testing was performed at the Intelligent System Creation Environment Laboratory in the Kochikan Building of Doshisha University, and in the test, 15 white fluorescent lamps and 4 wireless sensor nodes were used. Of the wireless sensor nodes, 3 were used as illuminance meters, and 1 was used as a sync node.

A scene of the testing site is shown in Fig. 2, and the testing environment is shown in Fig. 3. The wireless sensor nodes were installed in three locations, including directly beneath one light, between two lights, and between four lights. Fig. 3 shows the positional relationship between the fluorescent lamps and the wireless sensor nodes. The number to the side of each fluorescent lamp indicates the number of the each fluorescent lamp, and the letter to the side of each wireless sensor node indicates the symbol used to identify each sensor. For the wireless sensor nodes, MOTE MICAz [12] from Crossbow were used. An MDA088 general purpose exterior sensor base plate was installed on the MOTE MICAz, and a lead type Napica illuminance meter was embedded to enable acquisition of the illuminance value.

The target illuminance values were set at 450 lx for sensor node A, 500 lx for sensor node B, and 600 lx for sensor node C, and after 1,000 seconds, the target illuminance of sensor node C was changed to 800 lx. Note that when convergence for 200 seconds or longer within ± 50 lx of the target illuminance was achieved, convergence to the target illuminance was considered to be completed.

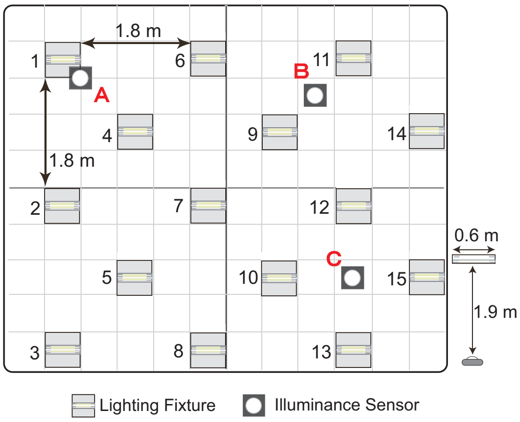


Figure 3: Experimental Environment.

Moreover, for the first 200 seconds, illuminance convergence was conducted with SHC for regression analysis, and after 200 seconds, the system shifted to ANA/RC. When Wait Algorithm was used, the shift to ANA/RC was delayed only for the number of times that packet loss was generated, and because Periodic Algorithm cannot be used during SHC, Wait Algorithm was used.

In order to simulate multiple environments in which packet loss occurs, packet losses were generated by the sync node. Three patterns were used for packet loss generation including a case with random generation 10% of the time, a case with generation 50% of the time, and a case that assumed packet loss with burst properties (loss rate of 58.3%). Ordinarily, packet loss of 50% or higher is thought to not easily occur, but a pattern with this type of high packet loss rate was used in order to verify to what extent the proposed algorithm can function properly. The testing applied Periodic Algorithm, Wait Algorithm and Estimation Algorithm to each of the three types of packet loss patterns described above.

4.2 Testing Results for a Case in which the Illuminance Value was Properly Transmitted and Received

Fig. 4 shows a history of illuminance for a case in which the illuminance value is properly transmitted and received. From Fig. 4, we can see that all of the illuminance meters entered the convergence range for the target illuminance in about 240 seconds. During the next 200 seconds as well, the illuminance remained in the convergence range, and thus one could argue that convergence was achieved. Moreover, sensor node C entered the convergence range about 80 seconds after the target illuminance was changed, and for the next 200 seconds as well, the illuminance remained in the convergence range, and thus we found that convergence was achieved.

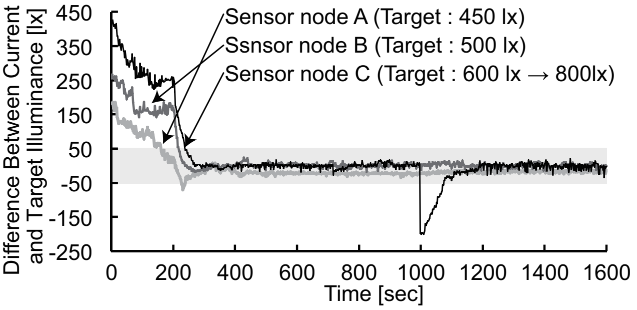


Figure 4: History of the illuminance data without packet loss.

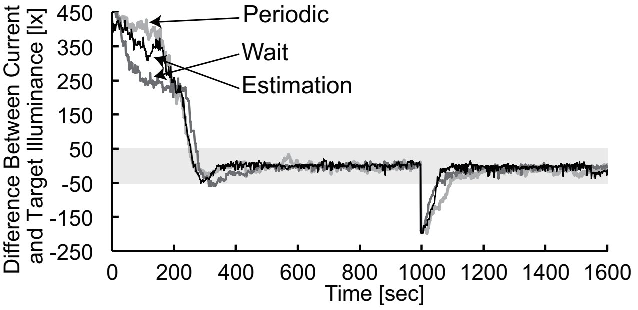


Figure 5: History of the illuminance of sensor node C with 10% packet loss(0 to 1600 seconds)).

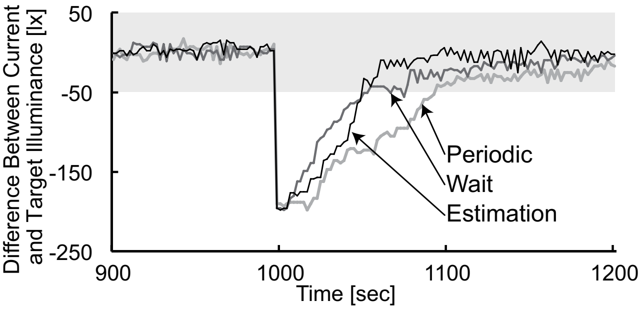


Figure 6: History of the illuminance of sensor node C with 10% packet loss(900 to 1200 seconds).

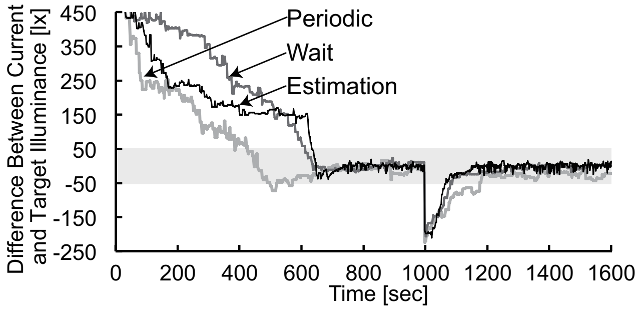


Figure 7: History of the illuminance of sensor node C with 50% packet loss(0 to 1600 seconds).

4.3 Testing Results for a Case in which Packet Loss was Generated with 10% Probability

Fig. 5 shows the illuminance convergence results for the sensor node C for a case in which Periodic Algorithm, Wait Algorithm, and Estimation Algorithm were used when packet loss was generated with a probability of 10%. In addition, Fig. 6 shows an expanded view of the graph of Fig. 5 from 900 seconds to 1,200 seconds. From Fig. 5, a significant difference between the three algorithms was not observed for a case in which packet loss occurred with a probability of 10%. However, from Fig. 6, we can see that with Periodic Algorithm, it took approximately 100 seconds to enter the illuminance convergence range after the target illuminance was changed, and that there was a delay of about 20 seconds in reaching convergence compared to the case in which transmission and reception of the illuminance value were performed normally. With Wait Algorithm and Estimation Algorithm, we found that the convergence range is entered at about the same time as the case in which transmission and reception of the illuminance value were performed normally even after the target illuminance was changed.

4.4 Testing Results for a Case in which Packet Loss was Generated with 50% Probability

Fig. 7 shows the illuminance convergence results for the sensor node C for a case in which Periodic Algorithm, Wait Algorithm, and Estimation Algorithm were used when packet loss was generated with a probability of 50%. In addition, Fig. 8 shows an expanded view of the graph of Fig. 7 from 900 seconds to 1,200 seconds.

From Fig. 7, it is clear that the convergence range was entered with Periodic Algorithm at about 500 seconds for the case in which packet loss was generated with 50% probability. In contrast, with Wait Algorithm and Estimation Algorithm, the convergence range was entered at about 620 seconds. With Wait Algorithm, this was because a delay in illuminance convergence was generated in order to return the luminance each time that packet loss occurred. With Periodic Algorithm as well, it is thought that a similar delay occurred because Wait Algorithm is used at the time of SHC.

From Fig. 8, we can see that with Wait Algorithm and Periodic Algorithm, the convergence range was entered and convergence occurred at about 80 seconds after the target illuminance was changed. In contrast, we can see that with Periodic Algorithm, the convergence range was entered and convergence occurred at about 180 seconds after the target illuminance was changed. The reason for the delay occurring in illuminance convergence with Periodic Algorithm after the target illuminance was changed is thought to be that the current illuminance value was not used in the objective function calculation when packet loss occurred, and thus accurate control could not be implemented. However, because no deviation from the convergence range occurred

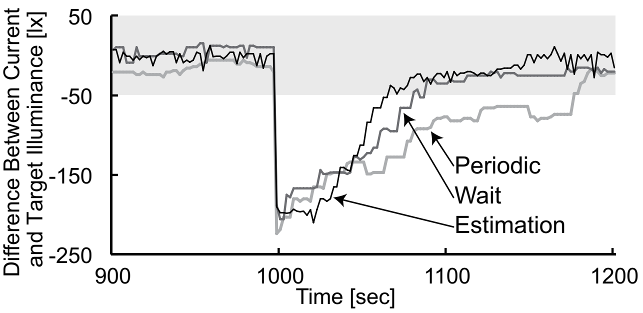


Figure 8: History of the illuminance of sensor node C with 50% packet loss(900 to 1200 seconds).

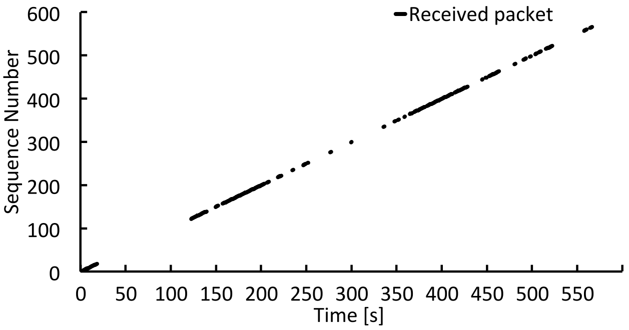


Figure 9: Packet loss obtained by experiment.

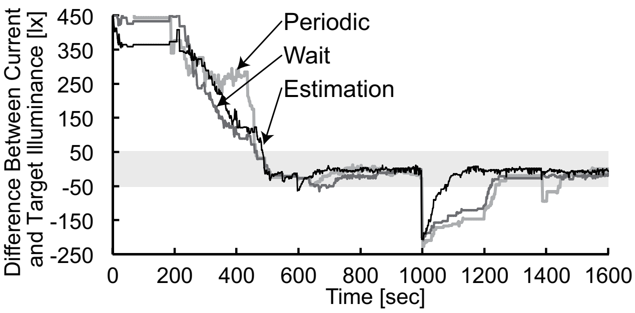


Figure 10: History of the illuminance of sensor node C with burst loss(0 to 1600 seconds).

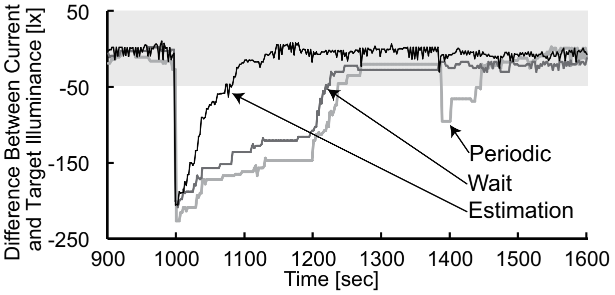


Figure 11: History of the illuminance of sensor node C with burst loss(900 to 1600 seconds).

for 200 seconds after the convergence range was entered with Periodic Algorithm, Wait Algorithm, and Estimation Algorithm, one could argue that convergence was achieved.

4.5 Testing Results for a Case that Assumed Packet Loss with Burst Properties

The pattern for the actually measured packet loss was used as the burst loss pattern. With a loss rate of 58.3%, packet loss like that shown in Fig. 9 was generated. This uses the loss pattern obtained through actual measurements.

The illuminance convergence loss results for the sensor node C for a case that used Periodic Algorithm, Wait Algorithm, and Estimation Algorithm when packet loss with burst properties occurred is shown in Fig. 10. In addition, Fig. 11 shows an expanded view of the graph of Fig. 10 from 900 seconds to 1,600 seconds.

From Fig. 10, it is clear that all algorithms entered the convergence range at about 480 seconds, and from Fig. 11, we can see that with Periodic Algorithm, about 300 seconds were necessary until the convergence range was entered after the target illuminance was changed. Furthermore, even after entering the convergence range, at about the 1,400 second, the illuminance deviated by about 50 lx from the illuminance convergence range. Therefore, convergence to the target illuminance was not completed until about the 1,600 second after the target illuminance was changed. The cause for control not converging is thought to be a problem with the calculation of the objective function. For the case in which Periodic Algorithm was used, when packet loss was continuously generated, the previous illuminance was used repeatedly, and thus, the illuminance became uniform, and the illuminance penalty term became a constant. As a result, it is believed that the objective function became a function that considered only power consumption, and thus, the illuminance fell below the target illuminance. Moreover, when Wait Algorithm was used, a time of about 260 seconds was required until the illuminance convergence range was entered after the target illuminance was changed. With Wait Algorithm, it is thought that time was required to reach convergence because illumination control was not implemented during the time in which packet loss was continuously generated. However, because no deviation from the convergence range occurred for a period of 200 seconds after the convergence range was entered, one could argue that convergence was achieved. On the other hand, when Periodic Algorithm was used, convergence occurred at about 80 seconds after the target illuminance was changed, and for the next 200 seconds, control remained stably within the convergence range. Thus, one could argue that convergence was achieved. This is the same convergence speed as the case in which the illuminance value is normally transmitted and received, and it is believed that this result was achieved because control was implemented by estimating the illuminance even during the time that packet loss was continuously generated. This result also indicates that Periodic Algorithm is effective in an environment that assumes continuous packet loss.

5. Related Work

A significant amount of research that takes into consideration the occurrence of packet loss in wireless sensor networks is being conducted, and research is being conducted on a technique that retransmits the transport layer as a technique that considers packet loss[14], [15]. However, techniques that retransmit packets place a load on the network. Research is also being implemented on a technique that temporally and spatially brings redundancy, installs sensor nodes, and takes packet loss into consideration[16], [17]. With this technique as well, the number of times for packet transmission and the number of transmitters both increase, thereby placing a load on the network. In contrast, with the proposed algorithm, the number of times for packet transmission is not increased, and thus network load can be reduced, and the algorithm is useful.

6. Conclusion and Future Work

In this study, we proposed a lighting control algorithm that takes into consideration packet loss in a wireless sensor network. We were able to confirm illuminance convergence with Periodic Algorithm for three cases including a case in which packet loss is generated with 10% probability, a case in which packet loss is generated with 50% probability, and a case in which packet loss with burst properties occurred. Moreover, even after the target illuminance was changed, we were able to obtain an illuminance convergence speed that was equivalent to a case in which the illuminance value was normally transmitted and received. Hence, Periodic Algorithm is useful in environments in which packet loss is generated. Based on the above, the effectiveness of the illuminance value estimation algorithm in environments such as offices where radio wave conditions are poor and bursty packet loss occurs was demonstrated.

In the future, a higher speed influence examination technique will be considered. With the technique used in this testing, delays are generated in lighting control only for the number of times that packet loss is generated during SHC. However, by using a higher speed influence examination technique, we anticipate that the shift from SHC to ANA/RC will be accelerated, and that the illuminance convergence speed will be improved. Moreover, when use of wireless sensor nodes in an actual environment is considered, battery operation time must also be considered. A conceivable technique to increase battery operation time is to reduce the number of times that the illuminance value is transmitted. It is also conceivable that the proposed illuminance value estimation algorithm can be used to compensate for illuminance values that are not transmitted due to reducing the number of transmission times.

References

[1] Kuorilehto, Mauri, et al. “A survey of application distribution in wireless sensor networks,” EURASIP Journal on Wireless Communications and Networking, Vol.2005, No.5, pp.774-788, 1900.

[2] Arampatzis, Th, J. Lygeros, and S. Manesis, “A survey of applications of wireless sensors and wireless sensor networks,” Intelligent Control, 2005. Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation, IEEE, 2005.

[3] Olli Seppanen, William J.Fisk, “A Model to Estimate the Cost-Effectiveness of Improving Office Work through Indoor Environmental Control,” Proceedings of ASHRAE, 2005.

[4] M.J.Mendell and G.A.Heath, “Do indoor pollutants and thermal conditions in schools influence student performance? A critical review of the literature,” Indoor Air, 15[1], pp.27-52, 2005.

[5] Boyce, Peter R., Neil H. Eklund, and S. Noel Simpson, “Individual lighting control: task performance, mood, and illuminance,” Journal of the Illuminating Engineering Society, pp.131-142, 2000.

[6] Francis Rubinstein, Michael Siminovitch and Rudolph Verderber, “Fifty percent energy saving with automatic lighting controls,” IEEE Industry Applications Society, 29, pp.768-773, 1993.

[7] P.J.Littlefair, “Predicting lighting energy use under daylight linked lighting controls,” Building Research and Information, 26[4], pp.208-220, 1998.

[8] D.H.W.Li and J.C.Lam, “An investigation of daylighting performance and energy saving in a daylight corridor,” Energy and Buildings, 35[4], pp.365-373, 2003.

[9] Maiko Ashibe, Mitsunori Miki, and Tomoyuki Hiroyasu, “Distributed optimization algorithm for lighting color control using chroma sensors,” pp.174-178, 2008.

[10] S.Tanaka, M.Miki, T.Hiroyasu, M.Yoshikata, “An Evolutional Optimization Algorithm to Provide Individual Illuminance in Workplaces,” Proc IEEE Int Conf Syst Man Cybern, 2, pp.941-947, 2009.

[11] M.Miki, Y.Azuma, H.Ikegami, “An Extraction of Influential Lightings for Illuminance Sensors and Lighting Off Mechanism in An Intelligent Lighting System,” The 15th International Conference on Artificial Intelligence 2013, 2013

[12] MEMSIC:MOTE MICAz, http://www.memsic.com/wireless-sensor-networks/

[13] 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, 85[5], pp.346-351, 2001.

[14] Park, Seung-Jong, et al. “A scalable approach for reliable downstream data delivery in wireless sensor networks,” Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing. ACM, 2004.

[15] Wan, Chieh-Yih, Andrew T. Campbell, and Lakshman Krishnamurthy, “PSFQ: a reliable transport protocol for wireless sensor networks,” Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications, ACM, 2002.

[16] Iyengar R, Kar K and Banerjee S, “Low-coordination Topologies for Redundancy in Sensor Networks,” the Sixth ACM Annual International Symposium on Mobile Ad-Hoc Networking and Computing (MobiHoc), Urbana-Champaign, IL, pp.332-342, 2005.

[17] Gao Y, Wu K, and Li F, “Analysis on the Redundancy of Wireless Sensor Networks,” Proc.2nd ACM Intl, Workshop on Wireless Sensor Networks and Applications (WSNA), pp.108-114, 2003.

A Network Structure for Medical Assistance in Rural and Urban Areas Using IoT Technology

E. Serafim, and S. Motoyama
Faculty of Campo Limpo Paulista - FACCAMP
Master Program in Computer Science
Campo Limpo Paulista, S. Paulo, Brazil

Abstract - *A network structure using IoT technology for application in healthcare is proposed in this paper. The main idea is to combine the body sensor network (WBAN) that monitors the vital signs of a patient with RFIDs. RFID readers placed strategically close to the patients, collect data from the WBAN, and through a clustered configuration of these readers reach the medical centers where they are processed and presented to physicians for monitoring purposes. In this paper, the configuration of proposed network structure is presented and main issues for good operation of network are discussed. The link capacity necessary for good operation of network is estimated using queuing theory and the number of RFIDs that can be accommodated in a given link capacity is also estimated.*

Keywords: internet of things, RFID, sensor networks, IoT cluster.

1 Introduction

The Internet of Things (IoT) consists of many smart objects interacting with people and other objects to achieve common goals. The IoT will allow simple objects such as air conditioners, refrigerators, cars, houses, etc. become intelligent and can be identified and accessed through the Internet. Such achievement will be possible due to technologies like RFID and sensor networks, which will provide these objects with intelligence, and thus can communicate [1].

One of the areas that can be benefited by IoT technology is healthcare. The applications of IoT in healthcare can be in areas such as tracking objects and people (staff and patients), identification and authentication of persons; automated data collection and patients monitoring [2].

The IoT technology can be beneficial to the medical care in country like Brazil, due to poor hospital infrastructure existing. The patients may stay in their homes without occupying hospital beds, but being monitored remotely and having prompt medical attention in case of emergencies. In this paper, we propose a network structure that is convenient for patient monitoring in areas of high human concentration, as well as in rural areas. This structure aims to integrate body sensor networks (WBANs) with the Internet, through the interconnection of RFID readers configured in a cluster. These readers should communicate with each other, transferring the data obtained from the sensor network through RFID tags, until they reach a sink node and through the Internet a medical center where they are processed.

The paper is structured as follows. In section 2 the concept of IoT is presented. The related works are described in section 3 and in section 4 the proposed network structure for medical care in urban and rural areas is presented. In section 5 the link capacity and number of RFIDs necessaries for good operation of proposed network are estimated. Finally, in section 6, the main conclusions and future work are presented.

2 Internet of Things

In a broad sense, IoT is the interconnection of the everyday objects of the real life environment with the Internet which is a virtual environment, becoming the objects smarts. This is possible thanks to the use of sensors and addressable RFID tags, attached to the objects, which communicate through a network, and then to the Internet [1] [2] [3] [4] [5] [6] [7]. This concept can be seen in Figure 1.

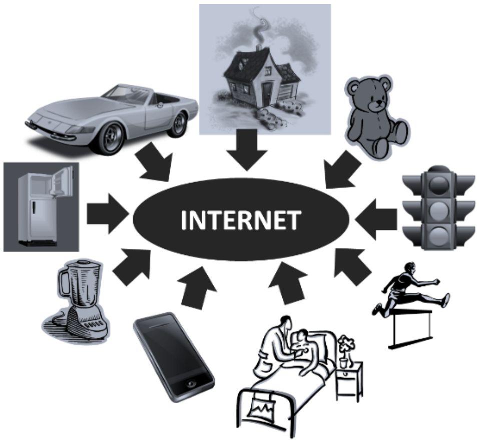


Figure 1. An overview of IoT concept.

Besides the use of RFID, the IoT should include technologies such as artificial intelligence, nanotechnology and embedded systems, which enable an interconnection machine-to-machine [1] [2] [6]. This will lead to a new form of ubiquitous communication, in which objects can communicate with people and other objects independently.

Although the IoT paradigm is very attractive, many questions remain open, which deserves special attention from researchers [1] [2]. The impact on security requires more attention and standards, especially for applications in health.

Because the Internet is an unsafe environment currently, patient privacy may be violated. Despite being the subject of efforts of various organizations such as the Auto-ID Center, EPCGlobal and Unique / Universal / Ubi-quitous architecture IDentifier (UID) in Japan, the lack of standardization is also a big issue today [1] [2] [3] [4] [8] [9] [10].

3 Related work

A platform for Remote Monitoring and Management of Health Information (Remote Monitoring and Management of Healthcare Information Platform - RMMP-HI) was proposed in [11] to monitor the health and preventing disease, improving quality of life, thus relieving the public health system. The project is to deploy sensors in patient body, through a WBAN network, and connect this network to the Internet through a cell phone or a router. Such information reaches the doctor who can track the health of patients.

In [13] a cooperative approach to IoT to monitor and control the parameters of health in rural area was proposed. In this approach, the vital signs such as blood pressure (BP), hemoglobin (HB) blood sugar, abnormal cell growth in any part of the body, etc. are monitored. The proposal is a mechanism for cooperative communication, which is more appropriate for Ad Hoc wireless sensor networks than cellular networks. Each node acts as a user (source), as well as transmitter, transmitting to multiple nodes forming an Opportunistic Large Array (OLA), being significantly flexible and scalable.

The use of RFID for identifying patients to improve health management in rural areas in India is proposed in [14]. The idea is to use an electronic medical record in consortium with RFID tags. The main objective is to enable easy and reliable identification of patients.

4 Proposed network structure

To improve the quality of patients lives, in this article a network structure for patients monitoring through the integration of RFID and WBANS is proposed. The proposed structure uses the monitoring concept presented in [11] associated with the cooperative Internet of Things presented in [13].

The proposed structure is presented in Fig. 2. The patients with WBAN receive an active RFID tag with high range. Active RFID tags should have its range from 5 to 200 meters, so that a near RFID reader can read the data. The tags carry information about the patient, and also serve as the interface between the WBAN and IoT cluster.

RFID readers should be close to the patient, such as in your home or near neighborhood. However, these readers may also be placed in other locations, seeking the wider possible coverage of IoT cluster. In urban centers, squares, parks, rural communities, among others, may be local to placing these RFID readers.

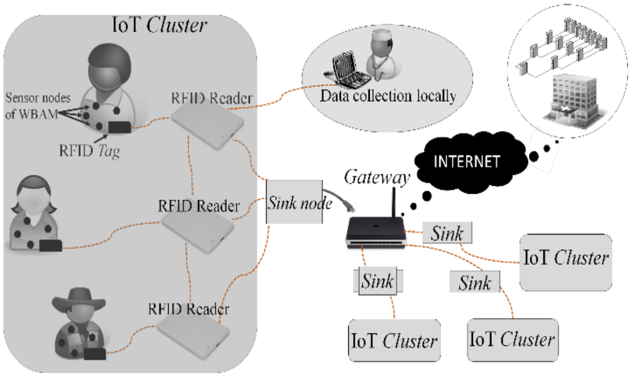


Figure 2. Proposed network model for patient monitoring in rural and urban communities using IoT cluster.

At the edge of the cluster, a sink node is used to collect and send the data to a gateway, which must be connected to the Internet through a conventional link. Through the Internet, the data are sent to a medical center or a larger hospital, which will collect the data, process them and store them.

The health professionals can also collect data locally directly from RFID readers. This procedure will allow the access to the patient data when the doctor is visiting the community. In this case the medical staff shall have a mobile computer that has a USB RFID reader attached, as well as a version of the system software installed.

At the medical center, a software interprets the information in real time, generating reports and alerts to the medical staff. In case where there is patient life-threatening, the alerts should be issued so that the rescue can be provided in a timely manner. Thus, to complement the proposed structure, the network should provide feedback to the doctor, any person responsible for the patient or the user himself. This feedback strategy is detailed in Fig. 3.

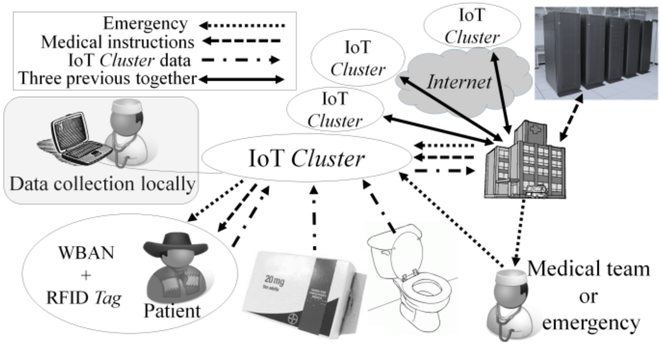


Figure 3. Feedback structure of the proposed network.

Furthermore, other "smart things" can be inserted, which justifies the use of IoT as the proposed center. With other everyday objects comprising this network, it is possible a better analysis of the patient's daily life and routine activities. Examples of this analysis may include measurements of bathroom use in people with kidney problems, control of medicine administration, moving between the rooms of the environment in which the patient lives, among many other possibilities. In Fig. 3, for example, there are a drug and a toilet with RFID tags. The use of one and the other can be measured, and its data can serve as a basis for further analysis in the medical center.

4.1 Components of the network structure

The network structure proposed in this section consists of many elements, which are detailed below.

At the health center, the end of the proposed structure should have a technological apparatus capable of receiving data generated by smart objects, process them and store them. Thus, in this center one or more servers must be provided and software that enables the real-time monitoring of the patients is required. As the data are received, the system must analyze them to determine whether data are classified as normal or need medical emergency.

These servers can be located in the medical center, a clinic, or even in some rented datacenter. Regardless of location, the system should always be running and available for access by health staff.

Since much information about the patients is stored in the server, often confidential, special care must be taken, ensuring the privacy and integrity of the system user. Thus, encryption and security mechanisms must ensure that data are protected and are only accessed by authorized personnel. This security protection should be applied to system software, sensors, readers, tags, etc.

The medical center system must be interconnected with the IoT cluster, preferably full-time. Such interconnection could be through the Internet using a conventional link. If a cluster grows large in size, the link may become overloaded at certain times of heavy use. If there are a lot of smart objects at the network, the throughput can also be a critical factor affecting the stability of the system. Thus, a careful link dimensioning must be made for interconnection between the IoT cluster and the health center. In section 5 a study is presented for this link estimation.

In rural communities, where the Internet is not available, an alternative would be to use the 3G network to connect the system to the IoT cluster. This alternative is attractive because usually the cell phone coverage is more present in these regions. Even in the absence of the cell phone structure, the radio can be installed to provide Internet access. In extreme situation, where even the use of radio is not possible, one option is the use of satellite which can greatly raise the cost of the system, but since has global cover, is an alternative solution.

The sink node should be a device capable of receiving traffic generated by IoT cluster, and convert it to protocols commonly used in the Internet. If possible, this functionality can be embedded directly into the gateway. The IoT cluster comprises the entire RFID infrastructure and the sensor network. Indeed, the sensor network only delivers the data to the cluster IoT, which serves as a bridge between WBAN and the health center.

In rural areas, each village may have only one cluster interconnected with the IoT gateway. In urban areas, several clusters can be provided, reaching many houses in a certain neighborhood. If the patient moves through the city, in public places may exist RFID readers to ensure greater coverage, such as plazas, subway stations and airports.

RFID readers have the function to capture the information of RFID tags placed in patients. Information from other RFID tags attached on smart objects should also be captured by readers covering the environment. If the patient is on the bed and cannot move, the reader can be positioned next to the bed.

Another function of RFID reader is the ability to communicate with others in order to expand the scope of coverage. This allows a data packet to be routed among the readers to reach the sink node. Thus, the readers can be configured as an ad hoc multi-hop network.

Readers should be coordinated by a routing algorithm that allows the choice of a leader or master reader node. The function of master node is the coordination of others nodes indicating the best way to reach the sink node.

The readers will form a set of transceivers operating as an asynchronous distributed joint communication system, constituting an array where work collaboratively, and are configured on an OLA (Opportunistic Large Arrays) of cluster of readers [13]. This arrangement of readers is shown in Fig. 4.

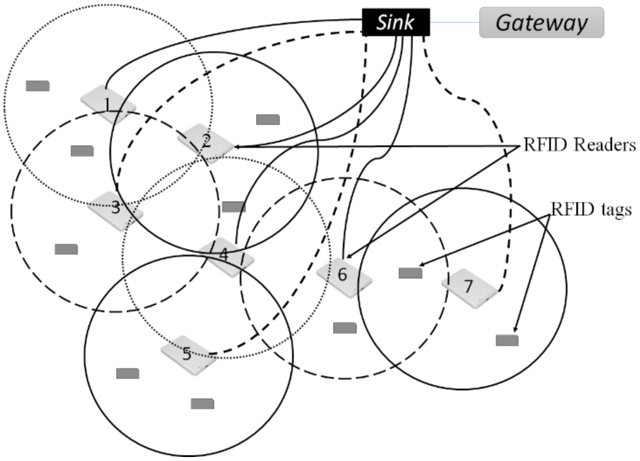


Figure 4. Configuration of cooperative communication among RFID readers.

As can be seen in Fig. 4, all RFID readers have potentially direct access to the sink node. Some readers do not have enough coverage to reach the sink node, because they are far from sink node, as shown in dotted lines. However, the readers can be moved from one place to another for better RFIDs readings and in this new rearrangement may have direct access to the sink node. This rearrangement ensures flexibility to the network configuration, even with frequent repositioning of readers.

Each RFID reader should check, initially, if there is a direct communication with the sink node. If it exists, forward the packet directly to the sink node, but must inform the closest reader that has direct communication.

The patient must have a long-range RFID tag. If the patient move out of range of the reader for some time, for example, to go to work in the field in rural areas, or move to some urban place without coverage of IoT cluster, the data from the sensor network must be stored in the tag, and when the patient reaches within range of the RFID reader, the data are transmitted to the medical center. Another possibility could be the use of Smartphones equipped with RFID readers that could store information while the patient is off-line.

5 Link dimensioning

The estimation of link capacity of the sink node is important factor for the good operation of the proposed network structure. In this section the link capacity estimation is carried out using simple queuing model. To model the proposed network, the WBAN can be considered as the packet generator and RFIDs readers constitute a network to deliver the packets to the sink node, as shown in Fig. 5.

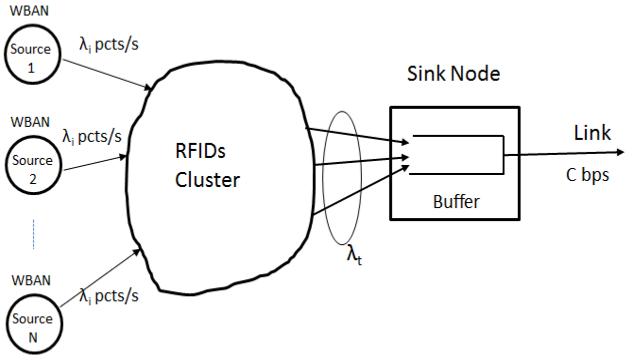


Figure 5. Network structure model for the estimation of link capacity of sink node.

For the estimation of link capacity of sink node, RFIDs cluster is just a delay network to deliver all the packets generated by WBANs. Thus, for the link capacity estimation, the analytical model can be just a buffer with a link and with a total packet rate $\lambda_t = \sum_i^N \lambda_i$ arriving to the sink node. Assuming that packets arrive to the sink node obeying Poisson distribution, as used in [15] and the packet length has negative exponential distribution, a simple M/M/1 queuing can be used to model the sink node.

5.1 Channel capacity estimation

For an M/M/1 queuing model, the waiting time in the system, that is, the queuing time in the buffer plus the packet transmission time, is given by

$$E\{T_S\} = \frac{1}{\mu - \lambda_t} \qquad (1)$$

where E{T_s}is the waiting time in the system, λ_t is the packet arrival rate and μ is the packet rate at the output of buffer.

The output rate μ can be written in function of the link capacity C as

$$\mu = \frac{C}{E\{X\}}, \qquad (2)$$

where E{x} is the average packet length in bits.
Rewriting E{T_s}in function of C and E{x}and separating C, the following expression can be written

$$C = \frac{(1 + E\{T_S\}\lambda_t)E\{X\}}{E\{T_S\}}. \qquad (3)$$

5.1.1 Numerical examples

For the illustration of above equation, some numerical examples will be given. As the design criterion it is assumed that the waiting time in the system, E{$T_{s,}$}, is small, so that, a packet spends short time at sink node. So, three values are considered, E{$T_{s,}$}= .1 sec, .5 sec and 1 sec. Assuming E {X} = 1000 bits, and varying the input packet rate λ_t the link capacity, C, can be estimated.

Figure 6 shows the estimation of capacity for three values of E{T_s} in function of input packet rate. As can be seen in the figure, for smaller waiting time in the system, a greater link capacity is required. For example, for λ_t = 10 the necessary link capacities are 20 kb/s, 12 kb/sec, and 11 kb/sec, for waiting time of .1 sec, .5 sec and 1 sec, respectively.

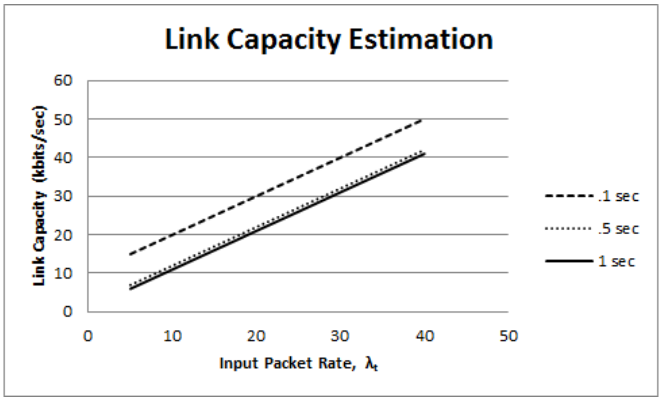


Figure 6. Link Capacity in function of input packet rate for various values of waiting time in the system.

5.2 Number of RFID tags estimation

The number of RFID tags can also be estimated, considering that the capacity is given. Using the Eqs. 1and 2, λ_t can be written as

$$\lambda_t = \frac{E\{T_S\}\dfrac{C}{E\{X\}} - 1}{E\{T_S\}} \qquad (4)$$

5.2.1 Numerical examples

Considering the waiting time in the system, E{T_s }= .1 sec, and the channel capacities of 50 kb/s, 100 kb/s and 200 kb/s, the total input packet rate, λ_t, using Eq. 4, will be 40 pkt/s, 90 pkt/s and 190 pkt/s, respectively. Considering an estimation of 1 pkt/s for each RFID tag, the numbers of RFIDs that can be accommodated are 40, 90 and 190. Table 1 shows the number of RFID tags that can be accommodated for other values of E{T_s}.

Table 1. Numbers of RFIDs tags in function of link capacity considering 1 pkt/s rate per RFID

	Capacity		
E{T_s} (sec)	**50 kb/s**	**100 kb/s**	**200 kbp/s**
.1	40	90	190
.5	48	98	198
1	49	99	199

6 Conclusions

In this paper, a network configuration using IoT technology for application in healthcare is proposed. The proposed configuration is appropriate for medical care of patients in their own homes. The main concept was to combine the WBAN network that monitors the vital signs of a patient with RFIDs. RFID readers placed strategically close to the patients, collect data from the WBAN, and through a clustered configuration of these readers, the data is transferred to a sink node, and then transmitted to a gateway. The gateway has an Internet connection, and thus the data of patients can reach medical facilities where they can be processed and presented to physicians for monitoring purposes.

In this article, the main points of the network are detailed, and some design considerations for good operation of network are pointed out.

The link capacity of sink node and number of RFID tags are estimated using simple queuing model for good operation of the network.

In future work, the best kind of routing in the cluster of RFID readers to deliver the packets to the sink node will be studied.

7 References

[1] Yang, D., Liu, F., Liang, Y. (2010). "A Survey of the Internet of Things". International Conference on E-Business Intelligence (ICEBI-2010): *Advances in Intelligent Systems Research*, pages 358 – 366. Atlantis Press.

[2] Atzori, L., Iera, A. and Morabito, G. (2010). "The Internet of Things: A Survey". *Computer Networks*, pages 2787-2805 Vol. 54, No. 15.

[3] ITU International Telecommunication Union (2005): "The Internet of Things. Executive Summary", http://www.itu.int/osg/spu/publications/internetofthings/, July.

[4] Fleisch, E. (2010). "What is the Internet of Things? An Economic Perspective". *Auto-ID Labs White Paper WP-BIZAPP-053*.

[5] Wu, M., Lu, T., Ling, F., Sun, J., Du, H. (2010). "Research on the architecture of Internet of things". *Advanced Computer Theory and Engineering (ICACTE).* 3rd International Conference, pages V5-484 - V5-487.

[6] Mazhelis, O., Luoma, E., Warma, H. (2011). "Defining an Internet-of-Things Ecosystem" *Springer-Verlag*, Heidelberg, Berlin.

[7] Auto-Id Labs. (2013) Available: http://www.autoidlabs.org/.

[8] Xue, X., Li, G., Liu, L., Liu, M. (2012). "Perspectives on Internet of Things and Its Applications". *2nd International Conference on Computer Application and System Modeling*. Atlantis Press, Paris, France.

[9] Vermesan, O., et al. (2011). "Internet of Things Strategic Research Roadmap". *European Research Cluster*. http://www.internet-of-things-research.eu/

[10] Bauer, M., et al. 2011. Introduction to the Architectural Reference Model for the Internet of things. First Reference Model White Paper. *IOT-i The Internet of Things Initiative.* [Online] Available: http://ww.iot-a.eu/ [Accessed 14 July 2013].

[11] Zhao, W., Wang, C., Yorie, N. (2011). "Medical application on internet of things". *Communication Technology and Application* (ICCTA 2011), IET International Conference, pages 660 – 665.

[12] Briseno, M., Cota, C., Garcia, E., Lopez, J. (2012). "A proposal for using the internet of things concept to increase children's health awareness". *Electrical Communications and Computers (CONIELECOMP*). 22nd International Conference, pages 168 – 172.

[13] Rohokale, V.M, Prasad, N.R, Prasad, R "A Coopera tive Internet of Things (IoT) for Rural Healthcare Moni toring and Control", Wireless Communication, Ve hicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE), 2011 2nd International Conference, pages 1-6.

[14] Chia, S., Zalzala, A., Zalzala, L., Karimi, A. 2011. RFID and Mobile Communications for Rural e-Health. A community healthcare system infrastructure using RFID for individual identity. IEEE Global Humanitarian Technology Conference. Computer Society, pp 371 – 376.

[15] S. Motoyama, "Hierarchical Polling-based MAC scheme for Wireless Body Sensor Network" in International Conference on Wireless Networks - ICWN-13, 2013, Las Vegas. pp. 103-109.

Detecting Boundary Nodes in WSN

Nihel SENOUCI, Moustafa KOUIDER EL OUAHED, Hafid HAFFAF
Department of Computer Science
University of Oran
Email: nihel.senouci@live.fr, moustafakouider@gmail.com, haffaf_hafid@yahoo.fr

May 31, 2014

1 Abstract

The Wireless Sensors Networks (WSN)are widely used in many domains such as security, health or environmental survey. WSN offer many advantages but some problems such as the geographical voids need to be solved. Our paper focuses on the detection of the network external border by using a new mechanism of identification of trigger nodes. The suggested mechanism is built on 3 phases. The role of the initial phase is to identify one or more border nodes, the second serves to discover the network border and the last one is devoted to exclude border nodes from the routing. This exclusion does not preclude the border nodes to play their role of sentinel against malicious intrusions. Compared to the previous methods, our approach uses more than one trigger node and the detection of the network border operates in two directions which reduce the detection time and increase lifetime of the network.

Keywords *Wireless sensor networks, routing, geographical void, open void.*

2 Introduction

Sensors permit to link the physical world with the digital environment. The evolution of wireless technology has led to the development of various derived architectures, such as cellular networks or wireless local networks. During the last decade, a new architecture has emerged: the wireless sensor networks (WSN).

A wireless sensor network (WSN) consists of a set of nodes able to communicate via wireless links. The overall goal of a WSN is to collect data from the environment around the sensors and communicate these data to a central processing station or sink. They have also storage capabilities.

WSN are often considered as the successors of ad hoc networks. Due to their ability to respond to real needs, WSN have attracted an increasing number of industrial applications. The need for continuous monitoring of a specific environment is important in various human activities: Industrial processes [4], monitoring of habitat [3], agriculture [5], natural ressources management [9], health monitoring [11], disasters [9], military applications of tracking [9] are the main domains of applications offered by WSN.

However, their development still faces obstacles that are real challenges for scientific research. Sensor networks suffer from several problems. Among these problems, we can quote first the energy constraint [16] because WSN operate with limited energy budgets. The problem of batteries is crucial and the choice of the energy design depends on the type of application in order to ensure the required efficiency. Another challenge is the self-management [7] because in many applications WSN must operate in remote areas and harsh environment, but they need to maintain their efficiency. WSN need also to ensure a transmission power which is a function of the design of the network. Regarding the problem of security [14] it is important to underline that WSN collect sensitive information. Then it is important to preserve them from malicious intrusions or attacks.

Propagation and delivery of data in a WSN is the most important feature of the network. Routing protocols for WSN have been studied extensively, and several studies have been published [1]. The methods can be classified according to either network topology criteria or to establishment of the road. In some realistic situations, existing routing protocols often become inoperative in certain situations such as the presence of a geographical void resulting from a random deployment of sensor nodes.

In this paper, we are interested in void avoidance problem. A geographical void can be concave or convex. It can be inside (internal void) or outside (external void) the network. Indeed, the presence of a void in a network leads to failure of the routing protocol and the packet loss during the transmission. This could have a disastrous impact on a strategic network such as a monitoring network. Hence, we propose an improved method of detection of external void as a support for a new routing protocol, which in turn aims, to identify the nodes forming the network border before starting the routing. We shall see that our method is able to reduce the failure of routing protocols, packet loss and then, increasing the lifetime of the network.

The paper will focus on the following sections. The section 1 presents the state of the art. In section 2 we describe our method in more details and present simulation results in section 3 followed by section 4 where our perspectives and conclusion are presented.

3 Related works

The main missions of WSN are to monitor a region and collect regular measures in order to send them back to the sink. To achieve these objectives, one of the major challenges in such networks is the development of efficient routing protocols that can ensure the establishment of effective routes at any time and quality communication between network nodes. There are three classes of routing protocols, flat routing [2], hierarchical routing [2] (LEACH (Low-energy Adaptive Clustering Hierarchy) [17]) and geographic routing [13]. According to the process of path discovery, there are 3 main types of routing protocols, proactive (OLSR [8], LEACH [17]), reactive (AODV [6]) and hybrid (ZRP)

Mainly dedicated to WSN, the Geographic Routing is based on geographic location information, instead of using the network address. The source sends a message to the geographic location of the destination. It differs from the classical topological routing since it uses the exact geographic locations of the nodes to take a decision on the routing of packets to be transmitted next in WSN. The location is generally obtained through a GPS system included in the sensors. This type of protocol doesn't require heavy routing tables and then, this approach is more efficient in terms of energy consumption.

GPSR (Greedy Perimeter Stateless Routing) [10] is one of the most used protocols. It uses the positions of routers and a packet's destination to make packet forwarding decisions. GPSR makes greedy forwarding decisions (represented in Figure 1 [10]) using only information about a router's immediate neighbors in the network topology, greedy choice in choosing a packet's next hop. Specifically, if a node knows its radio neighbors' positions, the locally optimal choice of next hop is the neighbor geographically closest to the packet's destination. When a packet reaches a region where greedy forwarding is impossible, the algorithm recovers by routing around the perimeter of the region. This algorithm uses the right hand rule (See Figure 2 [10]) by routing packets around the border of the void until they reach the other end of the void.

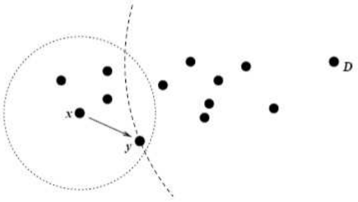


Figure 1: Greedy forwarding [10]

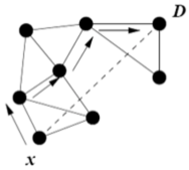


Figure 2: Perimeter forwarding [10]

PGR (Probabilistic Geographic Routing) [15] uses only local information to probabilistically forward the packet to the next hop. Every node relies on a beaconing process to keep track of the changes in the set of its neighbors. In order to forward a packet, the node selects a subset of its neighbors. These candidate nodes are then assigned a probability proportional to their residual energy and the link reliability.

These algorithms are effective and provide an easy use, but sometimes they cease to be operational in presence of a geographical void which represents a region where the nodes are no able to carry the package. This void is the result of a random deployment of nodes or failure of a circuit, energy depletion or destruction of a node.

Solving the problem of voids in geographic routing dedicated to WSN during information routing is still a technological barrier that remains open especially in survey networks. To overcome this problem, there are mechanisms of voids detection in WSN, namely Void Boundary Discovery (VBD) [12] and Network Boundary Discovery (NBD)[12].

The VBD mechanism (Void Boundary Discovery) is based on the identification of the nodes forming an internal void in a WSN and the calculation of its center and radius. The principle of this method is to route a packet VD (void discovery) around the void. The mechanism stops once the packet has done a complete turn around the void (See the Figure 3 [12])

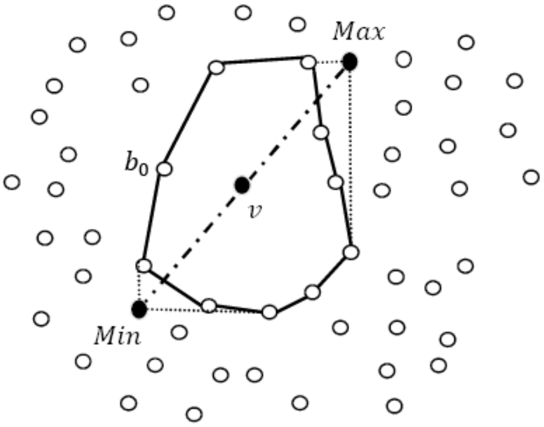


Figure 3: Void Boundary Discovery [12]

To identify the nodes forming the border of the network, the mechanism NBD(Network Boundary Discovery) is used to identify a boundary nodes and calculate the center and the radius of the external void.

The principle of the method is to indicate a fictive destination which will trigger the mechanism by routing a packet ND (Network Discovery) around the large void in one direction. (See the Figure 4 [12])

These two mechanisms are efficient but their performances are limited in the case of network monitoring where real time problem arises, in the case of malicious intrusions, or in the situation where there are a huge number of nodes. In all these cases, the void discovery could lead to a waste of time.

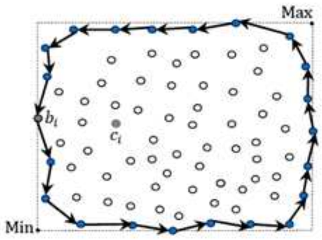


Figure 4: Network Boundary Discovery [12]

4 Our contribution

Our idea is to discover the external void before starting routing. The strategy is inspired by previous works on NBD (Network Boundary Discovery) as presented in the state of the art, but the specificity is to detect more than one trigger node and bidirectional routing scheme.

This method includes three complementary phases. The first one is the initiation of the process, the second is the discovery of the network border and the last is the exclusion of border nodes.

4.1 Phase 1: Initiation of the process

The role of this phase is to initiate the process by identifying one or more border nodes. Border nodes of the network are located at the farthest distance from the sink. Each sensor node in a network has its neighbor table.

In the first step, each node calculates its distance from both the sink and its immediate neighbors. The node that is the farthest from the sink represents a border node. The latter is called the boundary node. To accomplish this task, we add a boolean field (isBoundary) set to ”false”. A node is called ”boundary node” when the ”isBoundary” field is set to ”true”.

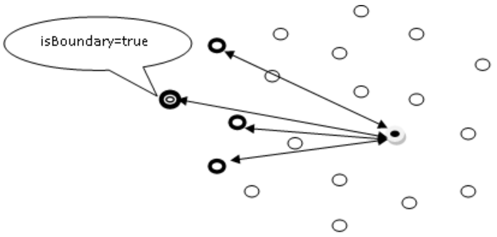


Figure 5: Initiation of the process

4.2 Phase 2: Discovery of the network border

When a border node is found, it must trigger the discovery process and bypass the external void.

4.2.1 Construction of the neighboring groups L and R (Left and Right)

The trigger node resulting from Phase 1 must know its right neighbors (R Right neighbor group) and left neighbors (L Left neighbor group). To do this, we use a geometric method to found points (whose coordinates are known) located below and above a specific line.

If $y_1 \succ ax_1 + b$ than A is on the LEFT of the line If $y_1 \prec ax_1 + b$ than A is on the RIGHT of the line If $y_1 = ax_1 + b$ than A is on the line (see the figure 6)

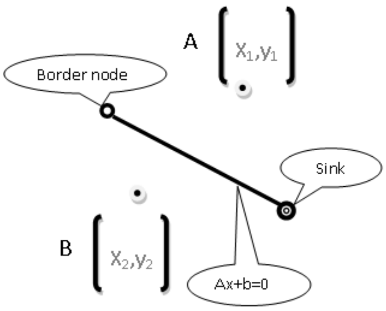


Figure 6: Geometric method

Once the two groups of neighbors L and R are formed, the trigger node creates a BP (Packet Boundary) package, by including a simple message "I'm boundary node" . To accelerate the process, the trigger node will send the packet to its immediate neighbor lying farthest from the Sink (border node) right and left.

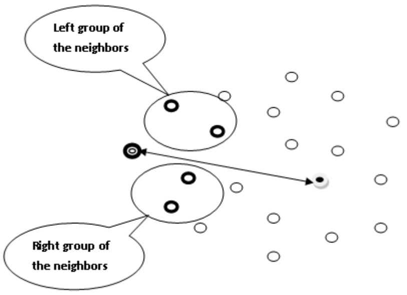


Figure 7: Right and left neighbor groups

4.2.2 Sending BP (Boundary Packet)

Once the trigger node has found its neighbor nodes located at the farthest from the sink (right and left), it must simultaneously send them the BP.

In a wireless network, when a node A sends a packet to another node B, all nodes located on the same radio range than A will automatically receive the packet, and check if it is intended to them or not by looking at the destination address specified on the package. For this, we add a column in the neighbor table. The neighbor table contains the identifier of the neighbor and its geographical position. This new column contains a new Boolean field "state" set to false. The node which receives the BP compares its address and the destination one. If they are the same, it understands that it is a border node and put its isBoundary field to true. Otherwise, it means that the node is not a border node but its transmitter is a border node, so it must make the neighbor transmitter state field true.

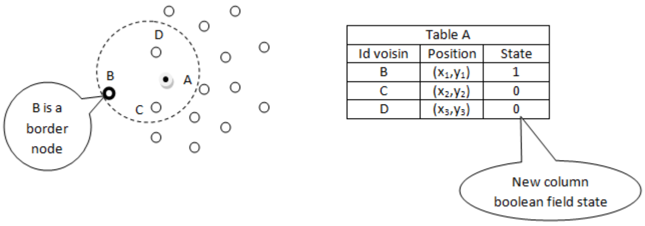


Table A		
Id voisin	Position	State
B	(x_1,y_1)	1
C	(x_2,y_2)	0
D	(x_3,y_3)	0

Figure 8: New table of neighbor

Our goal is to route a BP around the border of the network. The first node RIGHT which has received the BP packet forms its neighboring left and right groups, calculates the distance from its right neighbors relative to Sink and it transmits packet to the node farthest from the Sink (from the RIGHT neighbors) and so on. Regarding the first LEFT node which has received the BP packet, it forms its neighboring left and right groups too, calculates the distance from its left neighbors relative to Sink and it transmits the packet to the farthest node from the Sink (from the LEFT neighbors).

This mechanism allows us to stay on the border of the network; it will stop once a node receives the packet BP twice.

When a node is identified as a border node (Phase 1), it triggers locally the mechanism. In our approach, several triggers nodes may occur in the network and this makes our method robust.

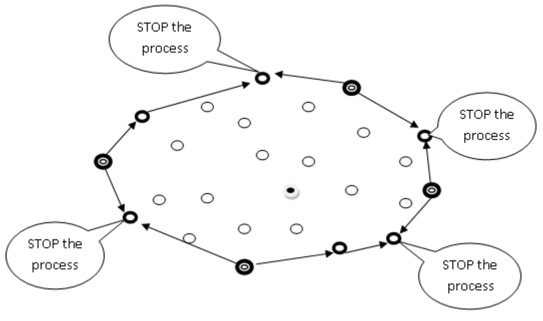


Figure 9: End of the process

4.3 Phase 3: Exclusion of border nodes from the routing

At this stage, the border of the network is detected and border nodes are identified. During routing, they will forward packets only to nodes whose state = false. This exclusion does not preclude the border nodes to play their role of sentinel against malicious intrusions.

Our method is sumarized by the organigram presented in Figure 10

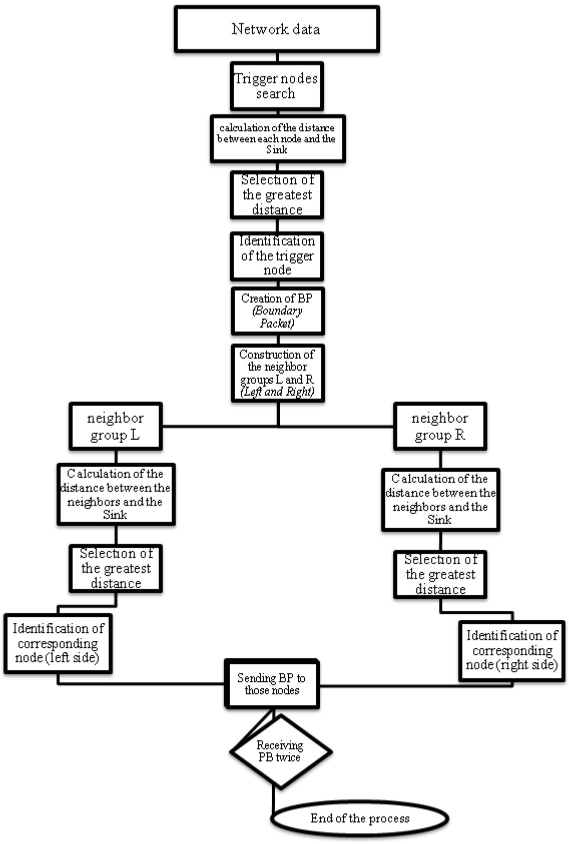

Network data
Trigger nodes search
calculation of the distance between each node and the Sink
Selection of the greatest distance
Identification of the trigger node
Creation of BP (Boundary Packet)
Construction of the neighbor groups L and R (Left and Right)
neighbor group L
neighbor group R
Calculation of the distance between the neighbors and the Sink
Calculation of the distance between the neighbors and the Sink
Selection of the greatest distance
Selection of the greatest distance
Identification of corresponding node (left side)
Identification of corresponding node (right side)
Sending BP to those nodes
Receiving PB twice
End of the process

Figure 10: Organigram

4.4 Experimentation

In order to test and implement our method, we used the Castalia / Omnet++ simulation platform. Castalia is a simulator for wireless sensor networks (WSN), networks of BAN (Body Area Networks) and generally networks of components with limited power. It is based on the platform OMNeT++. The structure of the source code of Castalia is hierarchical. Each module is a directory that contains a C++ code to describe the behavior of the module.

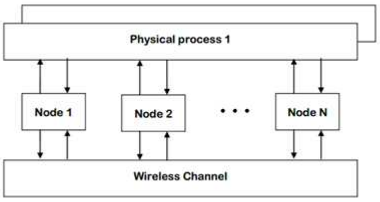


Figure 11: Castalia

OMNeT is a discrete event simulator object where the modules are programmed in C++. It a tool widely used for the simulation of communications networks. Omnet++ includes simple and compound modules that communicate with each other by sending messages that may represent packets, frames of a computer network...The structure of a module is defined by the user using the Omnet++ topology language NED. Also Omnet++ environment includes a graphical editor.

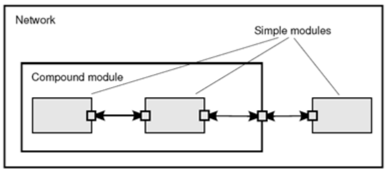


Figure 12: Omnet

To test our method, we used a video sensor type network (videosensor) consisting of 20 nodes. The implementation revealed 6 triggers nodes in the network, which generates a significant time saving.

To measure the effectiveness of our approach, we constructed a curve that estimates the time savings compared to the number of triggers nodes. Other curves will be needed to complete this analysis (energy saving ... etc)

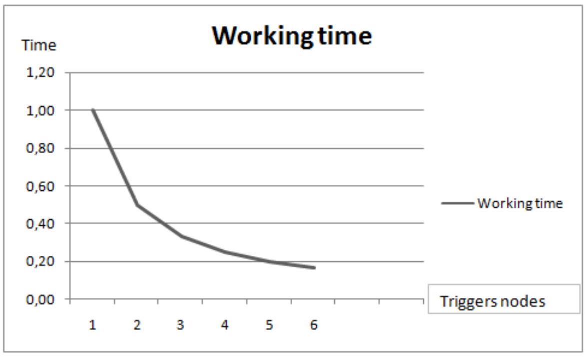


Figure 13: Curve

5 Conclusion and perspectives

Solving the problem of voids in geographic routing in WSN is the subject of considerable research and constitutes a technological barrier that remains still open. Our problem is linked to this specific scientific context.

This article is dedicated to the first part of our global experiment. A second part will be devoted to the identification of edge nodes and their exclusion before starting the routing itself.

With the early discovery of the border of the network, we avoid failure of routing protocols, we increase the lifetime of the network, and we reduce the risk of malicious intrusions. The identification of several triggers nodes reduces the census time of border nodes.

These results are encouraging and should guide us after finalizing experimentation towards improvements to adapt this approach to different types of topologies and solve the case of internal voids in a network. This is let to our future work.

References

[1] A.E. Al-Karaki, J.N. Kamal. Routing techniques in wireless sensor networks: a survey. *Wireless Communications, IEEE*, 2004.

[2] A.E. Al-Karaki, J.N. Kamal. Routing techniques in wireless sensor networks: a survey. *Wireless Communications, IEEE (Volume:11 , Issue: 6)*, pages 6 – 28, Dec. 2004.

[3] Joseph Polastre Robert Szewczyk John Anderson Alan Mainwaring, David Culler. Wireless sensor networks for habitat monitoring. *Proceeding WSNA '02 Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 88–97, 2002.

[4] Boon M.C. ; Green P.N. ; Green P.R. ; York T.A. Antoniou, M. Wireless sensor networks for industrial processes. *Sensors Applications Symposium, 2009. SAS 2009. IEEE*, pages 13 – 18, 2009.

[5] A Baggio. Wireless sensor networks in precision agriculture. *ACM Worshop Real-World Wireless Sensors Nerworks*, 2005.

[6] E. Royer C. Perkins and S. Das. Ad hoc on demand distance vector (aodv) routing. *Available : http://www.ietf.org/internet-drafts/draft-ieftmanet-aodv-03.txt*, 1999.

[7] D. Cerpa, A. Estrin. Ascent: adaptive self-configuring sensor networks topologies. *Mobile Computing, IEEE Transactions*, 2004.

[8] T. Clausen and P. Jacquet. Optimized link state routing protocol olsr. *IETF RFC 3626*, October 2003.

[9] Tian He John A. Stankovic, Anthony D. Wood. Realistic applications for wireless sensor networks. *Monographs in Theoretical Computer Science. An EATCS Series*, pages 835–863, 2011.

[10] Karp.B and H. T. Kung. Greedy perimeter stateless routing for wireless networks.

[11] P. Kulkarni and Y. Ozturk. Requirements ans design spaces of mobile medical care. sgmobile mob. *Comput. Commun. Rev. 11(3)*, pages 12–30, 2007.

[12] N. Badache M. Aissani, A. Mellouk and M. Boumaza. A novel approach for void avoidance in wireless sensor networks. *International Journal of Communication Systems (IJCS)*, 2010.

[13] Ke Liu Nael Abu-Ghazaleh, Kyoung-Don Kang. Towards resilient geographic routing in wsns. *Proceeding: Q2SWinet '05 Proceedings of the 1st ACM international workshop on Quality of service & security in wireless and mobile networks*, 2005.

[14] A.K. Hyung-Woo Lee ; Choong Seon Hong Pathan. Security in wireless sensor networks: issues and challenges. *Advanced Communication Technology, 2006. ICACT 2006. The 8th International Conference*, 2006.

[15] Shankar Sastry. Probabilistic geographic routing protocol for ad hoc and sensor networks,.

[16] Hien M. Nguyen Thai T. Vu, Viet D. Nguyen. An energy-aware routing protocol for wireless sensor networks based on k-means clustering. *AETA 2013: Recent Advances in Electrical Engineering and Related Sciences*, 2013.

[17] A. Chandrakasan W. Heinzelman and H. Balakrishnan. Energy-efficient communication protocol for wireless sensor networks. *Proceeding of the Hawaii International Conference SystemSciences, Hawaii*, January 2000.

An Energy based Routing Algorithm using the Centers of Local Clustering in Wireless Sensor Networks

Chung Sei Rhee
Department of Computer Science
Chungbuk National University Cheongju
Chungbuk, KOREA
csrhee@cbnu.ac.kr

Ming He Jin
Department of Computer Science
Chungbuk University
Chungbuk, KOREA
badistuta520@hanmail.net

Abstract

Recently, lot of researches for the multi-level protocol have been done to balance the sensor node energy consumption of WSN(Wireless Sensor Network) and improve the node efficiency to extend the life of the entire network. Especially in multi-hop protocol, a variety of models have been proposed to improve energy efficiency and apply it to WSN protocol. In this paper, we analyze LEACH algorithm and propose new method based on centers of local clustering routing algorithm in wireless sensor networks. We also perform NS-2 simulation to show the performance of our model.

Key Words : WSN, Clustering, Radio Wave Radius, Low-Power, Multi-hop, Centers of Local

1. Introduction

WSN(Wireless Sensor Network) technology has developed sensor nodes with small size, low price and low electricity due to the development of the information technology. Wireless Sensor Network has been studied for the military service, but recently it applied for a variety areas such as environment / ecology supervision, energy management, physical distribution / stockpile management, battle area management, medical monitoring and so on[1, 2].

Wireless Sensor Network has some problems such as low speed, fallacy, limited electric power, and replacement difficulty due to random sensor node distribution of wireless products. To overcome these kinds of problems, we should design to extend the life span of the whole network and distribute the energy which is concentrated on a few sensor nodes to the whole network.

Recently a variety of researches have been done for improving the energy efficiency. However, the existing techniques did not applied to the real sensor network construction. The reason is because the performance of the sensor node has been highly idealized. For example, to represent Wireless Sensor Network, LEACH assumes that the collected and integrated data communicate with sink node directly. And the sensor nodes also can control the sending energy actively according to the distance between the adjacent nodes. Therefore, to apply the wireless sensor network to the real sensor network, we have to consider not only simple network technique, but also restriction of the Electromagnetic Engineering.

In this paper, we compare and analyze the advantages and disadvantages of the existing techniques. Then, we will propose the routing algorithm which finds the center of local clustering in WSN. We consider the multi-level as well as the remaining energy of the node center of local clustering.

2. Hierarchical Routing protocols

Hierarchical or cluster-based routing, originally originated in wired networks well-known techniques with special advantages related to scalability and efficient communication as shown in figure 1. Therefore, the concept of hierarchical routing is also utilized to perform energy efficient routing in WSNs[3, 4]. In a hierarchical architecture, high energy nodes can be used to process and send the information while low energy nodes perform the sensing the proximity of the target. This means that creation of clusters and assigning special tasks to cluster heads can greatly contribute to overall system scalability, lifetime and energy efficiency. Hierarchical routing is an efficient way to lower energy consumption within a cluster and perform data aggregation and fusion in order to decrease the number of transmitted messages to the BS. Hierarchical routing is two-layer routing method where one layer is used to select cluster heads and the other layer is used for routing. However, most techniques in this category do not consider routing, but consider "who and when to send or process/aggregate" the information, channel allocation etc., which can be orthogonal to the multi-hop routing function [10].

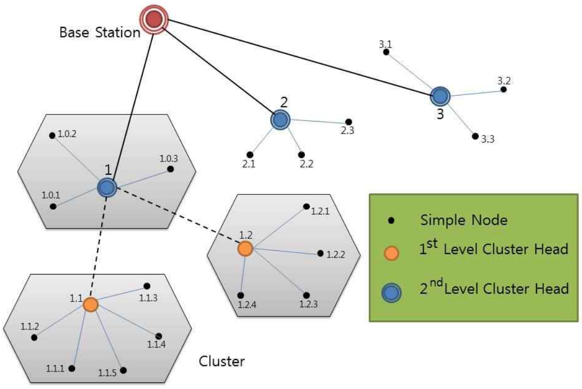


Figure 1. Hierarchical Routing Protocol

The typical methods of the hierarchical routing protocol can be found in LEACH [5, 6], TEEN [7], APTEEN [8], and PEGASIS [9]. LEACH (Low-Energy Adaptive Clustering Hierarchy) [4] is a clustering-based routing scheme in which the cluster head collects data from member nodes, gathers data through "data fusion" and directly sends them to the sink. This scheme is characterized by randomly circulating the cluster head which performs energy concentrative function to fairly distribute energy consumption to all sensors in network and by making local fusion of data collected in the cluster head from cluster to decrease total communication cost. Performance of LEACH depends on the fixed number of clusters in each round and it allows the cluster heads to be equally arranged, but it cannot be ensured with self-selecting method. Therefore, LEACH-C scheme was proposed to determine cluster head and cluster, depend on the location information on sensor nodes and amounts of preserved energy in the sink.

LEACH protocol decides cluster header using equation (1).

$$T_{(n)} = \begin{cases} \dfrac{p}{1 - p(r \bmod \frac{1}{p})} & \text{if } n \in G \\ 0 & otherwise \end{cases} \qquad (1)$$

The location of cluster headers are decided randomly from the censer field, therefore distribution of cluster headers is either uniformly distributed or congregated. If cluster headers are congregated, some cluster header contain too many member nodes. The distance between member nodes to cluster header are increased, therefore the energy efficiency is decreased rapidly. Furthermore, effective data merge and data quality in clustering are lost from ununiform cluster header. To solve this kind of problem, we propose new center of local clustering algorithm.

3. Centers of local clustering Routing Algorithm

We propose a cluster routing algorithm Which is operated in three steps such as cluster selection, creation of cluster and communication between clusters.

Algorithm is performed in round basis similar to LEACH algorithm. We create all the censor nodes on the network and perform first round. The first round is divided into information collection step, cluster creation step and data transmission step as shown in Figure 2. All the nodes send their location and energy information in the first round. Sink nodes perform simulator annealing and decide header list and each node's location identification after receive all the sensor nodes' message. Then sink node transmit this message to all the nodes in network. Sink node decides next hop based on header node's location and distance between header nodes. Cluster header performs header selection and cluster creation work instead of sink node from next round.

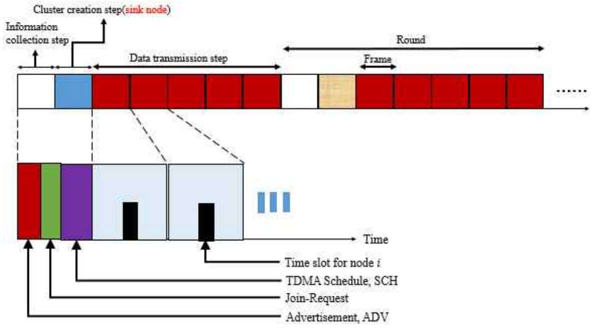


Figure 2. Structure of first round in center of local clustering Routing

Sink node uniformly cluster all the sensor network based on entire network information in first round. As we indicated earlier in the global central control method, all the sensor nodes transmit short message to the sink node then sink node does the clustering. But this method consumes too much energy. To solve this kind of problem, cluster header select new header instead of sink node.

Each node transmit node's geographic information as well as current energy to cluster header node during node's time slot in information collection stage. The received cluster header node executes dummy center algorithm and select new cluster header. New cluster node is selected based on two conditions. First, it is located on good geographic location. Second, it has the highest remaining energy. When cluster header run the dummy center algorithm, it decides energy threshold. We can find nodes which has an energy less than threshold. Generally, average energy is selected as a threshold. Selected new cluster header node is informed to header node. New cluster header node transmit ADV message to all the nodes using CSMA/CA MAC algorithm. All the other node know new cluster header node and previous header node is changed to normal node. All the nodes receive ADV message from new cluster header and decide to join some header based on RSSI(Received Singnal Strength Indicator). Then, nodes inform it to cluster header.

After all the cluster headers receive join request, they make a schedule and transmit it to all the nodes. After node receive the schedule, all the radio components are in relaxation stage except its time slot. Therefore, it can save lot of energy.

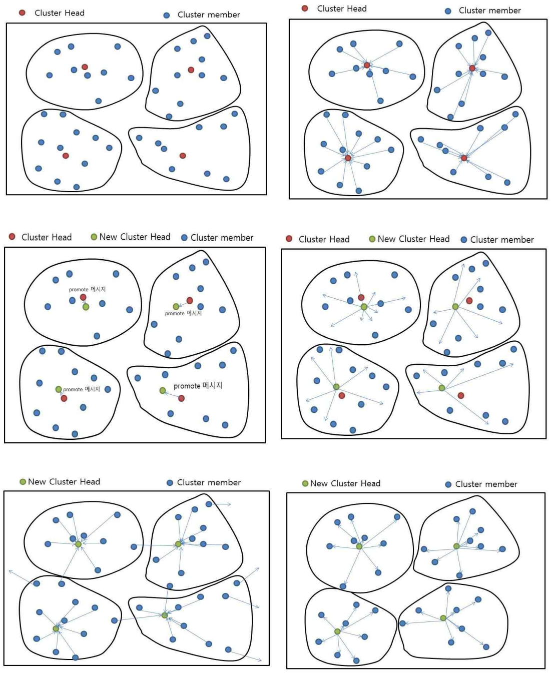


Figure 3. Cluster Creation Stage

The proposed algorithm use the hierarchical multi-hop technique which is different to original LEACH algorithm. As you find in Figure 4, the entire sensor network is divided into several clusters, then sensor nodes transmit data belong to it to cluster header node. Cluster header collects data from the member nodes, then transmit it to base station through the adjacent cluster header using hierarchical multi-hop. It is different compared to LEACH which directly transmit to base station.

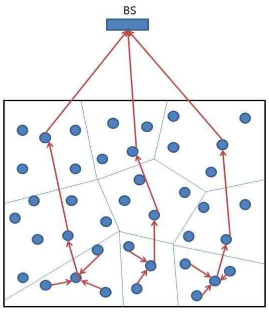


Figure 4. The structure of local-center routing method

4. Simulation and Performance

We use the NS-2((Network Simulator version-2)) tool to evaluate the performance of our method. We compare our method with multi-hop method and LEACH algorithm. Table 1. shows all the parameters and related values used in NS-2 simulation. Network size is 100m x 100m and 100 censor nodes are located randomly and we assume there are no node movements. The initial node energy is increased every 20 second for each round. Sink node is located in 50m in horizon and 175m in vertical. Cluster header node is selected from the highest energy nodes.

Table 1. System Environment and variables

Variable	Establishment
OS	Linux fedora 9.0
CPU	Inter Core Quad CPU Q8200
Memory	8GB RAM
Tool	NS-2
Size of Network	(0,0)-(100,100)
Total number of nodes	100
Base Station	(50,175)
Threshold distance	75m
E_{elec} (wireless Electronic Energy)	50nJ/bit
$\mathcal{E}$fs (Wireless energy increase in free space)	10 $pJ/bit/m^2$
$\mathcal{E}$mp (Wireless energy increase in multi-paths padding channel)	0,0013pJ/bit/m^4
Speed of data transmission	1Mbps
Data packet size	500 bytes
Broadcast packet size	25 bytes
Pack header size	25 bytes
Initial energy	2J/battery
Number of Nodes	100
Number of cluster	5

We perform simulation based on values in the table 1. Figure 5. shows the live nodes according to time span for the three cases. Figure 6. compare the energy consumption of three models. Finally, Figure 7. shows the data transmissions of LEACH, LEACH-C and the proposed model.

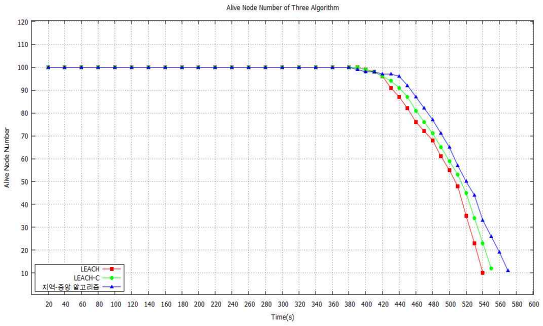


Figure 5. Live nodes of three model

The first occurrence of incompetent node is at 400 second, 390 second and 390 second for LEACH, LEACH-C and center of local algorithm. The last live node occur at 540 second, 560 second and 578 second. as you find in Figure 5. This implies the total sensor network' life is increased by 7 percent compared to LEACH algorithm and is increased by 4 percent compared to LEACH-C method. The energy consumption of proposed algorithm is 23 percent less than LEACH algorithm and 19 percent less than LEACH-C method. This occurs because the proposed algorithm distributes cluster headers more uniformly around the network.

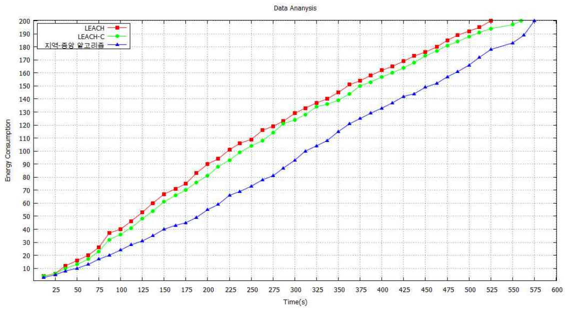


Figure 6. Energy consumption of three Model

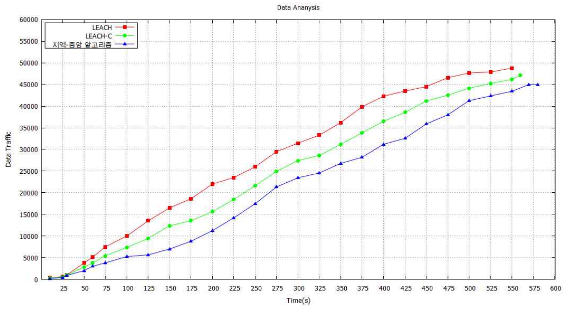


Figure 7. Data Transmission of three models

The data transmission of our algorithm is 31 percent less compared to LEACH algorithm and 16 percent less compared to LEACH-C method.

5. Conclusion

A lot of researches have been done to balance the sensor node energy consumption of WSN and extend the life span of the network.

In this paper, we extend the live span of network by using energy effectively in Wireless Sensor Network. We propose an algorithm which find the center of local clustering and compare with LEACH and LEACH-C algorithm for lfe span, energy consumption and data transmission. Simulation show our method has some improvement compared to LEACH and LEACH-C.

References

[1] V. Loscri, S. Marano, G. Morabito, “A Two-Levels Hierarchy for Low-Energy Adaptative Clustering Hierarchy (TL-LEACH).” Proceedings “VTC2005”, Dallas (USA), pp.1809-1813, Sept., 2005.

[2] Hnin Yu Shwe, Jiang xiao-hong, and Susumu Horiguchi, "Energy saving in Wireless Sensor Network", Journal of

Communication and Computer, vol. 6, no. 5, May 2009.

[3] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless Sensor Networks," Ad Hoc Network, Vol 3, pp. 325-349, 2005.

[4] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Network," proc of the IEEE wireless Communication, vol. 1, pp. 660-670, Oct,2002

[5] V. Raghunathan, C. Schurgers, S. Park and M. Srivastava, "Energy Aware Wireless Microsensor Networks," IEEE Signal Processing Magazine, 19(2) : 40-50, March, 2002

[6] W. B. Heizelman, A. P. Chandrakasan and H.Balakrishnan,"Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proc. of the Hawaii International Conference on System Sciences, pp. 1-10, January, 2000.

[7] O. Younis and S. Fahmy, "HEED : a Hybrid, Energy-Efficient, Distributed clustering approach for Ad-hoc Sensor Networks," IEEE Trans. on Mobile Computing, Vol. 3, no. 4, pp. 366-379, Oct. 2004.

[8] O. Buyanjargal, Y. Kwon, "Adaptive and Energy Efficient Clustering Algorithm for Event-Driven Application in Wireless Sensor Networks (AEEC)", Journal of Networks, Vol. 5, no. 8, pp. 904-911, Aug, 2010.

[9] A. Manjeshwar, D.P. Agrawal, "TEEN : A routing protocol for enhanced efficiency in wireless sensor networks," Parallel and Distributed Processing Symposium, Proc. 15th International 23-27, pp. 2009-2015, Apr, 2001.

[10] W. Heinzelman. " Application-Specific Algorithm Architectures for Wireless Network." Ph. D. thesis, Massachusetts Institute of Technology,2000.

[11] Y. Xu, J. Heidemann, and D. Estrin. "Geography-informed energy conservation for ad hoc Routing." In the Proceedings of the 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking(MobiCom'01). Rome, Italy, July 2001, pp. 70-84.

[12] Zhi-kun Wu. Research on Routing Protocols of Wireless Sensor Networks. M.S. thesis. Dalian University of Technology, 2006.

[13] H. Chan, A. Perrig. "ACE: An Emergent Algorithm for Highly Uniform Cluster Formation. Proc. First European Workshop Sensor Networks (EWSN), volume 2920 of LNCS. Springer, Berlin, Germany, pp .154-171, January, 2004

SESSION

WIRELESS NETWORKS AND SECURITY ISSUES + IOT

Chair(s)

TBA

Hierarchical-Based Measurement of Situation Awareness in the Internet of Things

Audrey Gendreau, Ph.D[1], Rita M. Barrios, Ph.D.[2]

[1] CS & CIS, Saint Leo University
Saint Leo, FL, USA
audrey.gendreau@saintleo.edu
[2] CIS-Cyber Security, University of Detroit
Mercy Detroit, MI, USA
barriorm@udmercy.edu

***Abstract*— Analysis of the security situation of networks is an important area in the information security research field. Furthermore, situation awareness is critical to the Internet of Things (IoT) given the limited lifetime, and autonomous nature of wireless sensor networks (WSN).**

In this work, a measurement for situation awareness for the IoT is presented. This measurement can then be used to determine the security situation of a multi-application self-determining network to facilitate the deployment of object applications to a secured environment.

Using a simulation to compare the presented measurement to a power based approach provides that the presented approach did select different zones than the zones selected within the controlled set. Thus, resulting in a more accurate indication of the current security state of the WSN.

Potentially, utilization of this approach can contribute to the situation awareness of the cyber space in general to support ubiquitous computing; however, fundamentally it is a diagnostic tool that can be used to facilitate security within IoT. Thus, on both a local and global scale, it has the potential to predict the effectiveness of the location when considering deployment of an object application to IoT

***Keywords*— Internet of Things, IoT, Situation Awareness, WSN, Wireless Sensor Networks**

I. Introduction

Situational awareness is the ability to identify a process and comprehend the critical elements of information to determine the network state [11]. The elements of information that are used to determine the network state are triggered by events, of which, location is a very important attribute [11]. Moreover, the risks and costs examined at each location on the wireless sensor network (WSN) contribute to a cost benefit analysis of the spatial situation.

This research presents a model of WSN security situation awareness, by providing a unique and efficient method to analyze the whole network sensor's security situation for application deployment to IoT. The proposed metric can be used as a holistic indicator to support situation awareness for a remote decisions concerning the quality of the network‘s ability to communicate. The spatial analyses of node placement, factors in risks and costs of the current object applications in the resource constrained WSN. By creating a minimum connected dominating set (MCDS) based on a metric that analyzes the monitoring capability of the nodes the network. IoT can be analyzed locally to collectively report its security situation in terms of device placement, power, and other risks and costs associated with the object applications on the network.

The effectiveness of the presented method was validated by comparing it against a type of a power-based scheme that used each node's remaining energy as the situation awareness indicator. To demonstrate that the metrics are different, the proposed IoT Index in those zones derived from the power-based approach was compared to the zones using the IoT Index as the metric. While available energy is directly related to the model used in the presented method, the study deliberately sought out nodes that were identified with having superior monitoring capability, cost less to create and sustain, and are at low-risk of an attack. This work investigated capturing the index after a temperature sensing object application had established the network traffic flow to the sink. In each scenario, after many packets were transmitted to the sink, the local IoT Indexes were derived from the communication history and the other network factors.

The simulation results show that the proposed approach is more sensitive of the network state than a general power-based metric. The power-based metric uses the network's remaining power as an indication of the security situation. Repeating the experiment by assigning risk values inversely proportional to the first test scenario, the results show that the network with less risk has a smaller index. The increased risk does not affect the power, therefore the remaining network power in both experiments is the same. Also, in both tests the more sparse networks displayed lower values. A sparse network can simulate an irregular horizontal plane. However, in a floor plan without obstructions the appearance of a sparse network can be an indication of other factors that may inhibit the sensors inability to communicate with one another.

A. Guide to the Paper

Section II present a brief review of the literature needed to set the foundations of the presented work. Section III presents a brief account of the research methodologies employed in this study. Section IV is an overview of the experiment. Section V presents the research findings while section VI presents a summary of the conclusions as well as avenues of future work.

II. *Prior Research*

A. Internet of Things (IoT)

The Internet of Things (IoT) is an integration of several technologies and communications solutions to enable complete end-to-end communication between digital objects. The basic concept is centers around that many “things” or objects are interconnected. These objects could be devices, networks, sensors, etc. which interact with each other to reach common goals. According to [3], IoT is considered to be one of the US National Intelligence Council’s (NIC) Six Disruptive Civil Technologies with the potential to impact the US’s national power. NIC states that by 2025, these nodes on the internet may reside in everyday objects such as food packaging, furniture, paper documents as well as many others [3].

While this highly connected environment poses increase risk, it also poses a great potential. We are expected to see changes in ongoing business models in healthcare, transportation, home appliances, critical manufacturing as well as the upcoming wearable technologies. [1] notes that in the last few years, the number of connected devices exceeded the number of people on the planet with some estimates topping out at over 10 billion devices.

With this great move in technology, many challenging issues need to be addressed concerning the security aspects. As more objects become connected, more vulnerabilities and threats will be identified. As a result of the increase in connectivity, the volume of data transported over the internet will also grow at an exponential rate. [1] tells us that as this growth occurs, we can expect to see more regulations around the security of the nodes as well as the data. In June 2013, the US FDA updated its 2005 Technology Draft Guidance on Cyber Security for medical devices to indicate the need to assure the security of medical device functionality with the increased usage of wireless environments [2]. Also, with this expanded usage of the internet, we can expect that as more and more devices become ‘intelligent’, there will be a widespread distribution of risk since some data may not go directly from the sender to the receiving server but rather to a local data collection hub that will store information temporarily and then upload it periodically to the receiving server [2]. This environment will give the malicious entity an increased level of targets available by attacking the hub which may not have the necessary security protocols.

B. Hierarchical-Based Model – Optimum Spatial Situation

Within each cluster of a hierarchical network topology, a clusterhead is a dedicated node that is the local coordinator for the cluster; it is responsible for the coordination of routing within a cluster, communicating between clusters, and supervising its members [9].

Since clusterheads are responsible for route coordination, the quality of clusterhead placement is a critical factor to ensure the local supervisory tasks are successful in achievement of the application’s goals. In a role-based hierarchical application, clusterhead placement is defined by selection criteria which are based on the rules of cluster formation as adopted by each application [8]. The objective is to place a supervisory agent in a position that will maximize the capability of the role, while minimizing the energy cost associated with the location [8]. This local control in a hierarchical WSN supports a self-reliant application that is free from a centralized command or human intervention.

As noted above, the success of the sensor agents are dependent upon placement. Several attributes have been known to influence the quality of sensor placement when positioning the supervisory role agents in a multi-application WSN [5]. It should be noted that a multi-application WSN is one where the WSN is used to facilitate more than one application. These attributes include:

- Minimization of inter-cluster redundancy: members that are two or more hops away from the clusterhead can receive the same broadcasted message more than once. To reduce the inter-cluster redundancy, one-hop memberships are required [5].
- Minimization of intra-cluster redundancy: To minimize intra-cluster redundancy, a fundamental objective of a clusterhead selection algorithm, a solution that most closely resembles a MCDS solution is recommended [18]. The MCDS solution seeks to minimize the number of clusterheads required to achieve full coverage [18].
- Cluster organization for communication optimization: To optimize communication links within a cluster, [6] found that a more centralized membership is favored. However, as shown by [17], some distributed and perimeter nodes are necessary to achieve improved monitoring.
- Adaptability: a self-configuring deployment algorithm that locally adapts to a change in the position of sensors by revaluating only the affected nodes is more energy efficient, a necessity for added mobility [15].
- Node vulnerability avoidance: All of the data collected in the network is funneled towards the sink. Therefore, the sink and the nodes close to it are vulnerable to attacks [4] [12].
- Monitoring cost awareness: related to risks, selecting clusterheads on high volume nodes can prevent real-time detection and limit scalability [4].

- Transmission energy consumption: packet transmission consumes the most significant amount of energy. The energy loss is directly proportional to the distance the packet travels [9].
- Clusterhead energy requirements: A clusterhead has lower energy requirements than its members. More energy is used by one node to relay all the inter-cluster packets to a clusterhead than the energy used to aggregate the data at the clusterhead and then send it as one packet [7].

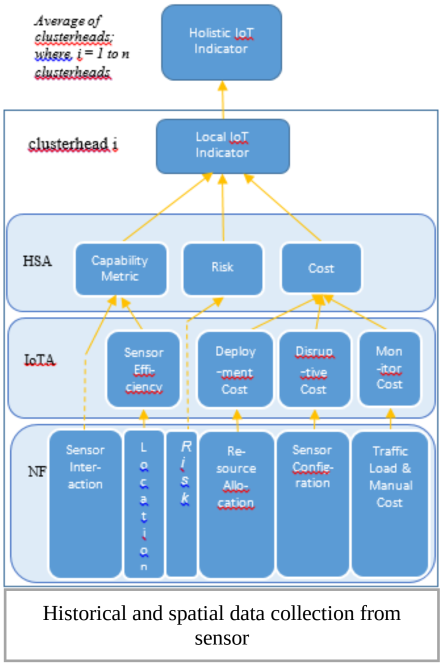
III. *Establishment of the Model*

An approach that analyzes the risks and costs to the established application [11] requires a self-assessment of a node's spatial situation. While this more comprehensive approach proposed by [14] for pairing IDS deployment with an optimum location on a Local Area Network (LAN), it was adopted for a hierarchically organized situational awareness application on a WSN. We will describe how a deployment value that represents compatible monitor-location pairings on a LAN was extended to facilitate agent placement for a situation awareness application on a WSN supporting the IoT.

To accomplish this on a different platform, the factors that distinguish the locations in a LAN were used to characterize the locations of clusterheads in a hierarchically organized virtual sensor network topology, thereby modeling the local index for the IoT. To organize a mobile cluster-based WSN, the IoT index was built on the following broad characterization (Fig. 1.) of the attributes and limitations of cluster formation. These included minimization of inter-cluster and intra-cluster redundancy, cluster organization for communication optimization, adaptability, node vulnerability avoidance, monitoring cost awareness, transmission energy consumption, and clusterhead energy requirements.

In figure 1, the first tier represents the Network Factors (NF) derived from the previously discussed optimum characterization of cluster formation. Built on the NFs, the perspective of the frameworks second tier was the effectiveness and efficiency of the Internet of Things Agent object applications (IoTA). The finial tier performed the Hierarchical Spatial Analysis (HSA) to form the parameters for the local IoT Index. The average of all of the clusterheads local IoT indexes created a holistic IoT Index to represent the whole network state (Fig. 1.).

Fig 1. Hierarchical-Based Measurement Model for Situation Awareness in the Internet of Things



A. Network Factors

The Network Factors included the sensor interaction abilities, location type, risk profile, resource allocation profile, sensor configuration costs, traffic load, and manual costs [14].

1) Sensor Interaction Abilities: To measure the interaction with the service layers, the fundamental functionality of each of the four TCP/IP suite service layers (Physical, Network, Transport, Application) of a LAN, were mapped to the ZigBee standards for a WSN. In a LAN, each of the four service layers was assigned the value [2.50] and was bounded by the minimum and maximum values [1, 10]. While A_I was denoted as the cumulative total of all the layers, the deterministic hierarchical spatial perspective approach is concerned with the application traffic flow generated by the virtual clusterheads. For this reason, it was represented for its overall contribution as 2.50 for its interaction at the application layer, $A_I = 2.50$.

2) Location Type: A sensor in a WSN acts as both a computer and router. Unlike IDSs distributed on a network to monitor a LAN, the agents designed for a particular role in a distributed application on a WSN tend to be the same. The same agent A, which is capable of monitoring, is deployed over the same WSN locations and is denoted as L_w. The subscript w is introduced to represent a WSN.

3) Risk Profile: In a LAN, the risk profile is the threat of a component being compromised. It is quantified based on the value of the asset and the likelihood of targeted attacks. Correspondingly, in a WSN, all of the data collected in the network is funneled to the sink; therefore, the sink and the nodes closest to it are most vulnerable to attacks and should be avoided when choosing clusterheads for monitoring capability [12].

As in the study by [14], a risk profile for location L_W in a WSN was denoted as $R(L_W)$ and expressed as a ratio relative to other locations within the network, with values ranging between [0,10]. Using the same values, but in a different context than a LAN, for the Local IoT Indicator it was suggested that for a WSN, there are four evenly distributed numbers between [1, 10] that signify the distance to the sink. Table 1 shows the ranked risk profiles of each node type.

Table 1: Risk profile values

Node Type	Risk	Value
Sink	High	1
1-hop from sink	Medium High	4
2-hops from sink	Medium Low	7
Remaining nodes	Low	10

A WSN is more vulnerable than a LAN, so none of the nodes are without some risk. It was therefore suggested that zero not be used. However, a more elaborate risk profile that takes into consideration the feed-back from an intrusion detection system and other systems could be used as defined by [19].

4) Resource Allocation Profile: A resource allocation profile must be established that identifies the amount of memory, storage, and CPU required for each node. An explanation is provided by Table 2. For a total range of values between [1, 10], it can be quantified by the summation of values defined as high, medium high, medium low, and low. For each attribute high = 3.33, medium high= 2.33, medium low = 1.33, and low =.33.

Table 2: Resource Allocation Profile

Attribute	Explanation
Memory	Amount of memory
Storage	Amount of storage
CPU	CPU usage

5) Sensor Configuration Costs: A sensor configuration profile assessed by the network factors that contribute to the configuration and downtime of a disruptive installation of an object application: This is restricted to a range of values between [1, 10]; where, the value 10 represents the highest level of disruption.

6) Traffic Load: The work load factor denoted the amount of processing, as a result of monitoring, at a specified location in a LAN. It is described as the capacity and usage on network links, or the processing load on a host [14]. In a role-based hierarchical multi-application WSN, the locations that have the highest load factor are:

- Sinks where WSN traffic is funneled towards the outside world;
- Neighboring nodes that are one or two hops from the sink;
- Sensors and neighboring nodes stimulated by a substantial amount of sensing activity in the area;
- Application clusterheads with large memberships;
- Application clusterheads with more than one-hop memberships, thereby creating inter-cluster redundancy that could generate duplicate messages;
- Overlapping application clusterheads creating intra-cluster redundancy to generate duplicate messages.

The load factor was restricted to a range of values [1, 10] to express the work load at different locations in a LAN [14]. For a WSN location, the load factor was denoted as L_F (L_W), which is the level of application traffic in each node.

Table 3: Traffic level load factor values

Traffic Level	Load
Very High	10
High	7
Medium	4
Light	1

7) Manual Intervention Costs: C_{MON}, in a LAN ranges between [1, 10]. Since a WSN must be independent of human, C_{MON}=1.

B. IoT Agent Object Applications

Building upon the wireless sensor network factors used to model the IoT Index are: Sensor Efficiency, Deployment Cost, Disruptive Cost, and Monitor Cost.

1) Sensor Efficiency: is the ability to accurately detect events of interest in a LAN [14]. Originally, it was subjective data based on the paired compatibility of the location on a LAN and the monitoring capability of the application. Alternatively, the measurement of monitoring capability in a cluster-based application on a WSN is dependent on the geographical positions. The cumulative sensing degree (CSD), a clusterhead selection value originally introduced in a study by [10], was extended to deterministically characterize the monitoring capability as a percentage in a hierarchical multi-application WSN. Good placement depends on the assignment of the selected node to a more advantageous position to monitor its neighboring nodes without wasting valuable energy. Instead of using exposure to model wireless monitoring, a variant of the CSD metric originally introduced in a study by [10], and named here as the cumulative monitoring degree (CMD) was used. It was modified for the IoT Index so that the nodes eligible for overlap were restricted to within the Euclidean distance of the wireless range. This algorithm typifies attenuation using a metric with a weighted computation. This quality of attenuation enabled it to be a representative of [6] preferred centralized membership. Also, shown to be important by the [17] study, the CMD metric was given weight to facilitate distributed and perimeter members which enhanced the monitoring capability. In this way, the monitoring percentage originally subjectively determined for a LAN was substituted with a deterministic one designed for a self-sufficient WSN. The sensor efficiency denoted as A_E = CMD, as in the original work by [14], ranged between [0.1, 1].

2) Disruption Costs: The disruption cost is denoted as $D(L_w)$ and is bounded by the range [1,10]. For locations in a LAN it is identified as the cost to deploy monitoring capability. Consequently, it represents the configuration changes and message delays associated with adding a sensor node to the Internet-Of-Things network to support the new object application.

3) Agent Deployment Cost: The cost of installing, configuring, and maintaining an agent is characterized in a LAN as the agent deployment cost, CDep(A), with values between [1, 10]. Unlike in a LAN, role specific agents deployed to a WSN tend to be all the same. However, on a WSN, the agent deployment costs are the added consequences of the configuration and message delays caused by the computing resources consumed. The effect of the act of intercepting the messages before passing them to their designated destination can be measured by the extra processing that agents on the newly deployed application specific sensor node perform. It was suggested that for the Local IoT Index agent deployment cost for an existing sensor is an average cost of value five, otherwise to justify the added work load it was a value as defined by the resource profile.

4) Object Application Monitoring Costs: The monitoring cost, $C_{Mon}(A)$, in a LAN is expressed as the product of the manual cost to monitor, multiplied by the level of traffic at the node, such as $cost_M(A,L_w) = C_{Mon} \cdot L_F(L_w)$. The level of traffic on the node is a cost issue because selecting clusterheads on high volume nodes can prevent real-time detection and scalability [12]. Therefore, $cost_M(A,L_w) = 1 \cdot L_F(L_w)$.

C. Hierarchical Spatial Analysis

To formulate the IoT index, the capability metric, risk and costs are determined.

1) Capability Metric: The capability metric is defined as the interaction ability of the monitoring capability and is denoted as $Cap(A)=A_I \cdot A_E$; therefore, $Cap(A) = 2.50 \cdot CMD$[14].

2) Risk: Risks are obtained from passed through risk profile.

3) Sensor Costs: In the denominator of the model for a LAN the total cost, $cost_T(A,L)$, represents the summation of the deployment and the monitoring costs. The deployment cost is the summation of the agent deployment costs and the network disruption costs for that location. Thus, each ranging [1,10] this cost can be denoted for a WSN location as:

$$cost_T(A,L_w)=(C_{Dep}(A) + D(L_w)) + cost_M(A,L_w) \qquad (1)$$

D. Local IoT Indicator

To model situation awareness agent placement in a WSN, the IDS A over a location L was converted to represent an object application agent A at location L_w. To represent this changed context, the Local IoT indicator for a WSN in place of the LAN value V(A,L) can be denoted as:

$$IoT(A,L_w) = (Cap(A) \cdot R(L_w)) / cost_T(A,L_w) \qquad (2)$$

The clusterhead selection algorithm proposed by [18] was used to define the zones and deploy the agents. The deployment algorithm selected clusterheads prioritized by higher Local IoT indicator values. When in competition for covering the same area, the clusterhead with the greater Local IoT indicator was chosen.

E. Holistic Perspective

The average of the Local IoT Indicators of the chosen clusterheads is used to compute a holistic perspective of the spatial state of the entire network while taking into consideration each local area's risks and costs. Based on the zones selected, the value of the Holistic IoT Indicator can be denoted as:

$$IoT_H = \frac{1}{n}\sum_{i=1}^{n}(IoT(A, L_w))_i \qquad (3)$$

$$where;\ i\ is\ clusterhead\ 1\ to\ n$$

IV. Experiment Overview

Using a Java program, two different simulated experiments were conducted based on data exported from OPNET. For each experiment, twelve different network sizes were tested using ten different WSN scenarios. The twelve test scenarios are listed under their associated dense, intermediate, and sparse deployment categories.

A. Simulation

For each test scenario, the process was separated into the following three identifiable phases:

To create the initial temperature sensing applications network topology, an exported OPNET file consisting of a list of node names and their coordinates was imported into a Java simulator. To select the clusterheads for the temperature sensing application, the nodes were ranked based on sensing accuracy. The resulting network topology with the selected clusterheads and their assigned members determined the routing tables, and thus, the initial network traffic flow in Java.

The established temperature sensing application transmitted a specific number of packets to the sink. After the packets reached the sink, the total number of bytes sent and received from each node and the distances they traveled were both used to compute the remaining power for each sensor.

Once the temperature sensing application had run, the network factors such as remaining power, traffic load, risk, and monitoring capability were defined for each node. These network factors were then used to compute the IoT index to determine the local network state for a hierarchical spatial perspective.

B. Metrics

In this process, a computer-generated temperature sensing application was established before the situational awareness application was deployed. Depending on the situational awareness experiment, the metric used was one of the following Power or non-Power (Local IoT Indicator) based equations to select the clusterheads that represented the zones in the hierarchically organized virtual topology:

1) Power Based: The transmission of the power consumed in transmitting and receiving a k-bit message, a distance d (using this radio model), was computed as:

$$E_{Tx}(k,d) = E_{elec} \cdot K + \varepsilon_{amp} \cdot k \cdot (d)^2 \text{ and} \qquad (4)$$

$$E_{Rx}(k) = E_{elec} \cdot k \qquad (5)$$

respectively, where E_{elec} is the energy the radio dissipates to run the transmitter or receiver circuitry and ε_{amp} denotes electronic energy expended in transmitting one bit of data. Therefore, to compute the remaining power for each node, the power consumed was subtracted from the initial amount of power given to each node.

2) Non-Power based (Local IoT Indicator): As discussed in section III, equation (2) can be used to compute a value that models security situation awareness at a particular location on a WSN. Using historical and spatial data collected from a sensor each node's monitoring capability, risk, and cost were analyzed to report on the locations network and security situation.

The purpose for the Power based clusterhead selection metric was to use it for comparison against the non-Power (Local IoT Indicator) based metric presented in this research. The remaining power was a popular metric used in extensions to LEACH and other hierarchical organized topology to increase the lifetime of a network. Moreover, as stated in [16], the IDS used Power to determine the duration a node can support the network monitoring role. Therefore, the Power metric was used as a benchmarking tool to determine if the objective had been met. Specifically, this objective was to find a situational awareness metric with the ability to determine the current state of a WSN. Therefore, after organizing the topology using the Power metric the Local IoT Indicator of those Power based clusterheads were averaged to determine if the situational awareness application using the Power metric picked the same zones.

C. Network Parameters

In this research, for every scenario of deployed nodes using OPNET's wireless network deployment wizard, ten different seeds were used to generate ten different deployments. As in [20], the transmission and receive radius for each node was 60 meters. In each node the radio dissipated Eelec=50 nJ/bit to the transmitter or receiver circuitry, the data rate was 20 kbps, and each node had εamp =100 pJ/bit/m. Modeling exposure, the sensing zone was defined to be 120 meters. To measure the consumed energy to transmit 1,536,000 bits over a specific distance, the power based formulas discussed above were used. As in [13], each node was initialized with 24,624 Joules (J) of energy utilizing two AA batteries as the source. Through repeated experiments, each value obtained represents an average of ten simulation runs, each using a different seed.

V. *Research Findings*

The results show that the proposed approach selected different zones than the zones chosen using a power-based metric. Moreover, the higher average of the zones using the proposed approach is a more accurate indication of the current security state of the WSN. After repeating the experiment by assigning risk values inversely proportional to the first test scenario, the results show that the network with less risk has a smaller index. Also, in both tests the more sparse networks displayed lower values. The sparse network simulated an irregular placement in a floor plan or an obstruction in the line of sight. Contrary, in a denser network the appearance of a sparse network can be indicative of other risk and cost factors that may inhibit the sensors ability to communicate with one another.

A. Approach Results

As discussed above, the IoT Index was weighted by the traffic load. The traffic load is proportional to the communication costs, and thus, the remaining power. Accordingly, the average of the IoT Indexes of the clusterheads selected based on Power and non-Power (IoT Index) were compared to ensure that they actually differed. Following this it was important to show that the non-Power based approach provide more accurate information about the network state from a hierarchical perspective than the Power based approach.

Here the opposite of the risk values from Table 1 are applied to demonstrate that the reduced risk present in the network lowered the value of the IoT Index.

Note: Applying the t-test to the data from the Power and non-Power (IoT Index) based simulations at alpha = 0.05 and 18 degrees of freedom, the p-values indicated that the averages are different

B. Approach Analysis

Apparently, as the networks increase in size, there appears to exist a commonality in energy between the non-

Power and Power based approach. This relationship is derived from the power variable in the IoT Index equation. In the equation, expended power equating to traffic load cost is a factor in the denominator. The number of packets that were sent and received from a node and the traversed distances were used when computing the remaining power. After sending and receiving many packets, the remaining power was depleted based on the power computations outlined above. While the metrics are different, both revealed the same result: they are individually inversely proportional to the percentage of energy expended.

VI. Conclusions and Future Work

A. Conclusions

The goal of this study was an approach that enabled a hierarchical situation awareness application on the Internet-Of-Things to utilize clusterheads as local zones whose average represent the global spatial situation of the network. To realize the goals of this research, the results of the test data showed that the main objective for a deployment of a situation awareness application has been accomplished. This was found by the effect of the approaches.

The results of this research indicated that the objectives of the study have been largely met. The tests support the premise of the main objective that the network factors derived from the optimal hierarchical spatial situation can be used examine the network state for situation awareness by taking into consideration monitoring capability, risk, and cost.

B. Future Work

These same procedures used in a dynamically changing network could be the subject of future WSN research. While a dynamic sample is beyond the limits of this study, it would have been preferred. Mobile nodes would notify their neighbors of changes to their status.

Finally, a potential area of improvement for the Local IoT Indicator could be to increase the interaction ability from 2.50 to a value of 10 or less. As discussed, node placement is a very important element of information that is used to determine the network state. If desirable, it would increase the importance of the monitoring capability while reducing the impact of the other network factors such as risk and cost. Such further research may also prove useful for many different types of situations, especially, the improvement of efficiency in dense networks and the efficacy of an application that shares a sparse deployment in a WSN.

Acknowledgements

Drs. Gendreau and Barrios would like to acknowledge the people in our lives who have assisted in the development of this study including our children who have grown up into fine young men and women during its creation.

References

[1] D. Burg, "The Internet of Things Raises New Security Concerns," PwC, 2013. http://usblogs.pwc.com/emerging-technology/the-internet-of-things-raises-new-security-questions/?utm_source=rss&utm_medium=rss&utm_campaign=The-Internet-of-Things-Raises-New-Securi; Feb. 06, 2014.

[2] USA Government, FDA Safety Communication: Cybersecurity for Medical Devices and Hospital Networks, F.D.A., 2013, USA Government: Silver Springs, MD.

[3] Atzori, L., A. Iera, and G. Morabito, *The Internet of Things: A survey*. Computer Networks, 2010. **54**(15): p. 2787-2805.

[4] F. Anjum, D. Subhadrabandhu, S. Sarkar, and R. Shetty, "On optimal placement of intrusion detection modules in sensor networks," First International Conference on Broadband Networks, BROADNETS'04, IEEE, pp. 690–699, October 2004.

[5] Q. Durresi, V. K. Paruchuri, S. S. Iyengar, & R. Kannan, "Optimized broadcast protocol for sensor networks," IEEE Transactions on Computers, 54(8), pp. 1013-1024, August 2005.

[6] L.M. Feeney, & M. Nilsson, "Investigating the energy- consumption model of a wireless network interface in an ad hoc networking environment," In Proc. IEEE Conference on Computer Communications (Infocom), pp. 1548-1557, April 2001.

[7] A. Forster & A. L. Murphy, "CLIQUE: Role-free clustering with q-learning for wireless sensor networks", 29th IEEE International Conference on Distributed Computing Systems, pp. 441-449, June 2009.

[8] C. Frank & K. Romer, "Solving generic role assignment exactly," Proceedings 20th IEEE International Parallel and Distributed Processing Symposium, p. 161, April 2006.

[9] K. Karl & A. Willig, "Protocols and Architectures for Wireless Sensor Networks," West Sussex, Hoboken, N.J.: Wiley, May 2005.

[10] M. Kochhal, L. Schwiebert, and S. Gupta "Role-based hierarchical self organization for wireless ad hoc sensor networks," In Proceedings of the 2nd ACM international Conference on Wireless Sensor Networks and Applications, pp. 98–107, September 2003.

[11] D. Kalashnikov, Y. Ma, S. Mehrotra, R. Hariharan, and C. Butts "Modeling and Querying Uncertain Spatial Information for Situational Awareness Applications", ACM-GIS'06, pp. 131-138, November 2006.

[12] E. Sabbah, A. Majeed, K. Kang, K. Liu & N. Abu-Ghazaleh, "A application-driven perspective on wireless sensor network security", ACM, pp. 1-8, October 2006.

[13] R. Sbrusch, "Authenticated messaging in wireless sensor networks used for surveillance", Retrieved

from ProQuest Digital Thesis & Dissertations. (AAT 1452686), May 2008.

[14] S. A. Shaikh, H. Chivers, P. Nobles, J. A. Clark, and H. Chen, A deployment value model for intrusion detection sensors, Lecture Notes in Computer Science in 3rd International Conference on Information Security and Assurance, vol. 5576., pp.250–259, June 2009.

[15] P. K. Sree & I. R. Babu, “Towards a cellular automata based network intrusion detection system with power level metric in wireless adhoc networks”, (IDFADNWCA) International Conference on Advanced Computer Theory and Engineering, pp. 1071-1075, December 2008.

[16] T. Srinivasan, J. Seshadri, J. B. Siddharth & A. Chandrasekhar, “A system for power-aware agent-based intrusion detection (SPAID) in Wireless Ad Hoc Networks”, In Networking and Mobile Computing, pp. 153-162, January 2005.

[17] P. Techateerawat & A. Jennings, “Energy efficiency of intrusion detection systems in wireless sensor networks”, International Conference on Web Intelligence and Intelligent Agent Technology - Workshops, pp. 227-230, December 2006.

[18] J. Wu & H. Li., “On calculating connected dominating set for efficient routing in ad hoc wireless networks”, In Proceedings of the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications, pp. 7-14, August 1999.

[19] T. Xiaobin, Z. Young, & X. Hongsheng, “Multi-perspective quantization model for cyberspace security situation awareness”, International Conference on Computational Intelligence and Security, pp. 853-857, December 2007.

[20] J. Xu, Lei Wang, C. Ma & L. Shu, “Impacts of duty-cycle on overlapping Multi-hop clustering in wireless sensor networks”, IEEE ComSoft E-Letters, 1(1), pp. 11-13, May 2012.

Investigating Performance of Extended Epidemic Routing Protocol of DTN under Routing Attack

Harminder Singh Bindra[1], **A L Sangal**[2]

[1] Research Scholar, Department of CSE, DR B R Ambedkar National Institute of Technology
Jalandhar, India

[2] Professor, Principal, DAV Institute of Engineering and Technology, Jalandhar, India

Abstract - *Nodes in the DTN work on the foundation of cooperation in the network. When working in a cooperative manner, these nodes consume some network resources like bandwidth, buffer space etc. Like any other networks, DTNs are also prone to the malicious nodes and different attacks. In this work, we have proposed an attack model comprising of falsification of extended routing protocol metadata information combined with drop all attack. We have proposed the attack model definition and analyzed the performance of extended Epidemic routing protocol of DTN under this attack model. From the simulation results, we analyzed that the delivery probability of extended Epidemic routing protocols is greatly affected by the proposed attack model whereas the DTN routing protocols are proved to be robust against the individual attacks when implemented independently of each other.*

Keywords: Delay Tolerant Networks, Epidemic, Spray & Wait, Prophet, Delivery probability, Average latency, Overhead ratio.

1 Introduction

Traditionally, data networks are sculpted by linking graphs whereby the continuation of at least one end-to-end passage among any source-destination duet is endlessly certain. In these networks, any random bond between two network nodes is thought to be bidirectional sustaining symmetric data rates with slight error chances and latency (i.e. Round-trip time is in the order of milliseconds). In these networks, packets are not thought to survive in a node's buffer for a prolonged time period. On the foundation of these fundamental suppositions, the Internet was planned and its most universally used protocols, predominantly the TCP/IP protocol suite, were planned.

On the other hand, these suppositions do not clutch when scheming existing and newly budding wireless networks, particularly those which are to be deployed in acute environments (e.g. Battlefields, volcanic regions, deep oceans, deep space, developing regions, etc.). Under such demanding environment, these networks suffer from extensive delays, acute bandwidth limitations, widespread mobility of nodes, recurrent power outages and frequent communication hindrance. Wireless networks operational under these demanding conditions experiences connectivity which becomes noticeably discontinuous and no uninterrupted end-to-end path(s) between any source-destination pair can be assured [1].

Popular examples of such irregularly connected networks (ICNs) scenarios are satellites, deep space probes, Wireless Sensor Networks (WSNs), Mobile Wireless Sensor Net- works (MWSNs) and Sensor/Actuator Networks (SANs) deployed in acute regions, Mobile Ad-Hoc Networks (MANETs) in general consisting of nodes (e.g. GPSs, PDAs, Cellular Phones, Tracking devices, Laptops, etc.) mounted over endlessly moving objects [1].

Numerous study interests spotlight on developing new approaches for routing in delay tolerant network atmosphere. These routing schemes in general use the store-carry-and- forward approach, where intermediate nodes keep the message until encounter other nodes to set up new links in the path to the destination [2]. DTN Routing protocols can be generally categorized on two bases: (1) on the basis of the number of copies and (2) on the basis of knowledge of future contact opportunities and message patterns. On the basis of the number of copies, we have Single-copy routing schemes which use only one copy per message and significantly reduce the resource requirements but suffer from long delays and low delivery ratios. Other one is Multi-copy routing schemes has a high probability of delivery and lower delays at the cost of buffer space and more message transfers.

This work is related to the security issues of the extended routing protocols [20] and studies the robustness of these protocols against the attack model proposed in this work.

Section 2 summarizes prior related work on routing in disrupted environment and the attacks on the delay tolerant networks.

Section 3 details the system model. In this section, the details about the security assumptions, routing model and simulation settings are discussed.

Section 4 detail the attack model proposed in this paper. This section emphasis on the step by step description of the attack model for this work.

In Section 5, results obtained from the simulation study are discussed. In this, the results are discussed for the three metrics i.e. delivery probability, overhead ratio and average latency under the varying buffer size of the nodes.

Section 6 concludes the study and lists the results obtained in this study.

2 RELATED WORK

Basic DTN routing algorithms rely only on node movement, and no other information is used for the establishment of the communication link. Examples of primary DTN routing are “Custody Transfer” and “Epidemic Routing”.

In order to get better performance of DTN routing, numerous mechanisms have been implemented in diverse DTN routing protocols [1], [2], [3], [4], [5], [6]. These mechanisms often take account of duplication of packets to several nodes so as to raise the probability of delivery and to lessen the delivery latency. In a sole contact, only restricted packets may be exchanged among two portable nodes. As an effect, the orders of packet transmit, which depends on the precedence a node acquaintances with every packet, has momentous impact on the general performance. Replication- based DTN routing protocols vary principally on how each packet’s precedence is determined.

Software of DTN study projects uses an arbitrary algorithm to reproduce node movement while mobility in actual existence has a knowable pattern. Certain DTN routing algorithms are designed to exploit this expected action of node mobility for predicting message delivery in a probabilistic approach [7].

There are numerals of additional proactive approaches to routing which are made achievable by stronger assumptions such as awareness of connectivity model and be in command of peer movement [8-14].

By and large the routing protocols of DTN deal with their buffers as first-in-first-out (FIFO) queues [15]. A further approach is Drop Least Encountered (DLE) algorithm [16] which proposed dropping messages with the lowest likelihood of delivering and various work deployed this dropping technique [7,10,12,17].

The problem associated to the existence of copy of the previously delivered communication in multi copy routing design was studied by Bindra et. al. in [18]. It was assessed that if the copies are removed at the same instance when the one of the data bundle is conveyed then there is the possibility for step up in the performance of the routing protocols.

Further Bindra et. al. in [19] proposed a message deletion policy for multi-copy routing scheme and analyzed the buffer occupancy of the nodes under this extended routing protocol (with proposed message deletion policy). Simulation results show that the extended routing protocol proposed in this work greatly relaxes the buffers of the nodes enabling them to handle more and more messages, which in turn improve the efficiency of the routing protocol. It also helps in preventing the nodes from buffer overflow problem and relaxes the resource utilization of the nodes.

In year 2013, Bindra et. al. in [20] studied the performance of different routing protocol with the proposed buffer management scheme. This scheme helped in preventing the nodes from excessive utilization of resources. It was analyzed that the extended routing protocols (with this new buffer management scheme) performed with improved delivery probability values with reduced overhead ratio and lower average latency value.

Performance of DTN routing protocols not only depend on the factors considered above but also depend on the attacks by the malicious nodes. There are numerous studies on securing the routing protocols of MANETs which focuses on securing the path establishment process [21], [22], [23], [24], [25], [26]. But these schemes cannot be used in securing the DTN as in DTN there is intermittent connectivity and no end-to-end path exist for all source-destination pair at all the time.

3 SYSTEM MODEL

In this section, we describe the system model of the network used for the analysis. This section also explains the security assumptions, mobility model, scenario, interface, node group, message creation specific settings. We evaluated the robustness of Extended routing protocols of DTN in the presence of attacking node. All the evaluations were performed using our simulator modified from ONE simulator [28], a simulator developed specifically for the DTN simulations.

1.1 Security Assumption

In this work, we have assumed that the relay nodes do not perform any authentication on the authenticity of the packets. Due to this non availability of this authentication service, the malicious nodes can add the fake metadata into the network i.e. the malicious nodes can add false delivery information of the packets into the network.

Another assumption made for this study is the lack of global knowledge of topology of the network to the nodes of the network. If this information is available in the network, the malicious nodes can perform much more damage to the routing performance. We have shown that, even in the absence of this information, yet our proposed attack model degrades the performance of the routing protocols to a great extent.

1.2 Routing Model

The routing protocol used in our evaluation is the extended version of Epidemic Routing Protocol [20]. The considered protocol is a replication-based DTN routing protocol. MaxProp which is also a replication based protocol has been shown to provide robustness against various attacks [17]. It offers better throughput than several other strategies such as Epidemic [1], Prophet [3] and Spray and Wait [2]. The overall routing model implemented is shown in figure 1 and 2.

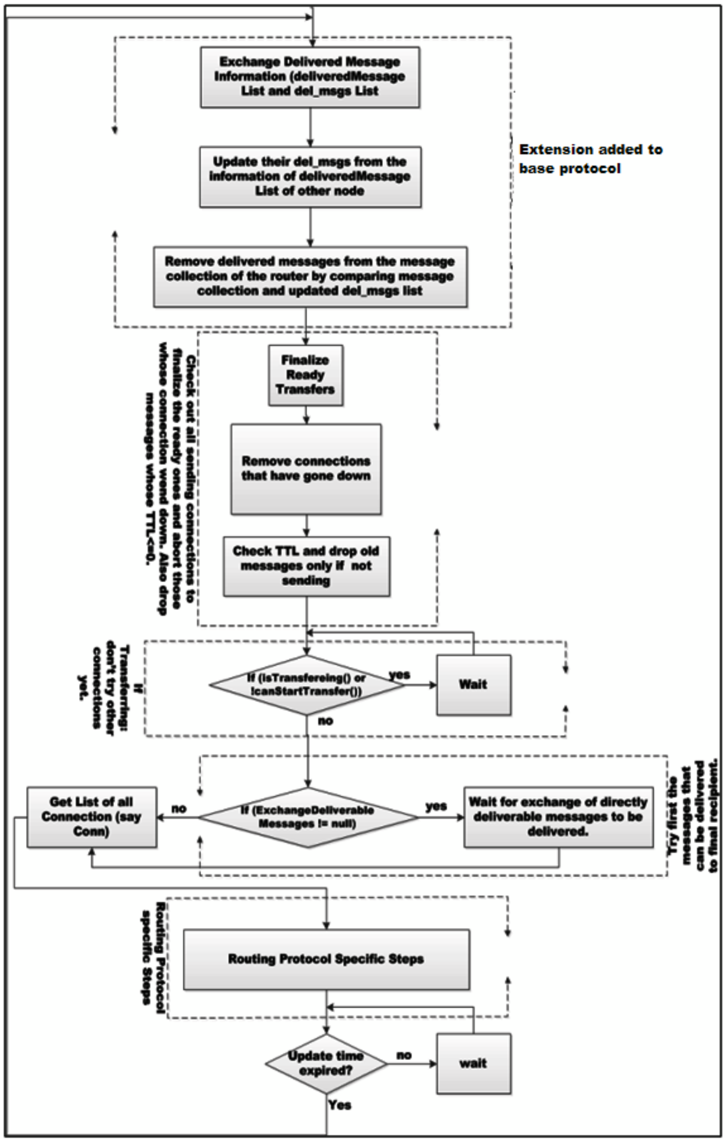

Exchange Delivered Message Information (deliveredMessage List and del_msgs List
Update their del_msgs from the information of deliveredMessage List of other node
Extension added to base protocol
Remove delivered messages from the message collection of the router by comparing message collection and updated del_msgs list
Check out all sending connections to finalize the ready ones and abort those whose connection wend down. Also drop messages whose TTL<=0.
Finalize Ready Transfers
Remove connections that have gone down
Check TTL and drop old messages only if not sending
If Transferring! don't try other connections yet.
If (IsTransfereing() or !canStartTransfer())
yes
Wait
no
Get List of all Connection (say Conn)
no
If (ExchangeDeliverable Messages != null)
yes
Wait for exchange of directly deliverable messages to be delivered.
Try first the messages that can be delivered to final recipient.
Routing Protocol specific Steps
Routing Protocol Specific Steps
Update time expired?
no
wait
Yes

Fig 1: Flowchart for the base Routing Protocols of DTN

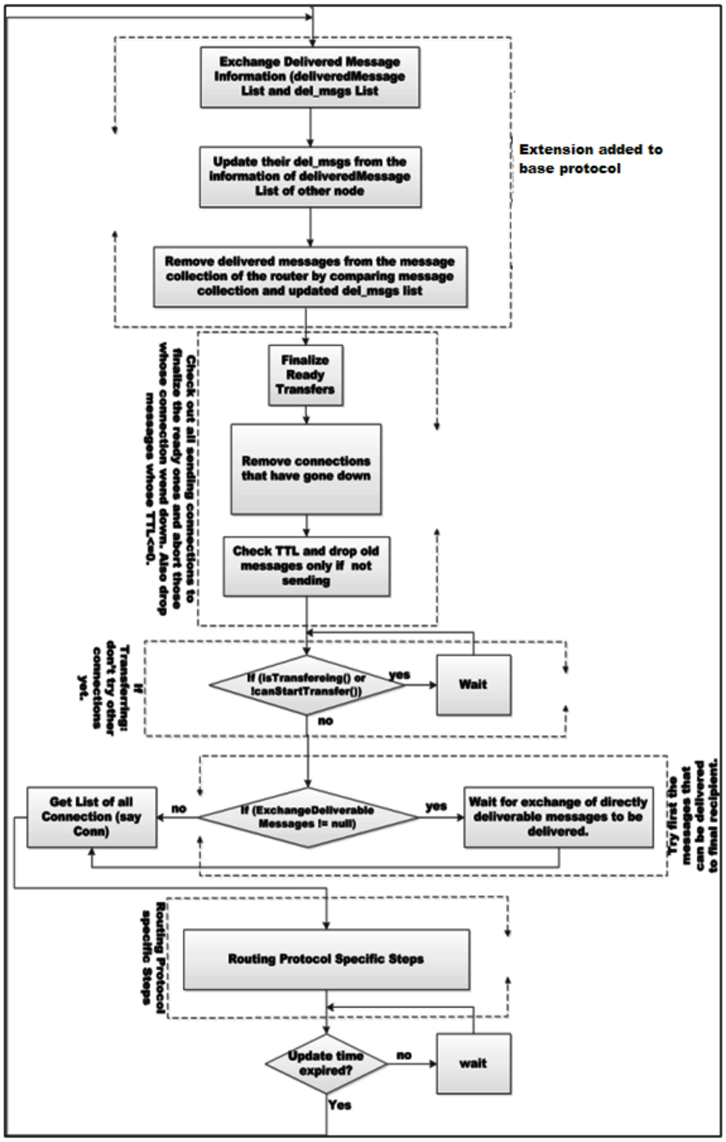

Exchange Delivered Message Information (deliveredMessage List and del_msgs List
Update their del_msgs from the information of deliveredMessage List of other node
Extension added to base protocol
Remove delivered messages from the message collection of the router by comparing message collection and updated del_msgs list
Check out all sending connections to finalize the ready ones and abort those whose connection wend down. Also drop messages whose TTL<=0.
Finalize Ready Transfers
Remove connections that have gone down
Check TTL and drop old messages only if not sending
If Transferring! don't try other connections yet.
If (IsTransfereing() or !canStartTransfer())
yes
Wait
no
Get List of all Connection (say Conn)
no
If (ExchangeDeliverable Messages != null)
yes
Wait for exchange of directly deliverable messages to be delivered.
Try first the messages that can be delivered to final recipient.
Routing Protocol specific Steps
Routing Protocol Specific Steps
Update time expired?
no
wait
Yes

Fig 2: Flowchart for the Extended Routing Protocol of DTN

4 ATTACK MODEL

In [27], four general attacks Drop All, Random flooding, Invert routing metadata, and Acknowledgement counterfeiting were experimentally shown to be ineffective.

Although the above attacks may be ineffective, many variant of these attacks are still possible. In addition, these attacks can be pooled to support each other.

Our proposed attack consists of falsification of extended routing protocol metadata information combined with drop all attack. In our extended DTN routing protocol, the network wide message delivery information is propagated to remove the existing replicas of the delivered information [20]. But if the malicious nodes are present in the network, they can inject the false delivery information in the network about the packets present in the buffer of this node and drop all the packets which are present in the buffer. The detailed attack model is represented below.

The figure 3(a) shows the behavior of normal node. In normal nodes, when two nodes come in the communication range of each other, they populate and exchange the del_msgs lists. Further details about the del_msgs lists and normal behavior of nodes are given in work by Bindra et al. [20].

But if one of the connected nodes or both are malicious, then the behavior of malicious node is depicted in figure 3(b). When these nodes get connected, first of all it reads all the messages from its collection and adds their ids in the del_msgs list and drop all the messages.

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

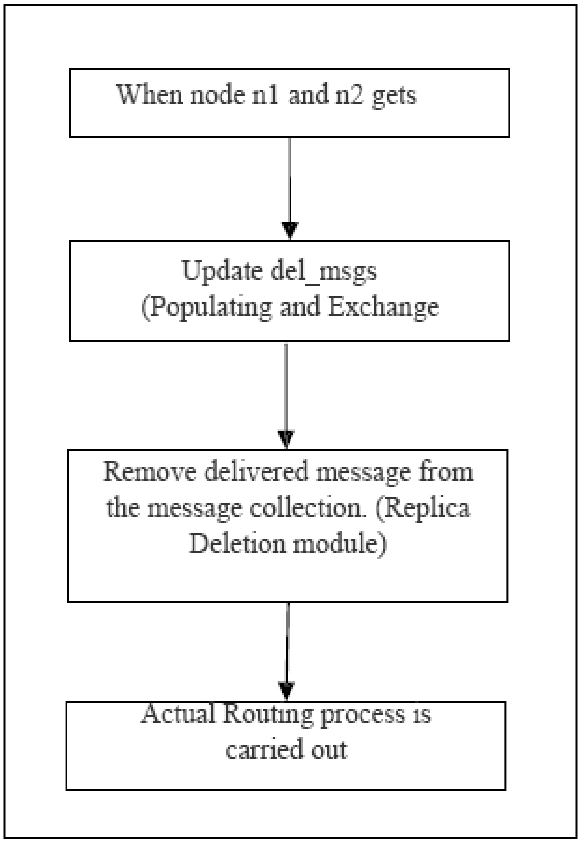

When node n1 and n2 gets
Update del_msgs (Populating and Exchange
Remove delivered message from the message collection. (Replica Deletion module)
Actual Routing process is carried out

Figure 3: a) Behavior of Normal Node

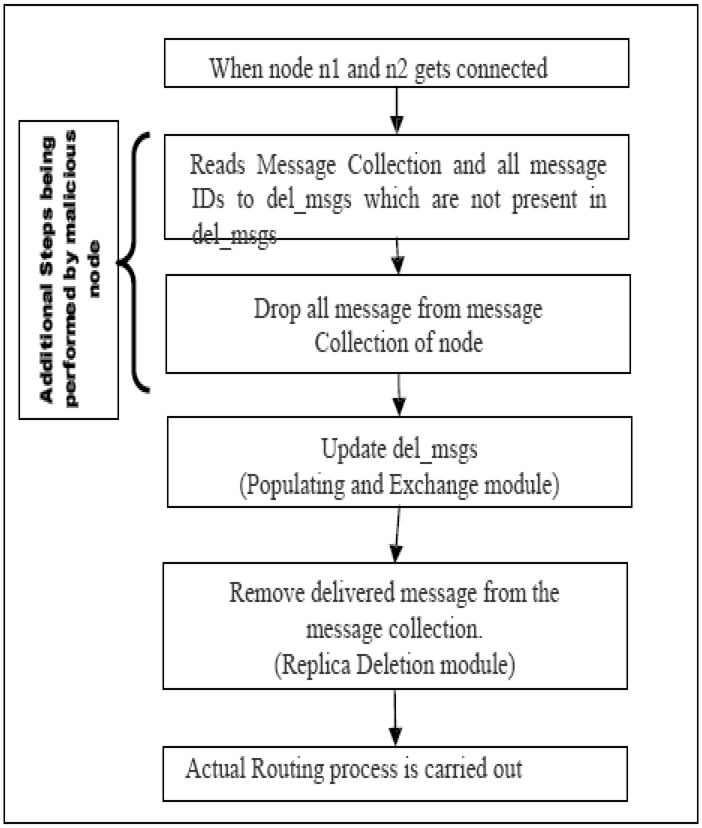

When node n1 and n2 gets connected
Additional Steps being performed by malicious node
Reads Message Collection and all message IDs to del_msgs which are not present in del_msgs
Drop all message from message Collection of node
Update del_msgs (Populating and Exchange module)
Remove delivered message from the message collection. (Replica Deletion module)
Actual Routing process is carried out

Figure 3: b) Behavior of Malicious Node

5 RESULTS AND DISCUSSION

In this section, we discuss the results of the simulations of the system model and attack model presented in section 3 and section 4. To study the attack model, simulations are carried out in our simulator that was modified from ONE simulator [28], simulator designed for DTN simulations. The detailed simulation setup is presented in Table 1.

Table 1: Simulation Configuration

Scenario Setting			
Name	simulateConnection	updateInterval	endTime
Default_scenario	True	0.1 s	43200 s

Interface Specific Setting			
Name	Type	Transmit Speed	Transmit Range
btInterface	SimpleBroadcastInterface	250k	30 m

Node Group Specific Settings								
Movement Model	Router	Buffer size	Wait Time	No of Interfaces	interface1	Speed	Msg Ttl	No of Hosts
Shortest Path Movement	Extended Version of [Epidemic]	5 - 35M	300, 900 s	1	btInterface	0.5, 1.5 m/s	120 mins	40

Message Creation Parameters

<table>
<tr><td>Events.nr of</td><td>Events1.class</td><td>Events1.interval</td><td>Events1.size</td><td>Events1.hosts</td><td>Events1.prefix</td></tr>
<tr><td>1</td><td>Message Event Generator</td><td>15,30 s</td><td>250k, 2M</td><td>0, 39</td><td>M</td></tr>
<tr><td colspan="6">Movement Model Settings</td></tr>
<tr><td colspan="2">MovementModel.rngSeed</td><td colspan="2">MovementModel.worldSize</td><td colspan="2">MovementModel.warmup</td></tr>
<tr><td colspan="2">1</td><td colspan="2">4500, 3400 m</td><td colspan="2">1000 s</td></tr>
<tr><td>Style</td><td>Bold and Italic : Category Heading</td><td>Bold: Attribute</td><td colspan="2">Italic : Attribute value</td><td>Attribute which is varied</td></tr>
</table>

A. The Impact on Delivery Probability

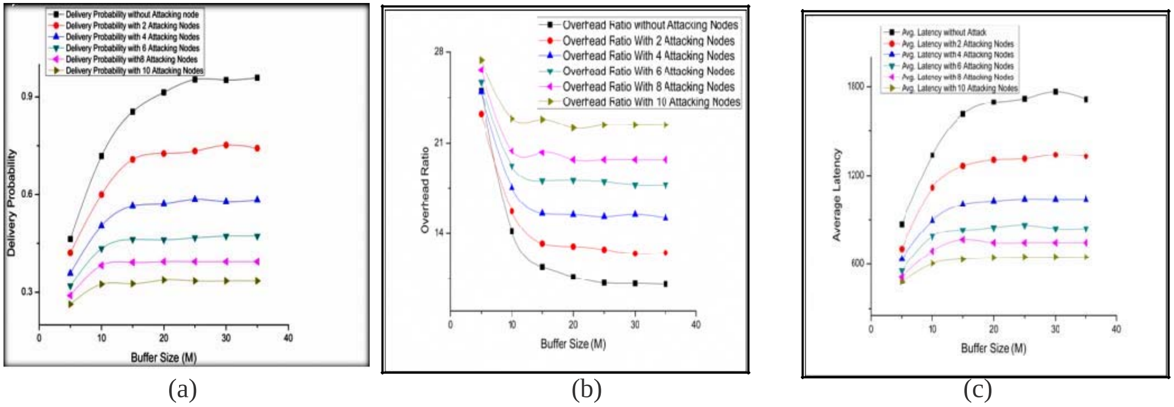
From the fig 4 (a), we analyze that the presence of the attacker node reduces delivery probability of extended protocols. In case of epidemic routing protocol, there is 23% decrease in delivery probability when only 4% of attacker nodes are present. If we increase % of attacking nodes, there is larger decrease in delivery probability.

B. The Impact on Overhead Ratio

From fig 4(b), it is clear that there is a significant increase in the Over Head Ratio when proposed attack is implemented. When 20% attacking nodes are present, there is 115% increase in overhead ratio for epidemic routing

C. The Impact on Average Latency

Results obtained from the simulative study show interesting results for the average latency. Figure 4(c) show that there is an improvement in Average Latency experienced by delivered message. This is because of fact that now buffer is further relaxed as more packets are being deleted due to proposed attack model. Thus remaining packets have to wait for less amount of time in buffer queue. It is observed that average latency value improves by 65% for epidemic routing when20% attacking nodes are added.



(a) (b) (c)

Figure 4: (a) Delivery Probability under attack model (b) Overhead Ratio under Attack model (c) Average Latency under attack model

6 CONCLUSION

Routing metadata that are employed in DTN routing to improve resource utilization can be exploited by attackers to improve the effectiveness of attacks. We have presented an attacks model - comprising of falsification of extended routing protocol metadata information combined with drop all attack - that demonstrates how attackers can exploit routing metadata to improve the effectiveness of attacks. Earlier works from the literature say that the DTN routing protocols are robust to the routing attacks. But the attack model proposed above which is a combination of two attacks is effective enough to degrade the performance of the extended routing protocols of DTN. The simulation results show that the addition of attacker nodes in the network decreases the delivery probability, increases the overhead ratio and decreases the average latency. From the results, it can be analyzed that the effectiveness of the attack increases when the combination of the attacks is employed in collaboration. So from these results we can conclude that there is the need of the authentication service in the routing protocols so that these attacks can be prevented.

In the future, we will try to provide the preventive measures or the authentication service to prevent the attacks and to preserve the performance of these extended routing protocols even in the presence of the attacking nodes.

REFERENCES

[1] A. Vahdat and D. Becker, “Epidemic routing for partially connected ad hoc networks,” Department of Computer Science, Duke University, Durham, NC, Tech. Rep., 2000.

[2] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, “Spray and wait: an efficient routing scheme for intermittently connected mobile networks,” in WDTN ’05: Proceeding of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking. New York, NY, USA: ACM Press, 2005, pp.252–259.

[3] A. Lindgren, A. Doria, and O. Schele´n, “Probabilistic routing in intermittently connected

networks," Lecture Notes in Computer Science, vol.3126, pp. 239–254, January 2004.

[4] B. Burns, O. Brock, and B. Levine, "Mv routing and capacity building in disruption tolerant networks," vol. 1, 2005, pp. 398–408 vol. 1.

[5] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networking," in Proceedings of IEEE Infocom 2006, Barcelona, Spain, April 2006.[Online]. Available:http://prisms.cs.umass.edu/brian/pubs/burgess.infocom2006.

[6] A. Balasubramanian, B. N. Levine, and A. Venkataramani, "DTN Routing as a Resource Allocation Problem," in Proc. ACM Sigcomm, Kyoto, Japan, August 2007. [Online].

[7] Lindgren, A., Doria, A., & Schel, O. (2003). Probabilistic routing in intermittently connected networks. SIGMOBILE Mobile Computing and Communications Review, 7 (3), 19–20.

[8] Jain, S., Fall, K., & Patra, R. (2004). Routing in a delay tolerant network. SIGMOBILE Computing and Communications Review, 34(4), 145–158.

[9] . Wenrui, Z., & Ammar, M. H. Message ferrying: Proactive routing in highly-partitioned wireless ad hoc networks. pp.308–314.

[10]. Sarafijanovic-Djukic, N., Grossglauser, M., & Mitrou, N., et al. (2004). Last encounter routing under random waypoint mobility NETWORKING 2004. Networking technologies, services, and protocols; performance of computer and communication networks; mobile and wireless communications, Lecture Notes in Computer Science, pp. 974–988: Berlin, Heidelberg: Springer.

[11] Zhao, W., Ammar, M., & Zegura, E. (2004). A message ferrying approach for data delivery in sparse mobile ad hoc networks. In Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing (pp.187–198). Tokyo, Japan: Roppongi Hills.

[12]. Burns, B., Brock, O., & Levine, B. N. (2005, March). MV routing and capacity building in disruption tolerant networks. In INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE (Vol. 1, pp. 398–408). IEEE.

[13] Yang, J., Chen, Y., Ammar, M., & Lee, C. (2005, March). Ferry replacement protocols in sparse manet message ferrying systems. In Wireless communications and networking conference, 2005 IEEE(Vol. 4, pp. 2038–2044). IEEE.

[14] Zhao, W., Ammar, M., & Zegura, E. (2005, March). Controlling the mobility of multiple date transport ferries in a delay-tolerant network. In INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE (Vol. 2, pp. 1407–1418). IEEE.

[15] . Vahdat, A., & Becker, D. (2000). Epidemic routing for partially connected ad hoc networks.

[16] James, A. D. Wearable computers as packet transport mechanisms in highly-partitioned ad-hoc networks. pp. 141–141.

[17] . Grossglauser, M., & Vetterli, M. Locating nodes with EASE: Last encounter routing in ad hoc networks through mobility diffusion. vol. 3. pp. 1954–1964.

[18] Harminder Singh Bindra, A L Sangal, Need of Removing Delivered Message Replica from Delay Tolerant Network - A Problem Definition I. J. Computer Network and Information Security, Vol. 4, No. 12, November 2012, 59-64

[19] Harminder Singh Bindra, A. L. Sangal, Analyzing Buffer Occupancy of the Nodes under Acknowledged Delay Tolerant Network's Routing Protocols, Communications in Computer and Information Science, Springer Berlin Heidelberg, 2013, 537-544

[20] Harminder Singh Bindra, A. L. Sangal, Extension of Delay Tolerant Network's Routing Protocols for Preventing Excessive Utilization of Resources, Wireless Pers Communication, Springer Science Business Media New York 2013.

[21] D. Djenouri and N. Badache, "Struggling against selfishness and black hole attacks in manets," Wirel. Commun. Mob. Comput., vol. 8, no. 6, pp. 689–704, 2008.

[22] D. Hongmei, L. Wei, and A. Dharma P., "Routing security in wireless ad hoc networks," IEEE Communications magazine, October 2002.

[23] Y. C. Hu, A. Perrig, and D. B. Johnson, "Packet leashes:a defense against wormhole attacks in wireless networks," vol. 3, 2003, pp. 1976–1986 vol.3.

[24] Cristina and H. Rubens, "An on-demand secure routing protocol resilient to Byzantine failures," in ACM Workshop on Wireless Security (WiSe), Atlanta, Georgia, September 2002.

[25] S. Yi, P. Naldurg, and R. Kravets, "A security-aware routing protocol for wireless ad hoc networks," in in: Proceedings of ACM Symposium on Mobile Ad Hoc Networking and Computing (Mobihoc), 2001, pp. 286–292.

[26] Y. chun Hu, "Rushing attacks and defense in wireless ad hoc network routing protocols," in in ACM Workshop on Wireless Security (WiSe), 2003, pp. 30–40.

[27]J. Burgess, G. D. Bissias, M. D. Corner, and B. N. Levine, "Surviving attacks on disruption-tolerant networks without authentication," in MobiHoc '07: Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing. Montreal, Quebec, Canada: ACM Press, 2007, pp. 61–70.

[28] A. Kernen, J. Ott, and T. Kinen, "The ONE simulator for DTN protocol evaluation," in Proceedings of the 2nd International Conference on Simulation Tools and Techniques, Rome, Italy, 2009, pp. 1-10.

TinyWatermark: a code obfuscation-based software watermarking framework for wireless sensor networks

Emanuele N. de L. F. Jorge[1], Luci Pirmez[1], Rafael de O. Costa[1,2], Igor L. dos Santos[1], Claudio M. de Farias[1], Davidson R. Boccardo[2], Luiz F. R. da C. Carmo[2], Lucila M. S. Bento[1,2]
[1]Programa de Pós-Graduação em Informática – Instituto Tércio Paccitti / Instituto de Matemática – Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil
[2]Instituto Nacional de Metrologia, Normalização e Qualidade Industrial - Av. N. S. das Graças, 50 – 25.250-020 - Xerém - Duque de Caxias – Rio de Janeiro, Brasil

Abstract - *One of the main challenges in the field of Wireless Sensor Networks (WSN) is related to the security of their nodes. This is because such nodes, called sensors, are arranged in unprotected areas where they are vulnerable to capture, reverse engineer and tamper with by malicious people (attackers). Once such sensors are captured, an attacker can distribute the software code or even use the code in his projects without being traceable. In this context we propose a "code obfuscation-based software watermarking framework" for protecting the authorship of the software embedded in the sensors, and thereby discouraging the theft of intellectual property. The proposed software watermarking framework is defined as a sequence of code obfuscations, which makes, simultaneously, the code more difficult to analyze and the watermark more difficult to be located and removed, because the watermark is interleaved with other software instructions. Experiments were performed for measuring the overhead imposed on the software after the watermark insertion and finally we present analysis related to the credibility and stealth of the proposed framework.*

Keywords: Software Watermarking, Intellectual Property Protection, Software Protection, Obfuscation

1 Introduction

Recent advances in microelectromechanical systems technologies and wireless communications have enabled the construction of devices, called sensors, endowed with processing and communication capabilities, used to monitor physical quantities in an environment. Due to their small size and low production cost, tens, hundreds or thousands of such sensor nodes, powered by batteries, can be connected together to create a Wireless Sensor Network (WSN) [1]. One of the main challenges in the field of WSN is related to the security of their sensor nodes, since such nodes are vulnerable to Man-At-The-End (MATE) attacks [2]. In this type of attack, an attacker can capture sensor nodes in a WSN, which are usually deployed in unprotected areas, in order to gain advantage through violating the software or hardware of such nodes. For instance, an attacker in possession of one sensor can dump its internal memory, disassembly it and distributed it or even he could extract its proprietary modules to use in other projects, without being traceable. Hence, it is important to provide means for protecting the intellectual property of the software embedded on the nodes of a WSN.

According to Collberg and Nagra [3], software watermarking is a solution for protecting the authorship of software, discouraging the theft of intellectual property. Such watermark is characterized as a unique piece of information that should be inserted into a software code, in order to recognize who is the author of this software, who holds its rights [4]. One of the most common strategies for software watermarking involves the inclusion of the name of the author at a specific point in the program [3]. Such strategy is advantageous because it does not significantly increase the size of the software, neither its execution time. However, this strategy is not widely recommended because an attacker can easily tamper with the watermarking by changing the name of the author or by simply removing it. Therefore, a software watermark must be stealthy, i.e. difficult to be located within the code, avoiding its tampering with. A solution for making the watermark stealthy is combining the watermarking strategies with code obfuscation techniques [5, 8, 9]. Code obfuscation can be characterized as a set of code transformations that makes the code less intelligible, preserving, however, its original features, i.e. not modifying the software semantics.

Another challenge in the field of WSN is related to the resource consumption of the sensor nodes, which is divided in (i) memory limitation, requiring that the embedded software does not exceed a certain size and (ii) energy consumption, since sensor nodes are powered by non-rechargeable batteries and the elaboration of a replacement policy for such batteries is not common, since these sensor nodes are usually arranged in places where the access is difficult or limited. Therefore, the watermark inserted in the software embedded in these sensors should not overwhelm their resources, in terms of both memory and energy.

The objective of this work is to propose a framework for insertion and detection of a code obfuscation-based watermark, in order to prove the authorship of the software embedded in sensor nodes of a WSN. The proposed watermark is defined as a sequence of code obfuscations applied over the software, where the order of each code obfuscation action is responsible for identifying the author of such software. Thus, the proposed watermark is (i) stealthy, because it is difficult to localize the points of the program which were obfuscated, which compose the watermark itself, (ii) makes the code more difficult to be analyzed since the applied code obfuscations make the software less intelligible, and (iii) generates little overhead in terms of resource consumption, because the applied code obfuscations are based on simple instruction replacements.

Experiments were performed for verifying (i) the credibility of the watermark, i.e., if the proposed watermark is able to identify the author of the software, (ii) stealthy, i.e., if the watermark is not easily disguinshed from the program instructions, and (iii) the overhead in terms of resource consumption of the sensor nodes after inserting the watermark.

The remainder of this paper is organized as follows: Section 2 presents the related works, Section 3 presents the framework for watermark insertion and detection, Section 4 presents the experiments performed and, finally, Section 5 presents the conclusions and future work.

2 Related Works

In this section we present some works in the literature that propose watermarking strategies combined with code obfuscation for hindering the watermark location [5,6,9]. Altough, such works present code-obfuscation based software watermarking strategies, as our proposal does, they differ from the present work since they do not use code obfuscation for creating the watermark itself, as the present paper proposes.

Zeng et al. [5] proposed a watermark based on the interpretation of a vector containing the frequencies of each kind of instruction in the software. For inserting the watermark, the program is modified through the replacement or insertion of instructions. In the replacement mode, the code is modified using equivalent instructions recognized by the interpreter. When it is not possible to replace one instruction by an equivalent one, the code insertion is needed, and such insertion is performed using opaque predicates. After inserting the watermark, a new frequency vector is extracted from the program. For watermark recognition and extraction, the interpreter compares this new frequency vector with the original vector. Our proposal differs from the work of Zeng et al. because it does not need a vector with the frequencies of each kind of instruction in the program for handling the watermark.

The idea of Xu [6] is to store the watermark information in the position marked by the difference of height between two jump statements, where the height is defined as the number of instructions between the jump statement and its target. The watermark key is converted to a binary digit sequence, which is stored in a code table. Then, the binary digit sequence goes, as a parameter, through a code obfuscation function, and the obfuscated information is stored in another vector. One of the differences between our proposal and the work of Xu is that the process of watermark insertion in our proposal uses the sequences of obfuscations as the watermark itself, while in the work of Xu, the watermark is obfuscated before being inserted in the software and the software itself is not obsfuscated.

Chen and Chaoquan [9] proposed a code obfuscation-based software watermark technique to protect software developed on the .NET platform. The technique proposed by them, inserts the watermark in the intermediate code generated by the .NET virtual machine. It divides the intermediate code into variable and random sized blocks. Next, those blocks are reordered, while preserving the software semantics, and all the possible combinations are listed, for randomly choosing one. The watermark is inserted as an adittional sequence of bits at the end of each block. After, an unconditional jump instruction is inserted between each block to ensure that the watermarked software behaves properly, as the original one. Finally, a new executable is created. The difference between our proposal and the Chen and Chaoquan proposal is that in our proposal various code obfuscation techniques are used, while the work of Chen and Chaoquan uses only the order of the blocks for code obfuscation.

3 TinyWartemark Framework

In this section a framework, called TinyWatermark, which is able to insert and detect a code obfuscation-based watermark is presented. The objective of this framework is to inhibit intellectual property theft. This section is organized as follows: Section 3.1 presents the logical architecture of this framework; Section 3.2 describes an implementation instance of the components: Key Generator and Obfuscator; Section 3.3 describes the TinyWatermark operation.

3.1 Logical Architecture

The logical architecture of TinyWatermark, shown in Figure 1, comprises 7 components (Intellectual Property Manager, Compiler, Key Generator, Watermarking Embedder, Watermarking Detector, Obfuscator and Obfuscation Recognizer) and two data structures (Obfuscation Rules and Watermark Key). The TinyWatermark components are interconnected through the following interfaces: genKey, embedWM, detectWM, reqObf, recogObf and compile.

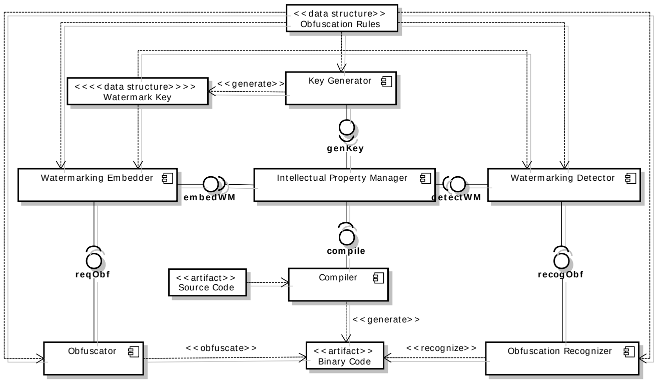


Figure 1. TinyWatermark logical architecture

The **Obfuscation Rules** data structure specifies the obfuscation techniques supported by TinyWatermark and the rules that define how to employ each of these techniques to one or more hardware platforms in certain pieces of the software. In other words, this component specifies the set of instructions that must replace a given instruction located in a particular point of the software, in order to employ a particular obfuscation technique for a specific hardware platform. This data structure must have the following information: (i) the identification of the obfuscation technique, (ii) the hardware platform, (iii) the set of instructions to be obfuscated and (iv) the set of new instructions with the same semantics to replace the original instructions. This structure should be filled by a security expert who is able to customize each obfuscation technique to a given hardware platform.

The **Watermark Key** data structure defines the sequence of obfuscations (based on instruction replacement or insertion) to be applied on a piece of software, which characterizes a certain watermark. Each item of this data structure has the following fields: (i) location of the piece of code where the obfuscation technique must be applied and (ii) the obfuscation technique to be applied at that location.

Intellectual Property Manager is the main component of this framework, and is responsible for (i) booting TinyWatermark, (ii) coordinating the actions of other components, and (iii) receiving the following input information from the user: operation mode, which indicates if the user wants to embed or recognize a watermark in the software; hardware platform, which indicates the hardware platform of the device; software author, which stores the identification of the software author; source code, which is the software to be marked; watermark key, which identifies the key that composes a given watermark, containing the sequence of obfuscations and their locations within of the marked software; binary code, which contains the address of the binary code of the marked software.

The **Intellectual Property Manager** requires the following information from the interfaces of this framework: the Binary Code generated by the Compiler component via the compile interface, the watermark key generated by the Key Generator component through the genkey interface, the marked Binary Code provided by the Watermarking Embedder component via the embedWM interface, and the Software Author recognized by the Watermarking Detector component through the detectWM interface.

The **Compiler** component is responsible for compiling the Source Code for a given hardware platform. The Compiler must handle the Source Code to insert NOP instructions to open space for instruction insertion. After handling the Source Code for a given hardware platform, the Compiler selects the correct compiler for generating a Binary Code since this component is able to compile a Source Code for different hardware platforms. For providing such functionality, the Compiler provides an interface, called **compile, which** requires the following parameters: Hardware Platform and Source Code.

The **Key Generator** component is in charge of generating the Watermark Key. This key is generated according to the identification of the software (Software Author) and must be able to uniquely detect a watermark, i.e., to uniquely detect the sequence of obfuscations applied in the software program, in order to prove that Software Author is the owner of the software. The Key Generator provides an interface named **genKey**, which takes the parameter Software Author, to generate the Watermark Key, which contains the authorship identification.

The **Watermarking Embedder** component is responsible for inserting a watermark in the Binary Code according to the Watermark Key. The Watermarking Embedder determines which code obfuscation techniques, which are listed in Obfuscation Rules, should be applied in the Binary Code. Next, it calls the Obfuscator component that obfuscates the Binary Code taking into account the code locations specified by the Watermark Key. It is important to mention that these locations are covert to the attacker, then making difficult for him to discover a watermark location by only looking the Binary Code. The Watermarking Embedder component provides an interface called **embedWM**, which requires the following parameters: Binary Code, Hardware Platform and Watermark Key.

The **Obfuscator** component is responsible for performing the obfuscations in the Binary Code for a given Hardware Platform. For performing such obfuscations, this component provides the **reqObf** interface, whose parameters are Binary Code, Hardware Platform, and the obfuscation technique to be applied in the Binary Code, together with the address in the Binary Code where such technique must be applied. For performing an obfuscation technique, the Obfuscator uses the Obfuscation Rules data structure, which specifies the instructions that must be replaced and the instructions that should be inserted in the Binary Code.

The **Watermarking Detector** component is responsible for detecting if a Binary Code has a particular watermark, which is identified by a certain Watermark Key. For performing such detection, the Watermarking Detector calls the Obfuscation Recognizer component. If all obfuscations are recognized, the Watermarking Detector informs who the Software Author of this software is. This functionality is provided by Watermarking Detector through the interface **detectWM**, whose parameters are: Binary Code, Hardware Platform and Watermark Key.

The **Obfuscation Recognizer** component is responsible for recognizing, for a given Hardware Platform, if an obfuscation technique was applied at a given address of the Binary Code. For performing such recognition, this component queries the Obfuscation Rules structure, identifying which instructions should be found in a given address. If such instructions are found, the Obfuscation Recognizer reports that the recognization of the obfuscation technique was successful, and, otherwise, the component reports a failure. This functionality is provided by the Obfuscation Recognizer through the interface named **recogObf**, whose parameters are: Binary Code, Hardware Platform, obfuscation technique and the address of the first instruction to be verified.

3.2 Implementation

In this section we describe the implementation of the Obfuscator (Section 3.2.1) and Key Generator (Section 3.2.2) components of the prototype developed for inserting and detecting the proposed code obfuscation-based watermark in a code compiled for the AVR hardware platform [11].

3.2.1 Obfuscator

For implementing the Obfuscator component we used only obfuscation techniques based on instruction replacements, because such kind of obfuscation does not cause major impact on the computational resources of the device running the obfuscated software, like the software size and processing cycles. This choice was made because the watermark will be inserted into the embedded software of a type of sensor that has memory and energy constraints.

The Obfuscator depends of the information stored on the Obfuscation Rules structure for performing obfuscation techniques in the specified addresses of this software. Therefore, for the hardware platform selected in this work (AVR), it was necessary to fill the Obfuscation Rules structure according to the obfuscation techniques chosen: call obfuscation (CallObf) [7] and return obfuscation (RetObf) [7]). Table 1 shows the Obfuscation Rules structure inclunding the folowing informations: the adopted Hardware Plataform, which instructions must be replaced (Original Instructions) and which instructions will be used for replacement (Obfuscated Instructions).

Obfuscation Technique	Hardware Platform	Original Instruction	Obfuscated Instructions
CallObf	AVR	call	ldi/push/ldi/push/ ldi/push/ldi/push/ret
RetObf	AVR	ret	pop/pop/ijmp

Table 1. Obfuscation Rules

Upon receiving a request through the reqObf interface, the Obfuscator component checks if the address provided via this interface contains the value specified in the Original Instruction column. In positive case, the Obfuscator replaces the instruction found by the instruction specified in the Obfuscated Instructions column. Table 2 shows an example of how the Obfuscator can apply the call and return obfuscations. The left column of this table (Original Code) shows the original code snippet containing the Main and Sum functions. In the Main function, 0xA and 0xB constants are stacked and then a function call to the Sum function is performed, which in turn performs the sum of the constants 0xA and 0xB. The right column of this table (Obfuscated Code) shows how the instructions responsible for the call (*call*) and return (*ret*) of the function were obfuscated, in accordance with the information from Table 1.

First, the call instruction, located in the L5 address is replaced by the instructions located between the M5 and M13 addresses. The instructions ldi, push, ldi push, located between M5 and M8 are responsible for stacking the return address of the function to be called (M14). And the instructions ldi, push, ldi push located between M9 and M12 are responsible for putting on top of stack the address of the function to be called (M17). Finally, the Obfuscator inserts, in address M13, the ret instruction that will be responsible for redirecting the control flow to the address at the top of the stack, i.e. M17 (the address of the Sum function). After performing the instructions of the Sum function, the control flow is then redirected to the return address previously stacked (M14). Then the return instruction (ret), located in L6 address, is replaced by the instructions pop, pop, ijpm. The pop instructions are responsible for storing, in the specified registers (r30 and r31), the return address of the function, which is located at the top of the stack, and the ijmp instruction is responsible for redirecting the control flow to the addresses specified by registers r30 and r31. After applying those obfuscation techniques, the code analysis tools could not create properly the control flow graph (CFG) and call graph (CG). The CFG is used to represent all paths that might be traversed during runtime. On the other hand, the CG represents the relationship betwen the functions of the software. When the call and ret are obfuscated, the code analysis tools could not create the right edges of those graphs (CFG and CG), because they are not capable to identify correctly the location of the call and ret instructions.

Original Code	Obfuscated Code
`Main:`	`Main:`
`L1:  ldi r16, 0xA`	`M1:  ldi r16, 0xA`
`L2:  push r16`	`M2:  push r16`
`L3:  ldi r16, 0xB`	`M3:  ldi r16, 0xB`
`L4:  push r16`	`M4:  push r16`
`L5:` **`call Sum`**	`M5:` **`ldi r30, low(M14)`**
`L6:` **`ret`**	`M6:` **`push r30`**
`Sum:`	`M7:` **`ldi r31, high(M14)`**
`L7:  ldd r16, Y+3`	`M8:` **`push r31`**
`L8:  ldd r17, Y+4`	`M9:` **`ldi r30, low(M17)`**
`L9:  add r17, r16`	`M10:` **`push r30`**
`L10: ret`	`M11:` **`ldi r31, high(M17)`**
	`M12:` **`push r31`**
	`M13:` **`ret`**
	`M14:` **`pop r31`**
	`M15:` **`pop r30`**
	`M16:` **`ijmp`**
	`Sum:`
	`M17: ldd r16, Y+3`
	`M18: ldd r17, Y+4`
	`M19: add r17, r16`
	`M20: ret`

Table 2. Examples of how TinyWatermark can apply the call and return obfuscation techniques

3.2.2 Key Generator

In this section we present the procedure for creating the Watermark Key. This procedure comprises three steps. The first step is responsible for transforming the string that identifies the author of the software (Software Author) into a binary code (called BinStr). For generating the BinStr, the ASCII code of each character from the Software Author string is concatenated. For instance, if the software is identified by a Software Author string of value "UFRJ", then the following ASCII codes should be concatenated: 01010101, 0100 0110, 01010010 and 01001010, which refer to the U, F, R and J characters, respectively. Thus, after this first step, BinStr contains 32 digits (01010101010001100101001001001010). The BinStr indicates the number of obfuscations that should be employed in the software, in order to embed the watermark in this software, whose Software Author is "UFRJ".

The second step consists in selecting which obfuscation techniques must be employed in the software. For performing this second step, the Key Generator must select one of the obfuscation techniques listed in Obfuscation Rules for each binary digit of BinStr. This selection obeys to the following rules: (i) if the digit from BinStr equals to 0, then we randomly select an obfuscation technique in Obfuscation Rules, among the obfuscation techniques whose search index is zero or pair; (ii) if the digit from BinStr equals to 1, then we randomly select an obfuscation technique in Obfuscation Rules, among the obfuscation techniques whose search index is odd. For instance, the first obfuscation technique chosen, according to the BinStr related to the Software Author "UFRJ" and to the Obfuscation Rules data structure presented in Table 1, will be CallObf, since the first digit of this BinStr is equal to 0. Therefore it is necessary to choose the single possible technique whose index is 0 or pair in Obfuscation Rules. This same procedure can be applied to choose the remaining 31 obfuscation techniques for the remaining 31 binary digit of BinStr.

The third step consists on populating the Watermark Key structure with information about the obfuscation techniques and the addresses of the software in which each of these techniques must be applied. For populating this structure, it is first necessary to decide how to represent the information about the obfuscation techniques and addresses. In the prototype developed in this work we chose to represent the Watermark Key through a directed graph, called G = (V, E), where V is the set of vertices of this graph, $V = \{v_1, v_2 \dots v_n\}$, and E represents the set of edges of this graph $E = \{e_1, e_2 \dots e_m\}$. Each vertex v_i stores the code address where an obfuscation technique is applied, and e_i represents the connection between two vertices of this graph, identifying who are the source vertex and destination vertex. Although an edge alone does not express concrete information, the sum of the edges arriving at a given vertex, i.e., the degree of entry of that vertex, is used to represent which obfuscation technique must be applied at the address stored in the vertex. The number of vertices in G must be equal to the number of digits in BinStr, and the number of edges must be equal to the sum of the indices of the obfuscation techniques listed in Obfuscation Rules, which were chosen in the second step. Thus, following the example of the Software Author of value "UFRJ", the Watermark Key should be represented by a graph with 32 vertices, and the degree of entry of the first vertex will be 0, since the obfuscation technique chosen in the second step was CallObf, located at index 0 of the Obfuscation Rules table. If a technique were chosen whose index in Obfuscation Rules were greater than 0, it would be necessary to create edges for ensuring that the degree of entry of this vertex would be equal to the index of the chosen technique. For instance, if the index of the chosen technique to obfuscate the address stored in the vertex v_k were equal to 1, it would be necessary to create an edge whose destination vertex were v_k.

One constraint that should be considered, if the graph representation is used, is that the index that identifies the obfuscation technique chosen from those listed in Obfuscation Rules can not be greater than the number of digits of BinStr, since the total number of vertices of this graph equals to the number of digits of BinStr and it is not possible to guarantee a degree of entry of a vertex that is greater than the total number of vertices of this graph.

It is important to mention that, because the choice of indices of the obfuscation techniques listed in Obfuscation Rules was made randomly, it is possible to create distinct keys for a same given Software Author, so that each of such keys are able to identify the same Software Author.

3.3 TinyWatermark Operation

This section shows the operation of TinyWatermark according to the chosen operation mode, which can be insertion or detection of the watermark. TinyWatermark operation is performed offline, i.e. before the Binary Code is deployed in a device or while the software is not running.

Initially, the Intellectual Property Manager component receives the input information, which varies with the operation mode. If the operation mode is insertion of the watermark, the user must provide the following input information: Hardware Platform, Software Author and Source Code. On the other hand, if the operation mode is detection of the watermark, the user must provide the following input information: Hardware Platform, Watermark Key and Binary Code.

The Intellectual Property Manager, after identifying that the operation mode is insertion of the watermark, requests the Compiler, via the interface compile, the compilation of the Source Code for the Hardware Platform specified, generating the Binary Code. Then, the Intellectual Property Manager requests to the Key Generator component, through the genKey interface, the creation of the Watermark Key, given the Software Author provided as input by the user. After the Watermark Key is created, the Intellectual Property Manager requests to the Watermarking Embedder component, through the embedWM interface, the insertion of the watermark (identified by the Watermark Key created) in the Binary Code. The Watermarking Embedder, in turn, requests, through the reqObf interface, that the Obfuscator component performs the obfuscation techniques in the Binary Code, in accordance with the information specified in Watermark Key. For performing such obfuscation techniques in the Binary Code, the Obfuscator queries the Obfuscation Rules structure for checking the obfuscation rules for the Hardware Platform specified. After the Obfuscator applies the obfuscation techniques in Binary Code, the Watermarking Embedder informs to the Intellectual Property Manager that the watermark was inserted. Finally, the Intellectual Property Manager returns to the user the Binary Code with the watermark inserted, ready for being deployed in a sensor node. The Intellectual Property Manager also returns the Watermark Key, which can be used later for proving the authorship of this software.

If the operation mode is detection of the watermark, the Intellectual Property Manager asks the Watermarking Detector, via the detectWM interface, to check whether the watermark, specified by Watermark Key, can be recognized in Binary Code. Then the Watermarking Detector calls, through the recogObf interface, the Obfuscation Recognizer, for recognizing whether a particular obfuscation technique was applied in a specified address from the Binary Code. For performing this task, the Obfuscation Recognizer queries the Obfuscation Rules structure, obtaining the instructions, which must be found for the Hardware Platform specified. If it is possible to identify all the obfuscations performed, the

Watermarking Detector component informs the Intellectual Property Manager component, who is the respective Software Author. Otherwise, the Watermarking Detector informs that it was not possible to prove the authorship of the software.

4 TinyWatermark Evaluation

This section describes the metrics and scenarios used during the experiments performed for evaluating TinyWatermark and theirs results.

4.1 Metrics and Scenario

The **overhead** is measured in terms of (i) the diference between the number of processing cycles required by the software with the watermark and the same software without watermark, called simply Difference of Processing Cycles (DPC) and (ii) the diference between the amount of memory consumption (the memory size in bytes) required by the software with the watermark and the same software without the watermark, called simply Difference of Memory Consumption (DMC).

The **credibility** indicates the capacity of TinyWatermark of detecting the watermark in the software. This metric is assessed in terms of the rates of **true positives** (TP), **false positives** (FP), **true negatives** (TN) and **false negatives** (FN). A **TP** means that TinyWatermark is able to detect that the given software contains a watermark embedded in its code, when in fact the watermark is embedded. A **FP** occurs when a watermark is detected in the given software, but in fact no watermark is embedded in the software. A **TN** situation occurs when TinyWatermark is not able to detect the watermark, which is really not embedded. A **FN** happens when TinyWatermark says that the given software has no watermark, but in fact the watermark was inserted in the software. It is important that TinyWatermark provides a low FP+FN rate, i.e. a low probability of **false detection**.

In the watermarking context, **stealth** means the difficulty to identify the code that composes the watermark. This is useful to resist against attacks intended to modify or remove the watermark because such attacks needs to identify the location of the watermark code before perform any action. Thus, to making the watermark code stealthier, it should look more like original code and vice versa.

Concerning the scenarios, the experiments were performed using the AvroraZ [12] simulator, for simulating a WSN. This WSN was based on MICAz motes. The MICAz hardware platform uses 8-bit AVR (ATmega128) microcontrollers, which are manufactured by Atmel. The disassembler chosen to evaluate the proposed framework was IDA PRO, a commercial disassembler tool, which is based on a recursive transversal algorithm.

Every software used for testing the watermark insertion and detection is available and distributed together with the development environment of TinyOS [13], version 2.1.2, which uses the nesC language [13]. The software programs chosen to be used and installed in the experiments were Blink, Sense, RadioCountToLeds, RadioSenseToLeds, and BaseStation. The experiments were repeated 30 times for achieving a reliable 95% confidence interval for the results.

4.2 Experimental results

In this section, the results of the experiments for evaluating TinyWatermark will be described.

4.2.1 Watermark embedding cost

For each software program (Blink, Sense, RadioCountToLeds, RadioSenseToLeds and Base Station), experiments for evaluating the overhead in terms of the amount of additional computational resources of the software with watermark in relation to original software was performed. Table 3 shows the additional memory comsumption (percentage) and the extra processing cost (percentage) of software with watermark in relation to the original software (software without watermark). In the case of software with watermark, some instructions can be changed or inserted in the original program, and this cost has to be evaluated.

Software	DMC			DPC		
	Orig.	Water.	%	Orig.	Waterm.	%
Blink	7.9	8.1	2.2	295	295	0
Sense	9.4	9.5	1.7	295	295	0
Radio CountTo Leds	39.0	39.2	0.4	295	295	0
Radio SenseTo Leds	40.2	40.4	0.3	295	295	0
BaseStation	45.6	45.7	0.3	295	295	0

Table 3. Effects of the proposed watermark on overhead

For each software program we have inserted a random 3-byte Software Author key. For all software programs in Table 3, we can observe that the respective watermarked program consumed more memory (measured in bytes) than the original one. For the Blink, Sense, RadioCountToLeds, RadioSenseToLeds and Base Station software programs, the respective software programs watermarked by TinyWatermark consumed 2.2%, 1.7%, 0.4%, 0.3% and 0.3% more memory than the original software.

It is important to observe that each software program with the watermark consumed the same amout of processing cycles of the respective original software. The changes performed by TinyWatermark for inserting the watermark in each software program did not reflect negatively in their execution times, in relation to the respective original programs. This fact is explained because only a few instructions were added to each program for inserting the code obfuscation-based watermark in it and, therefore, a negligible more processing cycles are needed by the software for performing the same task. We therefore conclude that the overhead (in terms of memory size amount of processing cycles) that TinyWatermark imposes in the original software is irrelevant.

4.2.2 Credibility Experiments

The credibility experiment aims to evaluate the ability of TinyWatermark to correctly detect a watermark in a given software program. In this section we describe the two experiments performed for assessing credibility. In the both experiments, we evaluated if TinyWatermark is able to detect the watermark of each program. For this purpose, for each program we randomly chose a 3-byte key for inserting in Blink, Sense, RadioCountToLeds, RadioSenseToLeds and

BaseStation applications. Then, we set the operation mode in TinyWatermark as detection, in order to verify whether our proposal is capable of recognizing the previously inserted random 3-byte key as a valid watermark or not. We repeated this procedure 30 times for each program, and in each time we chose a new random 3-byte watermark. In this first experiment, for each program TinyWatermark was able to correctly detect 30 times the random 3-byte watermarks. In other words, for every program TinyWatermark obtained 100% of TP and 0% of FP.

In the second experiment, for each program we also randomly chose a 3-byte key for inserting in (Blink, Sense, RadioCountToLeds, RadioSenseToLeds and BaseStation). Then each one of these watermarks was inserted in its respective program. After that, we simulated an attack to change the original watermark. In this attack we have changed a randomly chosen obfuscation technique in the inserted watermark. The technique belonged to the obfuscation techniques sequence. Then, we tried to extract watermark values, which differed from the current random 3-byte watermark, which was actually inserted in the first place. In every round of every program, TinyWatermark returned that the watermark could not be detected in this experiment, leading to the result of 100% of TN and 0% of FN. Therefore, TinyWatermark presented 0% of false detections (FN+FP) in the experiments performed. These results showed that the proposed watermark scheme is credible.

4.2.3 Stealth

We claim that the proposed watermark is stealthy because the watermark code can not be easily distinguished from the original code, this is because the watermark code is an essential part of the software, giving no indication that such code represents one watermarking. This is due we only make few changes in the code by replacing and inserting few instructions with the same semantic. Also, if this watermark code is modifyied or removed, the software behaves unexpectedly.

5 Conclusion

In this work, we presented a code obfuscation-based software watermarking framework for wireless sensor networks. The contributions of our framework are the proposal of a code obfuscation-based watermark and the mechanisms for insertion and extraction of such watermark. Both contribute for protecting the intellectual property of software programs developed for running on WSN. The effectiveness of TinyWatermark is supported by our simulation results. The proposed watermark is credible and has a low memory/processing overhead.

The future directions of our work are to investigate: (i) the use of tamper proofing techniques and (ii) the use dynamic watermarks, which are watermarks that change according to time and according to sofware behavior, also enhancing software security.

Acknowledgements

This work has been partially supported by Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq) through processes 477223/2012-5, 304941/2012-3 and 473851/2012-1; by Financiadora de Estudos e Projetos (FINEP) through process 01.10.0549.00; by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Instituto Nacional de Metrologia (Inmetro).

References

[1] J. Yick, B. Mukherjee and D. Ghosal. “Wireless sensor network survey”; Computer Networks, Vol. 52, No. 12, pp (2292-2330), (2008).

[2] C. Collberg, J. Davidson, R. Giacobazzi, Y. Xiang Gu, A. Herzberg and F. Wang. "Toward digital asset protection"; IEEE Intelligent Systems, Vol. 26, No. 6, pp (8-13), (2011).

[3] C. Collberg and J. Nagra. “Surreptitious Software”; Addison - Wesley Pearson Education, 2010.

[4] D. Boccardo, R. Machado, L.F. Carmo. “Transformações de código para proteção de software”; Minicursos SBSEG, (2010).

[5] Ying Zeng, Fenlin Liu, Xiangyang Luo, and Chunfang Yang. “Software Watermarking Through Obfuscated Interpretation: Implementation and Analysis”; Journal of Multimedia, Vol. 6, No. 4, pp (329-340), (2011).

[6] G. Xu and X. Guangli. “A Method of Software Watermarking”; IEEE International Conference on Systems and Informatics, pp (1791–1795), (2012).

[7] A. Lakhotia, D. Boccardo, A. Singh, A. Manacero Jr. “ Context-sensitive analysis without calling context”; Higher-Order and Symbolic Computation. Springer. Vol. 23, No. 3, pp (275–313), (2010).

[8] S. Thaker. “Software Watermarking via Assembly Code Transformations”; Thesis presented to the Department of Computer Science of San Jose State University. (2004).

[9] L. Chen and Z. Chaoquan. “A novel algorithm for .NET programs watermarking based on obfuscation”; International Symposium on Instrumentation & Measurement, Sensor Network and Automation, Vol. 2, pp (583-586), (2012).

[10] M. Smithson, K. Anand, A. Kotha, K. Elwazeer, N. Giles and R. Barua. “Binary rewriting without relocation information”; University of Maryland, (2010).

[11] ATmega128, available in http://www.atmel.com/dyn/products (Last accessed 01/2014).

[12] R. Alberola and D. Pesch. “AvroraZ: extending Avrora with an IEEE 802.15.4 compliant radio chip model”; 3nd ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks, pp (43-50), (2008).

[13] P. Levis and D. Gay. “TinyOS Programming”; Cambridge University Press, (2009).

Source-Location Privacy Protection in Wireless Sensor Networks using AZR Routing

Zhiwen Zeng[1], **Meiling Zeng**[1], **and Hui Liu**[2,*]
[1]School of Information Science and Engineering, Central South University, Changsha, Hunan, China
[2]Department of Computer Science, Missouri State University, Springfield, Missouri, United States

Abstract - *In environments where sensor networks are deployed to monitor certain events or sensitive objects, attackers may deduce the approximated location of monitored objects by hop-by-hop backtracking the traffic. This paper proposes a routing technique to provide adequate source-location privacy with balanced energy consumption, the Phantom Routing based on Annular Zone (AZR). With this technique, the entire network is divided into several layers according the distance between the nodes to the SINK node. The source node randomly selects a phantom source within the same layer when the source node is far away from the SINK node. Otherwise, the phantom source is selected in the FAR layers randomly. The message will then be routed from the source node to the phantom source through the annular routing path. The AZR protocol ensures that the phantom node is distributed uniformly in the entire network, and the routing path from the source to the SINK avoids the one-hop dangerous zone from the SINK. While ensuring source-location privacy, our simulation results prove that the proposed scheme can provide performance better than the existing schemes.*

Keywords: source-location privacy; wireless sensor network; annular zone routing; safety period;

1 Introduction

A wireless sensor network is composed of spatially distributed autonomous, low-cost and energy-efficient sensors to monitor physical or environmental conditions, and to cooperatively forward their data through the network to a main location, the SINK node. It has gained more popularity in recent years and been widely used in both military and civilian applications. However, one of the primary concerns that hinder the successful deployment of wireless sensor networks is to protect privacy information running on the network due to its broadcast wireless media. Privacy in a network consists of not only the privacy of the message content but also the privacy of the source locations. A variety of mechanisms such as encryption can protect the confidentiality of the message content. Even if a powerful encryption algorithm is used to protect the source identity, the well-equipped attacker may still be able to determine the location of the source by monitoring the traffic patterns and routing paths, which is a significant concern especially in environments to monitor sensitive objects. For example, WSN might be deployed in wild area to monitor endangered animals or on battlefield to acquire real-time military information. The protected animals or the soldiers will be exposed to attackers if the attackers find the source-location.

In the past two decades, a number of researches aiming at protecting source location have been proposed. Kamat et al. [1] propose a classic source-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 [3]. To solve this problem, several approaches have been proposed. Reference [2] designs directed walk through either sector-based approach or hop-based approach in order that the phantom source can be away from the actual source. Xi et al. [4] 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. The main drawback of this work is that the delivery time of messages is more instable.

The authors in [5] observe that the increasing of path length in random walks cannot necessarily make improvement on the safety period because phantom sources are not always placed in a secure location to initiate the routing phase. An attacker placed in the shortest path from the sink to the source is more likely to find the source if the angle of arrival of the message is usually less pronounced. Therefore, the authors proposed the phantom routing with location angle (PRLA) by introducing the inclination angles of sensor nodes in random walk area. PRLA chooses the next hops in random walk area with different probabilities, which can optimize the routing path and greatly increase the safety period.

The authors in [6] propose the random intermediate node (RRIN) scheme, where messages are sent to a random intermediate node that will eventually send the message to the

[*] The corresponding author: huiliu@missouristate.edu.

SINK node. The intermediate nodes are selected based on their relative location with the consideration that they are placed at least at a distance d_{min} from the source node. Normally the intermediate nodes are distributed outside the constrained are. With this scheme, the selected intermediate node is expected to be away from the actual source node. However, the energy consumption of this scheme is quite high.

In order to well balance the energy consumption and privacy protection, J. Ren et al. [3] 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. However, the energy consumption of the entire network is not balanced since nodes inside the toroidal region might drain out energy quickly while nodes outside the toroidal region seldom consume energy.

All of the above-mentioned researches do not consider the source-location privacy protection of nodes close to the sink node. In this paper, we 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. Energy consumption along with source-location privacy are two very vital components for the successful deployment of wireless sensor networks. Simulation results of our proposed protocol prove that our AZR protocol has better performance than the current existing schemes in those two components.

The remainder of this paper is organized as follows. In Section II, the system model is described. Details of the proposed source-location privacy scheme are illustrated in Section III. Section IV compares and analyzes the performances of AZR and STaR based on the simulation studies. Section VI concludes the paper.

2 Models

2.1 The System Model

The system is similar to the Panda-Hunter Game [1]. A wireless sensor network is deployed in a habitat to monitor the location of a panda in the Panda-Hunter Game, where sensor nodes are used to locate the general area of the panda. When the panda is discovered, the corresponding source node will observe and send data periodically to the sink node. The goal of our design is to make it infeasible for the hunters to determine the location of the panda by analyzing the traffic patterns in the network.

The following assumptions are made about the system:

- A wireless sensor network is deployed with equal density throughout a circular region. The whole network is fully connected through multi-hop communications [7]-[9].
- The only SINK node is located at the center of the circular network that is the destination location that data messages will be routed to. At any given time, there is only one source node, for example, there is only one panda in Panda-Hunter game, and then the sensor node closest to the panda will generate and send messages to the SINK node periodically through a multi-hop routing.
- Each sensor node not only knows their relative locations and the SINK node location, but also has the knowledge of its adjacent neighboring nodes. The information about the relative location of the sensor domain may also be broadcasted through this network for routing information update [10]-[11].
- Sensor nodes of the entire network are divided into several different layers with the SINK node as the center according to the relative distances between the sensors and the SINK node.

2.2 The Attackers Model

In this paper, the attacker has the following characteristics:

- Well-equipped: the attacker has enough memory space to store any information useful to him and adequate computation capability. The attacker could determine the immediate sender of the message by analyzing the strength and direction of the signal he received on detecting an event message. He is able to move to this sender's location without much delay.
- Passive: the attacker cannot tamper any contents of the messages transmitted in the senor network, or do any damage to the sensors. We assume the attacker is only able to monitor certain area of the sensor network and compromise a few network nodes, instead of all the traffic through the entire network. The monitoring radius of the attacker equals to the radius of transmission range of sensors in our assumptions.
- Initial status: the attacker is close to the SINK node and is observing the communicating messages between the SINK node and its neighboring nodes. On detecting the message event, the attacker moves the sender's location without any delay.

2.3 Energy Consumption Model

Energy consumption model [12] 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.

$$\begin{cases} E_{member} = lE_{elec} + l\varepsilon_{fs} d^2 & d < d_0 \\ E_{member} = lE_{elec} + l\varepsilon_{amp} d^4 & ifd > d_0 \end{cases} \quad (1)$$

$$E_R(l) = lE_{elec} \quad (2)$$

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. ε_{fs} and ε_{amp} are the energy required by power amplification in these two models, respectively. The above parameter settings are given in Table 1 [12].

Table 1. Network parameters

Parameter	Value
Threshold distance (d_0) (m)	87
Sensing range r_s (m)	15
E_{elec} (nJ/bit)	50
ε_{fs} (pJ/bit/m^2)	10
ε_{amp} (PJ/bit/m^4)	0.0013
Intial Energy (J)	0.5

3 Source-location privacy scheme - AZR

The proposed scheme has three phases. Firstly, when a source node has a message to transmit, it randomly selects a phantom source node in the same layer if the source node is far away from the SINK node, and otherwise, the phantom source is selected in the FAR layers randomly. Then the source node forwards this message to the selected phantom source via annular routing path or radial routing path. Finally, the message is forwarded from the phantom source to the SINK node through the shortest path. The detailed description of the proposed scheme will be described in the following.

3.1 The Selection of Phantom Source

The entire network is divided into different layers according to the relative distances between sensors and the SINK node, and these layers are labeled as Layer 1, Layer 2…Layer k, respectively, from inside to outside as shown in fig. 1. We call Layer 1 and Layer 2 as NEAR layers, and other layers as FAR layers. However, the idea of NEAR layers and FAR layers is a relative definition that depends on privacy protection strength. For example, if requirement of privacy protection is low, outside layers from Layer 2 can be counted as FAR layers, while only the most outside layer is considered as FAR layers, the network provides the strongest privacy protection.

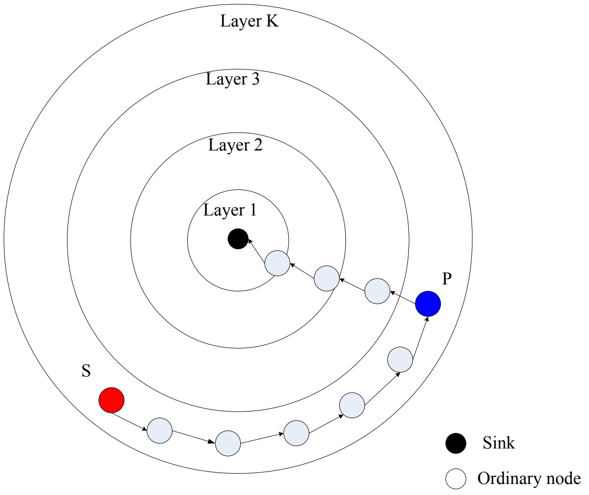


Fig1. Source node in FAR layers selects the phantom source

When the source node is located inside FAR layers, one of all the sensors within the same layer with the source node will be selected as the phantom source randomly as shown in fig.1. When the source node is located inside NEAR layers but not within one-hop area from the SINK node, one of all the sensors inside FAR layers will be denoted as the phantom source randomly as shown in fig. 2. The randomly selecting the phantom node aims at increasing the length of routing path from the source to the SINK node, and providing more varieties of routing paths, further, improving the safety period. If the source node is within one-hop area of the SINK node, there exist no schemes that can protect source-location privacy.

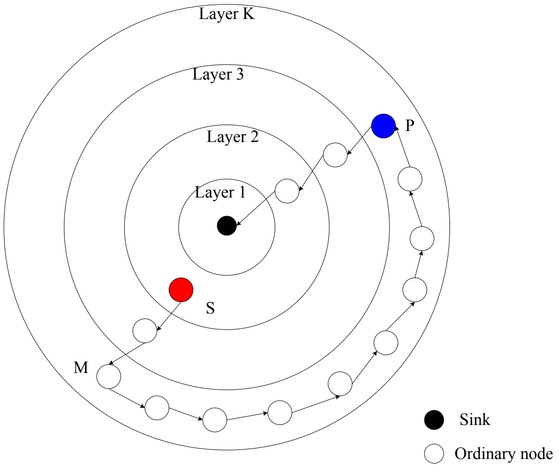


Fig 2. Source node in NEAR layers selects the phantom source

3.2 Routing strategy

Definition 1. Suppose nodes A and B are located within the same layer in a WSN, *the annular routing path* is the shortest routing path between node A and B, while all the intermediate nodes along this path are located within the same layer with nodes A and B.

Definition 2. Suppose node *S* is in Layer *I*, node *D* is in Layer *J*, $I < J$, the SINK node, node *S* and node *D*, these three nodes, form a straight line *L*, while the distance between node *S* and node *D* is less than the distance between the SINK node and node *D*, and all the intermediate nodes are located in *L*, thus *the radial routing path* is the path between node *S* and node *D*.

Definition 3. The transmission radius of sensor node is *r*, the circular region with the SINK node as the center and *r* as the radius is called as *the dangerous region*.

When the routing path from the source node to the phantom source passes through the dangerous region, the attacker is able to backtrack the source node directly from the SINK node without tracking the phantom source. Therefore, traffic bypassing the phantom source only increases the message transmission delay and energy consumption, but does not improve source-location privacy at all.

The routing algorithm from the source node to the phantom source (Phase 2) is described as the following:

If the source node is located within FAR layers, one node in the same layer with the source node is randomly selected as the phantom source, and the message is sent from the source node to the phantom source by the annular routing path as in fig. 1. The path from *S* to *P* in fig. 1 is called an annular routing path. If the source node is within the NEAR layers, one node located within the FAR layers is selected as the phantom source randomly, with the assumption that the selected phantom source is in layer *K*, the message is routed from the source node to the intermediate node *M* in layer *K* by radial routing path, and then the message continues to be sent to the phantom source from the intermediate node *M* by annular routing path as shown in fig. 2, and then the path from *S* to *P* in fig. 2 is named as a radial routing path.

The combination of the annular routing path and the radial routing path not only increases the routing path by consuming the energy of sensors located in outside layers, but also make the routing path avoids the dangerous region eventually, therefore, improve the safety period.

The message is routed from the phantom source to the SINK node by using the shortest routing path in Phase 3.

4 Simulation and performance analysis

We use the discrete event-based simulator-OMNET++ [13] to simulate the two protocols, AZR and STaR, comparing their safety period, message latency and energy consumption.

4.1 Simulation environment

Assume there are 1800 sensor nodes uniformly distributed in a circular area with the radius of 400 meters, *R*. Every node can communicate with those nodes no more than 50 meters far from it. The entire network is divided into 8 layers, and the width of every layer equals the transmission radius of a sensor node to simulate AZR protocol. The constrained area is defined in the toroidal region with the radius of inner-edge, 0.28R and the radius of the outer-edge, 0.38R in simulating STaR [3].

A simple MAC protocol is used in sensors so that they have the ability to retransmit messages when wireless communication collides.

In our simulation, Layer 1 and Layer 2 are counted as NEAR layers, while all the outside layers from Layer 3 including Layer 3 are denoted as FAR layers.

The longest distance on which attackers could eavesdrop is the same as the transmission range of sensors. We randomly select 50 sensors in every layer as source nodes, and every source node generates messages 200 times and sends them according to AZR and STaR routing schemes, respectively. That means that 200 times trace back simulations will be done for each test point in the results.

4.2 Performance analysis

4.2.1 Safety period

Safety period is defined as the number of messages being sent by the source from beginning to the end of one capture, which is usually used to evaluate the capability of source privacy for communication protocols. In our simulation experiments, the attacker is located close to the SINK node when the source node begins to transmit messages. The number of messages being sent by the source to the SINK node is recorded for evaluating the safety period until the attacker finally backtracks the source node.

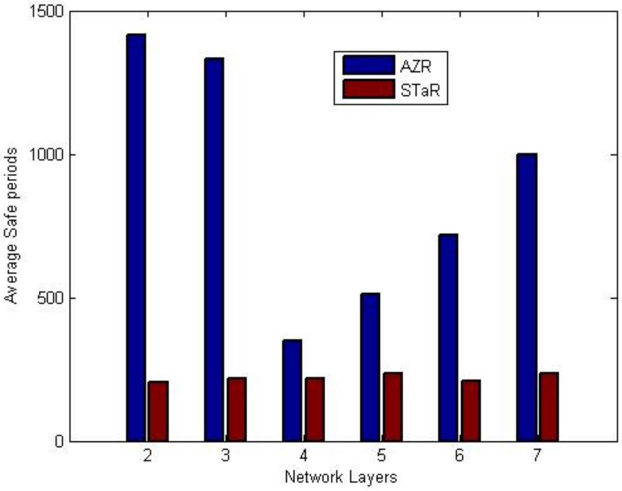


Fig 3. The average safety period of AZR and STaR

As shown in fig. 3, it is obvious that AZR protocol achieves much larger safety period than the STaR protocol with the source nodes located in the same layer. When the source nodes are in Layer 2 and Layer 3, the safety period of AZR protocol is 1450 and 1300 respectively, while that of STaR is 221 and 225. Averagely, AZR improves the safety period by up to 7 times in these two layers, which can be explained by that in AZR, source nodes within NEAR layers select the phantom sources in FAR layers, which enormously increases the length of routing path, therefore AZR can obtain better safety period than STaR protocol. AZR well designs the source-location privacy protection for nodes that are very close to the SINK node. We also observe in fig. 3 that the safety period of AZR is 380 and 998 respectively when the source node is in layer 4 and layer 7, while that of STaR stays in the range from 220 to 230. The safety period of STaR protocol is relatively stable no matter which layer the source node is located within, however, the safety period of AZR is increasing significantly when the locations of sensors become farther from the SINK node, because AZR protocol provides more variety of routing paths for source node compared with STaR protocol.

4.2.2 Message latency and energy consumption

The latency of a message is denoted by the length of routing path that the message goes through from source to SINK node being described as hops count. Fig. 4 indicates the variation of message latency of AZR and STaR with the variation of network layer. The length of routing path in STaR changes little, while that of AZR varies a lot with the source nodes locating in different layers. And averagely the length of routing path in AZR is in the range of 1.3 to 2 times of that of STaR. Since the phantom source is always selected in the constrained area in STaR, and the selection of phantom source is well designed. The longer the length of routing path is, the more difficult the attacker is able to track the source node. Thus, the length of routing path represents the protection strength for source-location privacy to some degree. In order to improve the source-location privacy protection, we prefer increasing the length of routing path from the source to the SINK node.

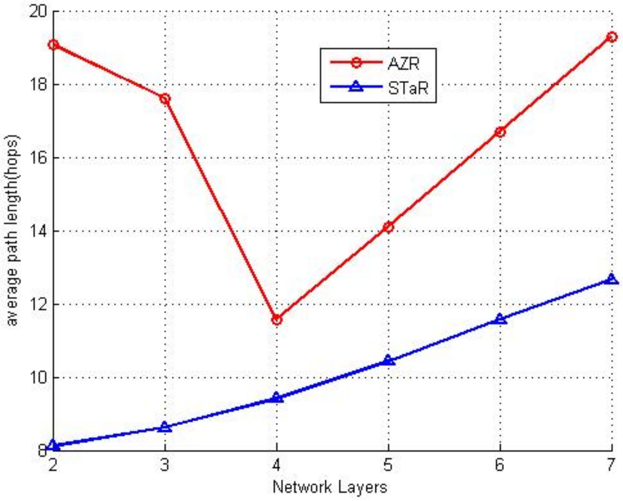


Fig 4. The message latency of AZR and STaR

Although the message latency of AZR is longer than that of STaR, the AZR protocol does not reduce the lifetime of the entire network as shown in Fig. 5 and Fig. 6. Simulation results indicate that nodes closer the SINK tend to carry heavier traffic loads than nodes farther away from the sink, which will deplete their energy faster in both AZR and STaR protocols. The average energy consumption of STaR is smaller than that of AZR, however, the energy consumption of AZR is more balanced than that of STaR. Therefore, AZR has better performance to avoid energy-hole problems [14] and improve the energy consumption rate in WSN than STaR.

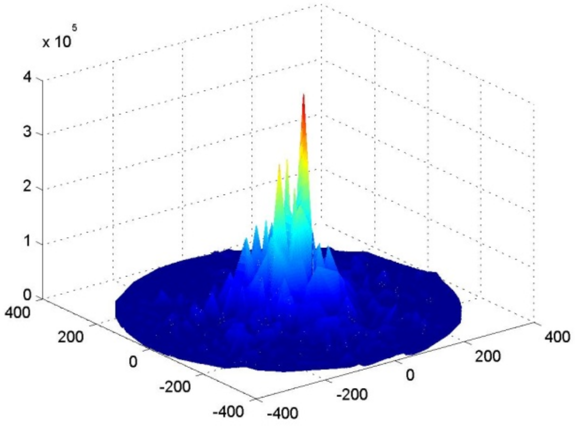


Fig 5. The energy consumption of STaR

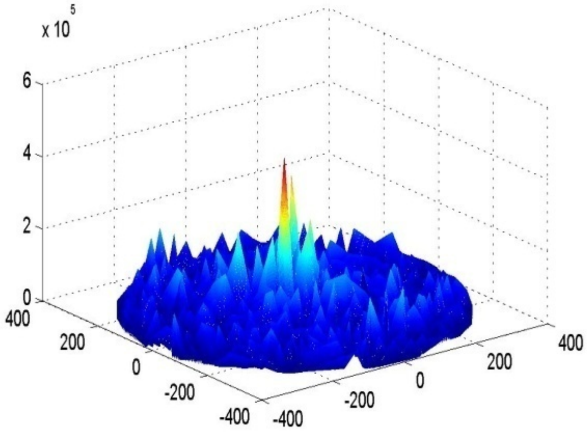


Fig 6. The energy consumption of AZR

5 Conclusion

Source-location privacy is significantly important to the successful deployment of wireless sensor networks. In this paper, we propose a novel routing protocol for protecting source-location privacy named as AZR. We carry out extensive simulation experiments to compare our proposed protocol with other existing schemes. Our simulation results demonstrate that the proposed AZR routing protocol can effectively enhance the safety period and achieve balanced energy consumption.

6 References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] Weiping wang, liang chen, jianxin wang, "A Source-location Privacy Protocol in WSN based on Location Angle", ICC'08: IEEE International Conference on Communications, 2008, pp. 1630-1634.

[6] Y. Li, J. Ren, "Providing source-location privacy in wireless sensor networks", WASA'09: proceedings of Fourth International Conference on Wireless Algorithms, Systems, and Applications, 2009, pp. 338-347.

[7] M. Ye, C. Li, G. Chen, J. Wu, "Eecs: an energy efficient clustering scheme in wireless sensor networks", IPCCC 2005, pp. 535-540, April 2005.

[8] J. Neander, E. Hansen, M. Nlion, M. Bjorkman, "Asymmetric multihop communication in larger sensor networks", 1st Internation Symposium on Wireless Pervasice Computing, 2006, Jan 2006.

[9] Q. Younis, S. Fahmy, "Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks", IEEE Transactions on Mobile Computing, vol. 3, pp. 366-379, Oct.-Dec. 2004.

[10] Y. Zhang, W. Liu, Y. Fang, D. Wu, "Secure location and authentication in ultra-wideband sensor networks", IEEE Journal on Selected Areas in Communications, vol. 24, pp. 829-835, April 2006.

[11] P. Zhang, M. Martonosi, "Locale: Collaborative location estimation for sparse mobile sensor networks", pp. 195-206, april 2008.

[12] 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.

[13] OMNet++ Network Simulation Framework, http://www.omnettp.org/.

[14] Aruna Pathak, Zaheeruddin, M. Tiwari, "Minimizing the energy hole problem in wireless sensor netowrks by normal distribution of nodes and relaying range regulation", CICN'12, the proceedings of the 2012 Fourth International Conference on Computational Intelligence and Communication Networks, 2012, pp: 154-157.

Applying TCP Profiling to Detect Wireless Rogue Access Point

James Yu
College of Computing and Digital Media, DePaul University
Chicago, Illinois USA 60604

Abstract - There are many studies addressing the risk and vulnerability of 802.11-based Wireless LAN (WLAN). However, Rogue Access Points (AP) attacks, although well known and widely studied, do not have a cost-effective solution yet. This paper presents our studies of three attacking scenarios, and several common industry approaches trying to address this problem. Our analysis shows the weaknesses of RF-based solutions, and we propose to use Round Trip Time (RTT) and jitter to profile TCP traffic. Because of different profiling between wired and wireless traffic, we could use this difference to identify wireless traffic on wired network, from which we could detect rogue AP attacks. Our proposed solution is validated on a network simulation environment (ns2).

***Key-Words* —**Wireless LAN, Rogue AP, Jitter, Delay, RTT.

I. INTRODUCTION

THE wide deployment of 802.11-based wireless LANs (WLAN) also increases the risk of security attacks. Most studies of WLAN security focus on crypto attacks, and the 802.11i-2004 standard effectively addresses this issue [1]. Denial of Service (DoS) attacks are also common on WLAN, and we conducted extensive studies on the DoS attacks and their protective measures [2]. The problem of Rogue Access Points (AP) is well known, and there are many white papers discussing its problems and effects [3]. The problem of Rogue AP is not just unauthorized access to steal proprietary and confidential information. A hacker could use Rogue AP to break into the corporate intranet and launch various attacks, such as ARP attacks, DHCP attacks, port scanning, etc. A hacker may use rouge AP to recruit *bot*s and to build a *botnet* within the corporate intranet. The hacker can then launch a Denial of Service (DoS) attack from the corporate intranet to any target on the public Internet. A study shows that about 20% of companies have Rogue AP problems at some time [3]. Almost every WLAN vendor has a solution to address this problem. In general, the industry solution is to deploy expensive wireless sensors at various locations to detect rogue APs, but there are few research studies discussing its effectiveness.

This paper presents different scenarios of the Rogue AP attacks and their threats to the enterprise environment, along with several proposed solutions. The purpose of our research is to analyze if these solutions are effective in detecting rogue APs, and then to develop a more efficient and cost-effective solution to address the problem.

II. ROGUE AP ATTACKING SCENARIOS

The authors of Reference [4] identify four types of rouge AP attacks which are *unauthorized* AP, *improperly configured* AP, *compromised* AP, and *phishing*. The advancement in WLAN management has properly addressed the issues of improperly configured AP and compromised AP. For example, CAPWAP provides a centralized management scheme with a wireless controller which could prevent misconfiguration of APs [5].

The case of phishing is that a wireless client is connected to a Rogue AP which then collects personal data from the user. This is also known as the *Evil Twin* problem of WLAN. A network side solution is to identify duplicate Service Set ID (SSID) with different MAC addresses. Reference [6] provides a solution to establish the profile of legitimate AP through statistical analysis, and use the profiling to detect phishing. Our preferred solution to phishing is *mutual* authentication. On the enterprise environment, the recommendation is to deploy a mutual authentication protocol such as Protected Extensible Authentication Protocol (PEAP) over 802.1X. At the hotspots, users also need to consider a mutual authentication mechanism from the wireless service provider, such as security code via text messaging. If a user cannot authenticate an AP from a wireless service provider, the user should avoid providing personal data on any untrusted web site.

Our study focuses on the case of unauthorized AP, and the scope covers three cases of commercial AP, workstation (soft) AP, and wireless ad-hoc station.

A. Unauthorized Commercial AP

The most common case of rogue AP is that an employee brings an AP to the corporate environment, and then connects the AP to an Ethernet switch port on the corporate LAN as illustrated in Figure 1.

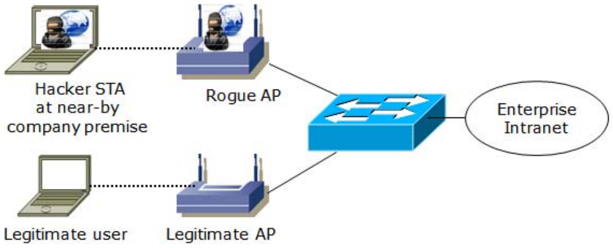


Figure 1. Rogue AP Attacks against Enterprise

The 802.11 standard references the wired LAN as the Distribution System (DS). With an AP on the LAN, multiple wireless clients can connect to the corporate network. This is a convenient and low cost solution to support multiple wireless users to access the corporate intranet and/or the public Internet at the company premises. There is a growing demand to use smart phones and tablets at the enterprise environment, and these devices have dual mode wireless connection: 802.11 and cellular. It is possible that some users may bring their own APs to the corporate environment for their personal mobile devices. However, this configuration poses a major security threat where hackers can also access the corporate network through this *unauthorized* AP. A corporate security policy should prohibit employees from using unauthorized APs. However, a security policy is not sufficient and a technical solution is required to enforce the compliance to the policy. A more critical issue is the case that an employee intentionally installs a rogue AP to hack the network for personal gain. The hacker may even configure the strongest security (AES for encryption) on the AP and hide its SSID from being detected. Security rules have no effect on hackers who intentionally break the rules.

B. *Soft AP*

The function of AP can be easily implemented by software, so a hacker can also configure his/her workstation (Linux or Windows) as an AP [7]. This is also known as soft AP or host AP.

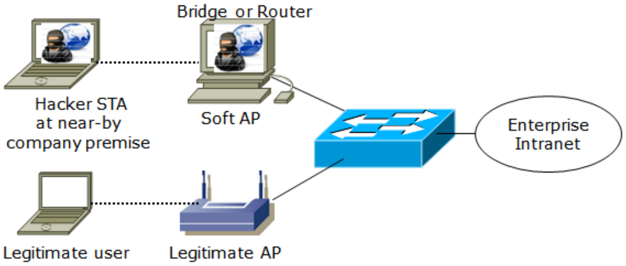


Figure 2. Soft Rogue AP Attacks against Enterprise

The soft AP has two interfaces: the wired interface (Ethernet) connected to the corporate LAN (DS) and the wireless interface open to any wireless client. A commercial rogue AP has the physical appearance to be recognized by others. A soft AP has the same appearance as a typical workstation, and it is a lot more difficult to detect a soft rogue AP.

C. *Ad-Hoc/STA*

Similar to soft AP, a hacker can also use the wireless *ad hoc* configuration to create a rogue AP like attacking scenario. The physical connection of this scenario is the same as soft AP, but there is no AP. The workstation, referenced as Ad-Hoc/STA in this paper, uses the Ethernet connection to the corporate LAN, and uses the wireless *ad hoc* configuration to connect other wireless clients. The Ad-Hoc/STA is running Layer-3 forwarding or routing between its wired and wireless interfaces. Because there is no AP, most rogue AP detection schemes cannot detect this attacking scenario. Also, a hacker may intrude a laptop, configure it as an Ad-Hoc/STA, and use it as a gateway to break into the corporate intranet.

III. COUNTER MEASURES TO ROGUE APS

A. *Wireless Sensor*

A common protection measure against rogue APs is to use wireless sensors and deploy them at various locations on the enterprise premises. Many APs and wireless clients also have the capability to act as wireless sensors to detect APs within their Radio Frequency (RF) range. The detection is based on the *beacon* and/or *probe request/response* messages which contain the SSID of the AP. However, a hacker may hide his/her AP by not sending these messages. Also, a rogue AP may use the same SSID as a legitimate AP. In a Multi-Tenant Environment (MTE), a wireless sensor may identify many unknown APs which are used by neighboring companies. In summary, if a wireless sensor relies on SSID, it is not effective in detecting rogue APs as it would have too many false alarms.

B. *Medium Access Control (MAC) Address Analysis*

In addition to identifying beacon and probe messages, a wireless sensor can collect wireless traffic to identify the MAC addresses of APs and wireless clients [8]. According to the 802.11 standard for infrastructure communication (i.e., communication between AP and wireless client), the AP's MAC source, the source MAC address (SA), and the destination MAC address (DA) are in the 802.11 frame header. A wireless sensor can sniff wireless traffic and collect the MAC addresses of the AP and wireless clients. The sensor then sends the data to the Wireless Intrusion Detection/Protection System (WIDS/ WIPS) for further analysis.

The WIDS has a local database which contains the MAC addresses of legitimate APs. When provisioning a new AP on the enterprise network, network administrator registers the AP in the WIDS database. The WIDS also collects the MAC addresses of stations on the wired LAN (DS). When the WIDS detects an unregistered AP, it would then check if SA and DA of the 802.11 frame are also detected on the wired LAN (DS). If there is no SA or DA on the DS, this AP is likely to be a neighborhood AP. This information of neighboring AP shall be recorded, but no alarms are generated. On the other hand, if SA or DA is also detected on the DS, this *unregistered* AP is likely a rogue AP and an alarm should be sent to the network administrator. An example of using MAC addresses to detect rogue APs is illustrated in Figure 3.

On this network, VLAN-1 is for wired traffic and VLAN-5 is for wireless traffic. If there is a Rouge AP, its traffic will be on VLAN-1 because the Rogue AP is connected to a LAN port for wired traffic.

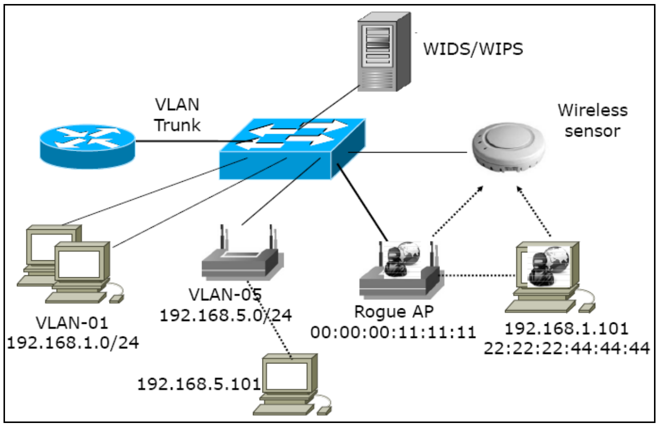


Figure 3. Detection of Rogue AP by MAC Addresses

An example of detecting scenario is given as follows: first, an AP address (00:00:00:11:11:11) is identified and this MAC address is not registered in the WIDS. The WIDS also detects a client MAC address (22:22:22:44:44:44) associated with the *unregistered* AP on the LAN side. Therefore, this AP is not a neighboring AP as it sends traffic to the enterprise LAN. The security rule then concludes the detection of a rogue AP, and an alarm is sent to the network administrator. This seemingly sound approach has at least three issues:

1. It is difficult and costly to deploy wireless sensors at every possible location on the company premises.
2. If a hacker configures the AP as a router or a layer-3 forwarding device, the WIDS will not detect any wireless client MAC address on the DS because the router hides layer-2 (MAC) information.
3. This approach does not work with the ad hoc configuration, as there is no AP. The wireless client addresses are hidden behind the Ad-hoc/STA, and cannot be detected on the DS.

C. Other Approaches of Rogue AP Detection

Cisco proposes a solution to send a deauthentiction or a disassociation frame to disconnect clients from a potential rogue AP, and then to establish a client connection to the AP to check if it is connected to the DS [9]. This approach would not work if the hacker turns on 802.11i for authentication and encryption, and enables 802.11w to authenticate the management frames. The proposed solution in [10] is similar to the Cisco solution and a hacker can easily reject those probe messages. Reference [11] proposes to dispatch a mobile agent to every wireless client and each client needs to be authenticated with a centralized network manager. This approach is effective to detect rogue wireless clients connected to legitimate APs rather than to detect rogue APs. Reference [12] proposes to use wavelet transformation to create a signature for each vendor's device, from which a rogue AP can be detected by checking the signature. However, it does not address how to distinguish rogue APs from neighbor's APs. In summary, sniffing RF traffic is not only costly but also ineffective in detecting rogue APs.

IV. Wireless Network Design

In order to effectively identify rogue AP, we first propose to separate the wired traffic from wireless traffic on the enterprise network as illustrated in Figure 4.

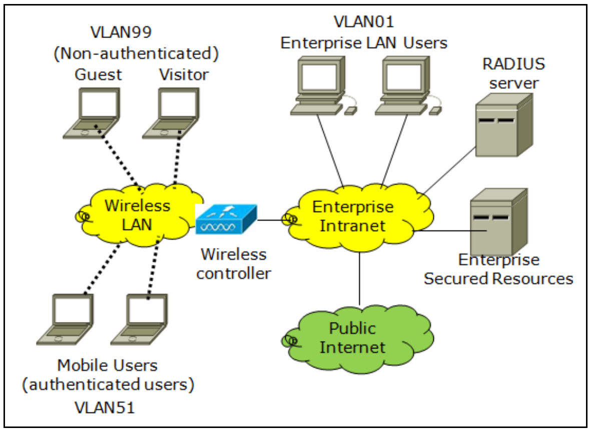


Figure 4. Separation of Wired and Wireless Traffic

In this network design, wireless traffic and wired traffic are on different Virtual LAN (802.1Q tagged VLAN) and different IP subnets. We can further distinguish guest (non-authorized) and authorized users of wireless LAN by the VLAN ID and IP subnets. If we observe wireless traffic on the wired VLAN or IP subnet, we can conclude the detection of a Rogue AP. This approach does not require expensive wireless sensors to sniff wireless traffic as all traffic is collected on the wired side. The security enclave for wireless traffic shall follow the strongest protection scheme as specified in 802.11-2012 [1]. In the next section, we will discuss how to distinguish between wired and wireless traffic.

V. Traffic Flow Analysis

It is well known that 802.3 and 802.11 traffic has significantly different characteristics, and the difference can be measured by delay or Round Trip Time (RTT) [13]. The 802.11 access method is Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). All wireless workstations share the same media (i.e., RF channel), and only one station is allowed to transmit data at a time. If two or more stations are transmitting the data at the same, there would be *collisions*. To avoid collisions, workstations need to follow the CSMA/CA procedure as illustrated in Figure 5. The procedure of collision avoidance requires *random* delay which would cause large variation of delay (i.e., jitter). On the switched Ethernet network, there is no collision and the delay can be calculated from packet size and network speed. Therefore, the jitter of switched Ethernet is expected to be very small.

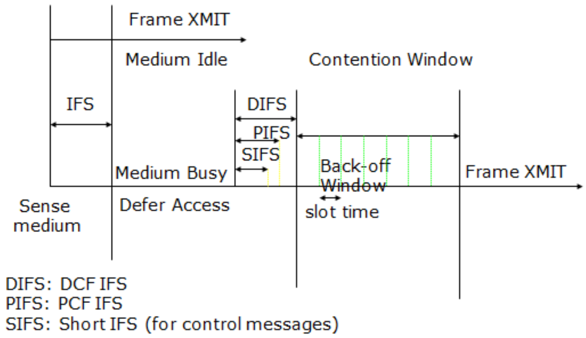


Figure 5. 802.11 CSMA/CA Operation

The delay of 802.11 frame can be calculated as follows:

$$T_{delay} = t_{IFS} + t_{defer} + t_{bf} + t_{data} + t_{retransmit}$$

where

- $t_{IFS:}$ InterFrame Spacing which has two components
- $t_{bf:}$ Delay due to the back-off window
- $t_{defer:}$ Waiting time due to the busy channel
- $t_{data:}$ Data transmission time which includes a fixed processing time and variable transmission time based on the packet size.
- $t_{retransmit}$ 802.11 Retransmission Time

The delay of wired Ethernet (802.3) frame has only t_{data} and $t_{ifs.}$ For wireless traffic, the major causes of delay variation are determined by t_{defer} (media busy) and t_{bf} (back-off and waiting) especially in the case of high contention. The reference of [14] uses inter-arrival time to distinguish between wired and wireless traffic. However, our analysis shows that T_{delay} would have much larger variation than inter-arrival time. In practice, we measure Round Trip time (RTT) instead of one-way delay which requires perfect clock synchronization. Both ICMP and TCP have a built-in mechanism in the protocol to support real-time RTT measurement. However, there are several issues with ICMP:

- ICMP is management traffic and not user data traffic. Network administrator needs to proactively send ICMP traffic to probe suspected hacker activities. In order to collect sufficient data for analysis, network administrator needs to continuously send ICMP traffic which would have negative performance impact on normal data traffic.

- RTT must be computed between the WIDS and the rogue wireless client, but a hacker may disable ICMP on the wireless client. WIDS can send ICMP *echo requests* but would not see any ICMP *echo reply* back.

- Hacker can easily block the ICMP traffic on the rogue AP. Therefore, WIDS cannot send ICMP to the wireless side of rogue AP.

- ICMP requires the knowledge of the wireless client IP address, which may not be available to the WIDS if the rogue AP enables Network Address Translation (NAT).

Our approach is to measure RTT of TCP segments because hackers cannot block or hide TCP traffic. It is well known that TCP congestion control is designed for wired network and not effective for wireless network [14]. Based on this and other studies, we can further assume that WLAN has longer RTT and larger RTT variation due to ineffective congestion control and acknowledgement scheme. TCP (RFC1323) defines RTT as the interval between sending a TCP segment and receiving its acknowledgement as illustrated in Figure 6. It should be noted that a TCP acknowledgement may acknowledge multiple TCP segments. In this case, RFC1323 specifies RTT by using the timestamp of the 1st segment being acknowledged.

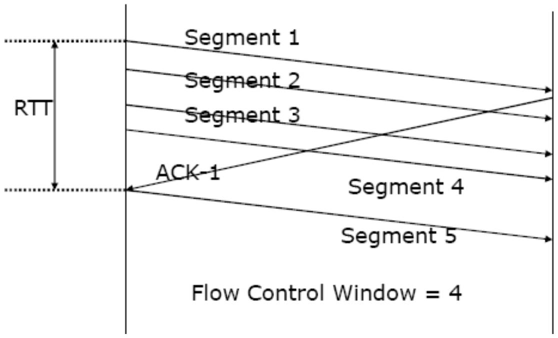


Figure 6. RTT Measure of TCP Segment

TCP assumes packet loss due to network congestion and adjusts its flow control window accordingly. However, the primarily reason of packet loss on 802.11 is contention or poor signal strength, and not congestion. As a result, TCP incorrectly adjusts its flow control window and causes poor performance [15]. In summary, we identify three major causes of delay variation of wireless TCP-RTT:

1. When there is contention, there will be a long delay of t_{defer} (waiting for busy media to become free).

2. After the long delay of t_{defer}, there is a random delay of t_{bf}. Because this is a delay of random back-off window, the variation is expected large.

3. The inconsistency of TCP and wireless flow controls will cause random adjustment of flow control window, which introduces large variation in delay.

The variation of RTT is referenced as *jitter*, and we follow the Real-Time Protocol (RFC 3550) to compute the jitter as follows:

```
for each packet[i] from 1 to n {
        RTT[i]  = RecvTime[i] – SendTime[i]
        delta[i] = | RTT[i] – RTT[i-1] |  # absolute value
        jitter[i] = jitter[i-1] + (delta[i] – jitter[i-1])/16
}
```

VI. Experimental Design and Measurement

We first collected RTT data on a live lab WLAN environment. Although we were able to apply TCP profiling to distinguish between wired and wireless traffic, the difference is not significant. The reason is the limitation of the lab environment to create practical contention scenarios. We then designed our experiment on a simulation environment using Network Simulator-2 (ns2). The trace data from ns2 supports the computation of RTT and jitter for wired and wireless traffic. The wired network (802.3) of this experiment is illustrated in Figure 7.

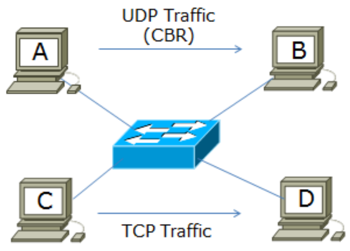


Figure 7. Network for Wired Traffic Measurement

The RTT and jitter measurements of wired traffic (802.3) are illustrated in Figure 8. We followed RFC2544 and measured RTT for the packet size of 64, 128, 256, 512, 1024, 1280, and 1500 bytes.

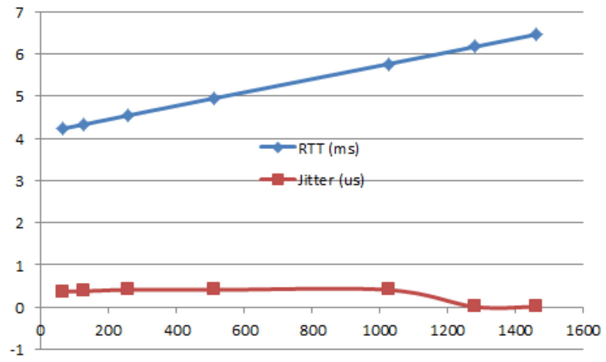


Figure 8. RTT vs. Packet Size for Wired (802.3) Networks

Note that RTT measurement is in ms (10^{-3}), and the jitter measurement is in μs (10^{-6}). In other words, there is almost no jitter observed on the wired traffic. In this experiment, we also tested different data rates of the background UDP traffic, and it has no effect on the TCP traffic as measured by RTT or jitter. This result is consistent with the behavior of *switched* Ethernet. On a full duplex switched Ethernet, each route is dedicated and not shared, so there is no contention on the network. RTT is determined by (a) a fixed processing time on the source and sink nodes, and (b) a variable transmission time based on the size. The data of Figure 8 shows what we expected – RTT curve is a straight line and its major cause of variation is packet size. The jitter is almost zero for all packet sizes.

The network topology of wireless experiment is similar to the wired experiment with an AP in the center. The traffic is between a wireless client and a wired client as illustrated in Figure 9.

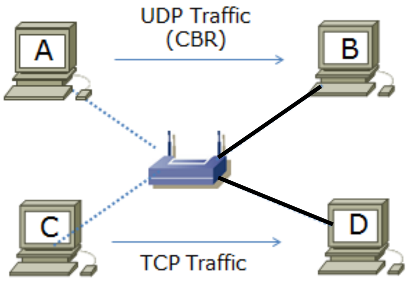


Figure 9. Network for Wireless Traffic Measurement

The TCP traffic is the same as the scenario of the wired experiment. The purpose of the UDP traffic is to create contention and to observe its effect on the TCP traffic. We conducted four different test scenarios of (a) no contention – no UDP traffic, (b) low contention (CBR of 80 kbps), (c) medium contention (CBR of 800 kbps), and (d) high contention (CBR of 8,000 kbps). The RTT chart is shown in Figure 10, and the jitter chart is shown in Figure 11.

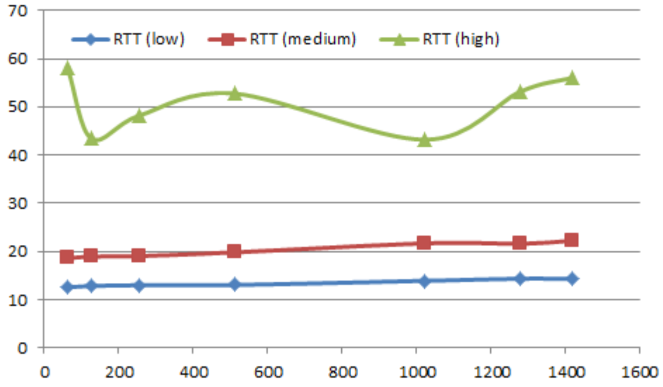


Figure 10. RTT (in ms) of Wireless (802.11) Networks

The results of no contention and low contention are almost identical. To make the chart easier to read, we do not show the RTT and jitter data of no contention in Figure 10 and Figure 11.

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

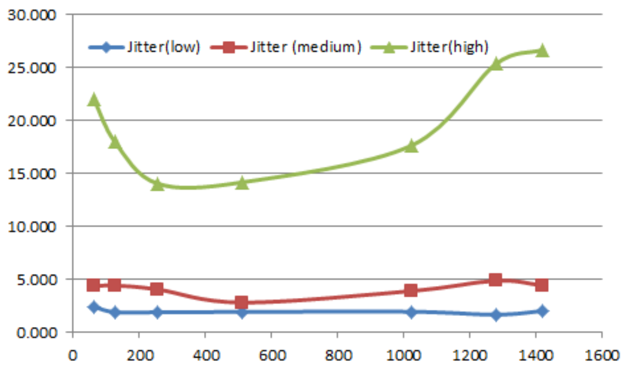


Figure 11. Jitter (in ms) of Wireless (802.11) Networks

From Figure 10, we observe that packet size is not a major cause of variation of RTT. In the case of low contention, there is minor increase of RTT from 12 ms to 14 ms when the packet size increases from 64 to 1420 bytes. The contention is a major cause of variation of RTT as it increases from 13 ms (low contention) to 50 ms (high contention). It should also be noted that the wireless RTT of no contention is comparable to the wired RTT. The most important observation is from the jitter chart where jitter of wireless RTT is significantly higher than jitter of wired RTT. Even in the case of no contention, we observe significant jitter (about 15% of RTT) while there is no jitter in wired traffic. In the case of high contention, wireless jitter is as high as 50% of RTT. The significance of jitter chart supports our proposed solution that jitter is an effective measure to distinguish between wired and wireless traffic.

Our proposed solution to detect rogue APs does not require any wireless sensor. Instead, we propose to collect all TCP traffic to the backend router as illustrated in Figure 12.

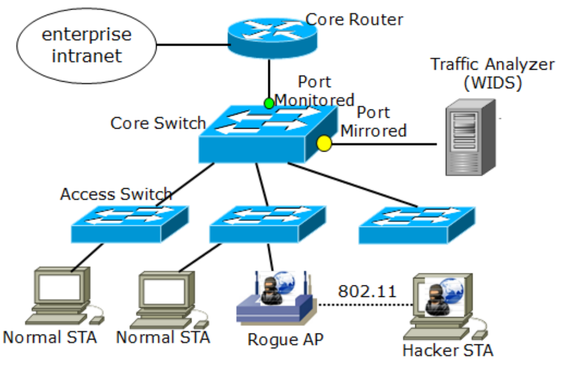


Figure 12. Network Configuration for Traffic monitoring

We already discussed the VLAN design (Figure 4) to separate wireless traffic from the wired traffic, and the use of security enclave to maximize the protection of wireless LAN. Figure 12 shows the network configuration to monitor the traffic and to check if there is wireless traffic on the wired VLAN. The steps of detecting rogue APs are summarized as follows:

1. Create TCP profiles of RTT and jitter for different packet sizes of 64, 128, 256, 512, 1024, 1280 and 1500 bytes.
2. Configure a mirroring port on the core Ethernet switch, and monitor all TCP traffic to the backend router (the default gateway). This configuration is referenced as Switch Port ANalysis (SPAN) for the Cisco switch configuration.
3. Compute RTT and jitter of each TCP traffic flow to and from the backend router.
4. Check if any traffic flow matches the wireless profile of TCP traffic.
5. If (4) is positive, WIDS detects a candidate of rogue AP and sends an alarm to the network administrator.
6. An optional step is to inject wireless traffic on the same channel. This is to create traffic contention. If the contention increases RTT and jitter of suspected traffic, it further confirms the traffic is from a Rogue AP.

The proposed solution works for all attacking scenarios, regardless of whether the AP is a layer-2 or layer-3 device. It is applicable to detect both rogue commercial APs and soft-APs. It also detects the non-AP case of wireless ad hoc configuration. A hacker may try different approaches, along with strongest security configuration on a rouge AP, to launch different attacking scenarios, but he/she cannot change the basic functions of CSMA/CA and TCP. Therefore, the proposed scheme could identify any Rogue AP attack.

Another advantage of our approach is the ease of removing rogue AP. When a rouge AP is detected, we can then check the MAC address tables on Ethernet switches to identify the physical port (interface) connected to the rogue AP. The WIDS can then use the MIB object of ifAdminStatus (1.3.6.1.1.1.2.2.1.7) and send an SNMP setRequest message to disable the Ethernet interface connected to the rogue AP:

```
snmpset –v2c –c community switch 1.3.6.1.2.1.2.2.1.7.x i 2
```

where *community* is the community string, *switch* is the IP address or host name of the Ethernet switch, and *x* is the interface index (ifIndex) of the physical port connected to the Rogue AP. The proposed solution of virtually removing rouge AP does not require to use any advanced technique to identify the physical location of the route AP [16], and it is more cost effective than dispatching security guards to physically remove rouge APs.

VII. CONCLUSION

This paper presents different attacking scenarios of rogue APs, and the common industry approach is to use wireless sensors to detect rogue APs. Our analysis shows that the use of wireless sensors, coupled with analysis of MAC addresses, is not effective as hackers can easily hide or change their MAC address behind rogue AP, soft AP, or Ad-hoc/STA. We propose a VLAN design to separate wired and wireless traffic and monitor TCP traffic on the wired LAN. Based on our analysis of TCP and wireless (802.11) protocols, we conclude that TCP-jitter on the wireless network is much larger than that on the wired network. We conducted many experiments on the ns2 simulation environment, and empirical data supports our theoretical analysis. We present a procedure to monitor network traffic on a backend router (a default gateway) and to collect and compute TCP RTT and jitter data. From the comparison of the profiling data, we can distinguish wireless from wired traffic. When wireless traffic is identified on the wired LAN, a rogue AP is detected.

REFERENCES

[1] IEEE 802.11-2012 Wireless LAN MAC and PHY Specifications. March 2012. It also includes 802.11i-2004 and 802.11w-2009.

[2] Chibiao Liu, James Yu, and Greg Brewster, "Empirical Studies and Queuing Modeling of Denial of Service Attacks against 802.11 WLANs," IEEE International Symposium on World of Wireless Mobile and Multimedia Networks (WoWMoM), Montreal, Canada, June 2010**.**

[3] “K. N. Gopinath and H. Chaskar, “All You Want to Know about WifI Rogue Access Points,” AirTight Networks, 2009.

[4] Liran Ma, A. Y. Teymorian, and Xiuzhen Cheng, “A Hybrid Rogue Access Point Protection Framework for Commodity Wi-Fi Networks, “ IEEE INFOCOM, pp. 1220-1228, April 2008.

[5] Control and Provisioning of Wireless Access Point (CAPWAP) Protocol Specifications, RFC5415, March 2009.

[6] K. G. Kyriakopoulos, F. J. Aparicio-Navarro, and D. J. Parish, “Detecting Misbehavior in WiFi Using Multi-Layer Metric Data Infusion,” Proceedings of International Workshop on Measurements and Networking, pp. 155-160, October 2013.

[7] Linux Host AP Driver, http://wireless.kernel.org/en/users/Documentation/hostapd

[8] David D. Coleman; David A. Westcott; Bryan E. Harkins; Shawn M. Jackman, “CWSP® Certified Wireless Security Professional Official: Study Guide,” Wiley Publisher, pp. 369-402, February 2010.

[9] “Rogue AP Detection under Unified Wireless Networks,” Cisco Document ID 70987

[10] Hongda Yin, Guanling Chen, and Jie Wang “Detecting Protected Layer-3 Rogue APs,” Proceedings of the Fourth IEEE International Conference on Broadband Communications, Networks, and Systems, Raleigh, NC. September 2007.

[11] V.S.S. Sriram, G. Sahoo, and K.K. Agrawal, “Detecting and eliminating Rogue Access Points in IEEE-802.11 WLAN - a multi-agent sourcing Methodology, “ IEEE 2nd International Advance Computing Conference (IACC), pp. 256-260, February 2010.

[12] Ke Gao, Cherita Corbett, and Raheem Beyah, “A passive approach to wireless device fingerprinting,” IEEE/IFIP International Conference on Dependable Systems & Networks (DSN), Chicago, IL, USA, June 2010.

[13] Sachin Shetty, Min Song, and Liran Ma, “Rogue Access Point Detection by Analyzing Network Traffic Characteristics,” IEEE MILCOM, October 2007.

[14] Hao Han, Bo Sheng, Chiu Chiang Tan, Qun Li, and Sanglu Lu “A Measurement Based Rogue AP Detection Scheme, “ IEEE INFOCOM, pp. 1593-1601, 2009.

[15] K. K. Leung, T. E. Klein, C. F. Mooney and M. Haner, “Methods to Improve TCP Throughput in Wireless Networks With High Delay Variability Proceedings of IEEE Veh. Tech. Conf., Los Angeles, CA, 3015-3019, September 2004.

[16] Tung M. Le, Ren Ping Liu, and Mark Hedley, “Rogue Access Point Detection and Localization,” IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communication, pp. 2489-2493, September 2012.

One-time Pad Cipher Based on Out-Key Distribution

Shengyuan Wu
Independent researcher, Ubit inventor, retired professor, Shandong University, Jinan, China

***Abstract** - This paper presents a one-time pad cipher based on out-key distribution. Key is divided into in-key and out-key; in-key is used in cipher and decipher; out-key is used in key distribution and in-key generation based on in-out number relation. Only out-key is transferred, in-key is not transferred. The relation between in-keys and out-keys is truly randomly characterized; that is out-keys doesn't contain any information of in-keys. So, it's impossible for Eve to crack the ciphered data by intercepting out-keys. In-out key combined with in-out password, in-out nonce can further strengthen computer security; in-out password makes password unbreakable; and in-out nonce makes replay attacks useless.*

Key words: Cryptography; key; password; nonce; cipher; one-time pad

1. Introduction

Computer security mainly consists of three aspects: confidential, authentication and integrity. Key, password and nonce are three kernel elements in computer security.

In cryptography, the one-time pad is the only unbreakable cryptosystem that exhibits what is referred to as perfect secrecy, and can be proven to be perfectly secure. [1-6].

However, this cryptosystem has a drawback: Since for each message a new key is needed, a large amount of random secret numbers have to be distributed between all parties that wish to communicate [1-3].

As we know, there are two ways to realize one-time pad:

First: distribute secret key by diplomatic suitcase, a physical method;

Second: distribute secret key by quantum cryptography.

“ Quantum cryptography” does not refer to quantum cryptosystems, but, somewhat misleadingly, to establishing a random secret key using quantum signals, i.e., implementing the one-time pad via quantum key distribution [1-3].

Theoretically, quantum key distribution is absolutely safe, one-time pad based on quantum key distribution is unbreakable and absolutely safe. However, the complexity and imperfectness of various parts in the quantum communication system makes it impossible to put broadly in use soon [2].

The difficulties stem from key, the key is used in cipher and decipher, the same key must be transferred; and the risk is just in key distribution.

Password is used in user authentication; password ideally should be easy to remember and hard to guess. Unfortunately these two goals are conflict with each other [6]; therefore, it isn't safe. The first line of security defense is password; however, weak and default passwords is a notable risk [7].

The difficulties of password is how to make password easy to remember and hard to guess.

Nonce is mainly used in against replay attacks, the most dangerous and the most difficult prevented attacks. The risk of nonce is the interception of nonce, because the same nonce must be transferred.

The weakness of key, password and nonce can be easily strengthened by in-out key, in-out password and in-out nonce based on in-out number relation.

This paper proposes to realize one-time pad cipher by three components: in-key is used in cipher and decipher; out-key is used for key distribution; only out-key needs to distribute, in-key is not transmitted.

The relation between in-keys and out-keys is truly randomly characterized, and in-out number relation is kept secret from Eve. Eve can intercept the ciphered data and the related out-key stream, but without the in-out number relation, she can't get the in-key stream by analyzing the out-key stream, so she can't crack the ciphered data.

This paper also proposes to realize user authentication by in-out password, out-password is easy to remember, and in-password is hard to guess.

This paper also proposes to against replay attacks by in-out nonce, challenge by out-nonce, and response by in-nonce.

The proposals can make key distribution and ciphered data absolutely safer, make password unbreakable; and make replay attacks useless.

2. ONE-TIME PAD CIPHER BASED ON OUT-KEY DISTRIBUTION

2.1 Review of one-time pad cipher

First, let us review one-time pad cipher, Vernam Cipher as showed in Fig. 1.

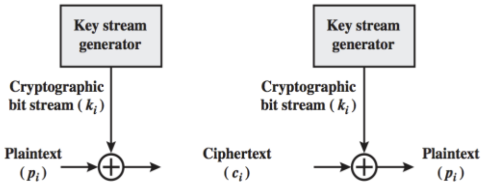


Figure 1. Vernam Cipher [5]

The system can be expressed as follows:

$$c_i = p_i \oplus k_i$$

$$p_i = c_i \oplus k_i$$

Where

p_i = ith binary digit of plaintext;

k_i = ith binary digit of key;

c_i = ith binary digit of ciphertext.

$\oplus$ = exclusive-or (XOR) operation

The essence of this technique is the means of construction of the key.

Vernam proposed the use of a running loop of tape that eventually repeated the key, so that in fact the system worked with a very long but repeating keyword [4].

Joseph Mauborgne proposed an improvement to the Vernam cipher that yields the ultimate in security. Mauborgne suggested using a random key that is as long as the message, so that the key need not be repeated. In addition, the key is to be used to encrypt and decrypt a single message, and then is discarded. The improved Vernam cipher is called as one-time pad, and is unbreakable [5-8].

However, Fig. 1 doesn't say how to generate cryptographic bit stream. Fig. 2 is a Vernam Cipher with key-controlled bit-stream generator, the two parties shares the key, and can generate the same cryptographic bit stream.

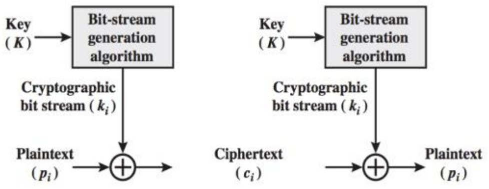

Key
(K)
Bit-stream generation algorithm
Cryptographic bit stream (k_i)
Plaintext (p_i)
Key
(K)
Bit-stream generation algorithm
Cryptographic bit stream (k_i)
Ciphertext (c_i)
Plaintext (p_i)

Figure 2. Vernam Cipher with key-controlled bit-stream generator [5]

However, the key stream must be provided to both users in advance via some independent and secure channel; this makes one-time pad cipher is hardly implemented [5]

Theoretically, one-time pad cipher based on quantum key distribution is absolutely safe; unfortunately, because the complexity and imperfectness of various parts in the quantum communication system makes it impossible to put broadly in use soon [2].

2.2 One-time pad cipher based on out-key distribution

Fig. 3 is a diagram of one-time pad cipher based on out-key distribution; here, key is divided into in-key and out-key.

The diagram relates to three kinds of ciphers.

The part below the short dash line relates to Vernam cipher as Fig. 1.

The part below the longer dash line relates to Vernam Cipher with key-controlled bit-stream generator as Fig. 2.

The whole diagram is one-time pad cipher based on out-key distribution. The extraordinary characteristic is that one-time pad cipher based on out-key distribution tells that two parties how to share keys, and how to generate the same cryptographic bit stream.

Here, only in-key is used to cipher and decipher; out-key is used in key distribution and controlling in-key generator.

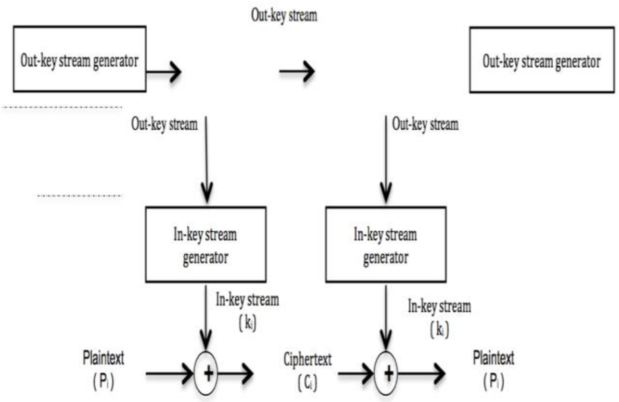


Figure 3. Diagram of one-time pad cipher based on out-key distribution.

Out-key stream generator is a random number generator; for Alice, it's used to control in-key stream generator to generate random in-key stream; out-key stream is also transferred to Bob, to control Bob's in-key stream generator to generate random in-key stream. Because Alice and Bob shares the same in-key stream generator, which is controlled by the same out-key stream; therefore, the in-key stream used in decipher is the same as in cipher.

The communication procedure is as following:

Alice generates out-key stream by out-key stream generator;

Alice generate in-key stream by in-key stream generator controlled by out-key stream;

Alice ciphers plaintext by in-key stream;

Alice sends ciphered text to Bob;

Alice also sends out-key stream to Bob;

Bob generates in-key stream by in-key stream generator controlled by Alice's out-key stream; Bob's in-key stream is the same as the Alice's in-key stream.

Bob deciphers the ciphered text by the in-key stream.

The difference between traditional one-time pad and one-time pad based on out-key distribution is as following:

In-key is used in cipher or decipher; but in-key is not transferred.

Only out-key is transferred.

Therefore, only out-key can be eavesdropped, but in-key is impossible to be eavesdropped, so in-key and ciphered text is absolutely safe.

However, if out-key stream contains any information about in-key stream; then Eve can know in-key stream by intercepting out-key stream, and crack the ciphered text.

2.3 In-key stream generator

In-out number relation is the core data structure in one-time pad cipher based on out-key distribution. Fig. 4 is a diagram of in-out number relation.

Out-number	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
In-number	8	F	3	B	2	C	3	4	1	6	2	E	5	A	9	7

Figure 4. Diagram of a simple in-out number relation

Out-number can be any length of binary digit, for example: 4 bit long, one byte long, two byte long, and so on. In Fig. 4, out-number is four bit long, represented in hex number; assume in-number is also hex number.

Usually out-number is arranged in order, ascending order as shown in Fig. 4; but in-number is set randomly

An in-key stream generated from an out-key stream is illustrated in the following simplified example.

Out-key stream generator generates an out-key stream,

Take each 4 bits from out-key stream, a hex number; assuming the out-key stream is 7846, 4 hex digits.

The first digit of out-key stream is “6”, search “6” in the out-number sequence in Fig. 4, find the related in-number is “3”;

The second digit of out-key stream is “4”, search “4” in Fig. 4, find the related in-number is “2”;

The third digit of out-key is “8”, search “8” in Fig. 4, find the related in-number is “1”;

The last digit of out-key is “7”, search “7” in Fig. 4, find the related in-number is “4”;

Then, in-key stream is 4123.

Similarly, for any random out-key stream, a random in-key stream can be generated based on in-out number relation.

Now, let’s discuss how to set up in-out number relation.

Out-number	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
In-number																

(a) At the beginning, in-number is empty

Out-number	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
In-number	8	F	3	B	2	C	3	4	1	6	2	E	5	A	9	7

(b) After in-number is filled

Figure 5. illustrating diagram of setting up in-out number relation

In Fig. 5 (a), out-number has been filled in ascending order; but in-number is empty, in-number can be filled randomly as the following:

Taking a16-sided die, labeled with the numbers 0 through F; each in-number is filled as the dice thrown or rolled, the die comes to rest showing on its upper surface a random integer from 0 to F, then the integer is filled in each column.

In-out number relation can be set by other method; but it must be kept to be random.

In-out number relation is random characterized, each out-number contains no information about related in-number, out-key stream consists of out-numbers, in-key stream consists of related in-numbers; therefore, the out-key stream contains no information about in-key stream.

Eve can intercept the ciphered data and the related out-key stream, but she can’t get the in-key stream by analyzing the out-key stream, so she can’t crack the ciphered data.

To keep ciphered message safe, in-out number relation must satisfy the three requirements:

1. For any random out-key stream, it can generate a related random in-key stream.

2. It can generate unlimited amount of in-key streams.

3. The relation between in-number and out-number is truly randomly characterized, so the relation of in-key stream and out-key stream is truly randomly characterized.

If out-key stream is generated by a true random number generator, then out-key stream is truly random, and the generated in-key stream is also truly random. A true random number generator can use a nondeterministic source to produce randomness, such as: by sound or video noise [5].

The relation between in-key stream and out-key stream is truly random, out-key stream bears no statistical relationship to in-key stream, and contains no information what so ever about the in-key stream.

Therefore, one-time pad based on out-key distribution is absolutely safe.

One-time pad based on out-key distribution can also be used in user authentication by sending an out-key to another party, and asking the party to send back the related in-key; the authentication fails if the party can’t send back the correct related in-key. This can efficiently against replay attack.

One-time pad based on out-key distribution can also be used to authenticate messages i.e. to identify their origin and integrity, and to identify user.

In practice, one-time pad based on out-key distribution integrates key, key distribution, encryption algorithm, data integrity, and user authentication as a whole.

2.4 The security of in-out number relation

In-out number relation is the core data structure in one-time pad cipher based on out-key distribution; its security depends on the security of in-out number relation.

First, like the master key distribution, an in-out number relation must be distributed in some noncryptographic way, such as physical delivery. However, the distribution of the in-out number relation is a bootstrap in-out number relation; because based on the in-out number relation, two parties can generates unlimited random keys; which can be used for one-time pad cipher; further, new in-out number relation can be safely transferred between the two parties by one-time pad cipher; therefore, the physically distribution of in-out number relation is similar to a bootstrap key in quantum key distribution for authentication. However, for mast key distribution, only a limited number of mast keys can be distributed one time; the limited master keys can't be used implementing one-time pad cipher; therefore, if new master keys needed, they still need to be distributed in some noncryptographic way, such as physical delivery.

Second, because in-out number relation is a bootstrap in-out number relation, the old in-out number relation can be replaced by a new one as needed; this strengthens the security of in-out number relation.

Third, in-out number relation is only kept by Bob and Alice, not transferred, it is impossible for Eve to intercept or to know it; this is different from master key, or session key; which must be transferred during key distribution although in encryption form.

Fourth, the in-number sequence and the out-number sequence of in-out number relation can be stored separately; for example in different media; this increases the difficulties if Eve tries physically or by malware to eavesdrop the in-out number relation.

2.5 A comparison between one-time pad based on out-key distribution and one-time pad based on quantum key distribution

1. For based on out-key distribution, in-key is not transmitted, and without the in-out number relation, Eve can't get the in-key stream by analyzing the out-key stream, absolutely safe; quantum encryption key must be distributed, theoretically, unbreakable and absolutely safe, but because of the complexity and imperfectness, vulnerabilities are inevitable;

2. Based on out-key distribution only uses classical channel; but quantum encryption uses both quantum and classical channels;

3. Based on out-key distribution integrates all five encryption elements, but quantum encryption doesn't;

4. Based on out-key distribution, the key management is simple and safe; however, quantum encryption key pool management is complicated;

5. Based on out-key distribution is implemented easily; however, quantum encryption theory and implementation is very complicated and difficult.

6. Based on quantum key distribution does not solve the key distribution problem without the need of a bootstrap key for authentication [1]. Neither based on out-key distribution, therefore, the first in-out number relation must be distributed physically.

In-out key can be combined with in-out password, in-out nonce in order to strengthen computer security.

3. IN-PASSWORD AND OUT-PASSWORD

Classical password, ideally should be easy to remember and hard to guess. Unfortunately these two goals are conflict with each other; therefore [6], it isn't safe.

The first line of security defense is password; however, weak and default passwords is a notable risk [7].

No one of the existed passwords can against all cracking methods today; for example, Finger print password is static, easily discovered, and easily attacked by replay; human body password, generated by password pill, is also breakable; therefore, personal, business's, and national confidential information eavesdropped and intercepted, huge money lost, moving vehicles in danger, even planted medical device in human's body can be cracked in danger.

However, based on in-out number relation, only owned by the user, password is divided into in-password and out-password, Out-password is short, simple and easily to remember, but in-password is long, extremely complex. Out-password is used to generate in-password; and in-password is used in authentication. For simplicity, assume out-password consists of digit 0-9 as shown in Table 1.

TABLE I. TABLE 1 IN-OUT NUMBER RELATION

Out-number	0	1	2	3	4	5	6	7	8	9
In-number	I_0	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9

Assume the in-number is not simple as in Fig. 5; and the related in-number is: I_1=3er78⌘, I_2=90 , I_3=45uip♪, I_4=w ⌫

Assuming out-password is 1234.

As input out-password: 1234; then, the related in-password is: 3er78⌘90 45uip♪w ⌫.

Here, the out-password is short, simple, and easy to remember; however, the in-password is long and complicated; the in-password above also include invisible characters; all these make dictionary attack, brute force attack, offline cracking (rainbow attack) useless.

Suppose Eve viewed the whole procedure of your inputting password; because out-password is only used in transforming to in-password based on the in-out number relation. Therefore, even Eve knows the out-password, for example here: 1234; she has no way to input your in-password without the in-out number relation. Key logger, screen scrubbing and shoulder surfing are all useless.

Assume an attacker asks your password by telephone posing as an IT security guy; at most, he can get your out-password, because you really don't know or can't remember your in-password.

4. IN-NONCE AND OUT-NONCE

Nonce is often used to against replay attacks. The use of random numbers for the nonces frustrates an opponent's efforts to determine or guess the nonce [5].

However, a lot of difficulties exist in applying nonce in computer security, for examples: the timestamp approach should not be used for connection-oriented applications because of the inherent difficulties with this technique; the challenge-response approach is unsuitable for a connectionless type of application [5].

The most important problem of nonce is the weakness of nonce. A nonce must be transferred in communication channel at least twice; Alice sends a nonce to Bob, Bob sends back the nonce to Alice. Eve can intercept the nonce to fool Bob, because the nonce is same.

The difficulties and the weakness problem can be easily solved by out-nonce and in-nonce by sharing in-out number relation by Alice and Bob. Out-nonce is used in challenge stage; in-nonce is used in response stage. Eve can intercept out-nonce, but it is useless in response stage, because without in-out number relation, she has no way to know the in-nonce; therefore this makes reply attack useless.

For example, two parties share in-out number relation as shown in Fig. 4, which is used to transform an out-number to related in-number; If Alice sends out-nonce=7846 to Bob, then Bob responses the in-nonce=4123. Eve can intercept the out-nonce, but she can’t response the correct in-nonce without the in-out number relation.

In-nonce and out-nonce can be used in user authentication. If client and server share an in-out number relation; mutual authentication can be done. For example, an attacker intercepted your in-password; later try to enter your bank account; the bank sends her an out-nonce and asks to send back the related in-nonce; because in-out number relation is only shared by the client and server, attacker can’t send back the related in-nonce without the in-out number relation. The client can also check if a web site is a fishing site by sends it an out-nonce, and asks an answer.

As two parties share in-out number relation; then replay attack is also useless, for example, an attacker intercepted in-password sent by car’s remoter controller, later tries to open the car using the in-password; however, the car can sends her an out-nonce and asks her sends back the related in-nonce, however she can’t; because only the car and the car owner’s remote controller share the in-out number relation [8].

5. CONCLUSION

In-out key, in-out password, in-out nonce greatly strengthens computer security. The methods presented in this paper are easily implemented.

Computer security can be further strengthened by Ubit theory [9]. In-out key, in-out password and in-out nonce combined with Ubit theory can lay a solid foundation of computer security.

Acknowledgments Thank Bin Wu, my wife, for her great contributions and fully supporting.

References

[1] Gilles van Assche, Quantum Cryptography and Secret-Key Distillation, Cambridge University press. 2006

[2] Yin Hao, Han Yang, The principles and technology of quantum communication, Electronic Industry Press, 2013

[3] W. Beiglb?ck, J. Ehlers, K. Hepp, H. Weidenmüller, Quantum Information, Computation and Cryptography: An Introductory Survey of Theory, Technology and Experiments (Lecture Notes in Physics), Springer, Berlin Heidelberg 2010, P 279

[4] Shannon, Claude, "Communication Theory of Secrecy Systems". Bell System Technical Journal 28 (4): 656–715, 1949.

[5] William Stallings, Cryptography and Network Security principles and Practice, Fifth Edition, Pearson Education, Inc., 2011

[6] Michael T. Goodrich, Roberro Tamassia, Introduction to Computer Security, Pearson Education, 2012

[7] 2013 Trustwave Global Security Report

[8] Shengyuan Wu，Methods and apparatuses of digital data processing, PCTIB2013060369, 11, 2013

[9] Shengyuan Wu，Introduction to Ubit Semantic Computing, Proceedings of The 2014 International Conference on Semantic Web and Web Services of Computer Science (SWW’14), 07, 2014

SESSION

MOBILE COMPUTING AND APPLICATIONS

Chair(s)

TBA

TTraffic: A Fully Integrated Accident Management System in Vehicular Networks Through Smartphones

Sevgi Kaya*, Burcu Gulfem Ozergin†, Halid Ozsoy† Taskin Kocak†, Cagri Gungor‡,
*Department of Computer Science, ETH Zurich, 8092 Zurich, Switzerland
†Department of Computer Engineering, Bahcesehir University, 34349 Istanbul, Turkey
‡Department of Computer Engineering, Abdullah Gul University, 38002, Kayseri, Turkey

***Abstract*—Recent years have witnessed new breakthroughs on emergency management systems and a reduced fatality rate caused by car accidents as a consequence of it. With the recent advances in wireless communication technologies, the trends in the automatic crash notification industry changed from individual reporting or infrastructure-based data sources to mobile data gathering devices such as smartphones. Our research primarily focuses on investigation of the most useful yet simple interfaces to develop an accurate and effective crash notification service based on detecting anomalities in the movement of the smartphone. For this purpose, a mobile application has been developed on both Android and iPhone platforms to detect the accidents and gather the relative data for determining the crash severity. In this paper, the proposed system architecture along with the detailed component descriptions and technical steps of the development phase are presented in detail. Furthermore, we introduce a web server application to process and illustrate the raw data transmitted from Android/iPhone mobile client applications. The paper also presents real life scenarios to demonstrate the underlying system architecture and its efficiency in emergency cases. We show that our fully-integrated emergency management system yields very promising results in the early detection of accidents.**

***Index Terms*—Real time services; Context aware service and applications; Information distribution services; Location-based services; Mobile web-based applications and services; Vehicular network applications and services**

I. Introduction

Over the last years, many research studies have been conducted to exploit the advances in sensing, communication and dynamic adaptive technologies for coping with the traffic related issues. In fact, one of the most crucial results of traffic congestion is the delay of the emergency services such as police, fire and rescue operations and medical services which may yield to the death of crash victims and huge amounts of financial loss. Every year, nearly 20,000 crash victims lose their lives before reaching to hospital[1]. According to the experts, the main reason behind these deaths is the failure of reaching the victim throughout the so-called “Golden Hour” where the treatment is most effective. Despite of the recent advances in wireless communications and technological infrastructure behind emergency management systems, crash response time is extremely lengthy to provide the most effective treatment for crash victims.

In the literature, there exist wide varieties of emergency management services which aim to reduce the crash response time as possible. These models can be classified as Pre-crash and Post-crash systems. While pre-crash models use radar, laser sensors and cameras to detect the damages caused directly by the crash, post-crash systems focus on eliminating the chances of crash injuries or fatalities due to the secondary effects of collision, such as fire. The role of emergency service models in this sense is beyond the foreseen. They play quite an important part in both determining the severity of an accident and reducing the crash response time by transmitting the crash information to concerned departments such as emergency, police and fire departments. The biggest impact of the emergency service models come from the fact that it provides the only possible way to help the crash victims and alleviate the results of a collision as effective as possible.

For many years, crash response systems relied on traffic reporters, static cameras and radars to detect the accidents on roads. Although individual reporting certainly helps to find out the crash place, it may take long hours to realize an accident especially in relatively unpopulated places. Hence with the recent advances in wireless communication technologies, the trends in the automatic crash notification industry changed from individual reporting or infrastructure-based data sources to mobile data gathering devices such as smartphones. The idea behind emerging emergency management systems is to deploy smartphones on vehicles to detect the accident via a mobile application which listens for accelerometer changes. Once the accident has been detected, the application automatically notifies the emergency medical personnel by transmitting the location information and a short video of the accident for determining the severity of it. Initial estimates reveal that the fatality rate can be reduced by 30% with the usage of recently emerged crash notification systems that rely on wireless communications.[2]

In this paper, we propose a mobile client application which gathers crash-related data and transmits them to a central web server for detecting accidents. In server side, we introduce our web application which detects crash location and notifies the emergency responders of accidents such as emergency service, police and fire departments in real time. Moreover, we describe the development phases of our emergency management system and evaluate its performance in different use case scenarios.

The rest of the paper is organized as follows. Section 2 describes related work. Section 3 presents a brief review of the system architecture. In Section 4, system components and development phases of the mobile application have been explained. Section 5 introduces the web server application and describes its key features. Section 6 illustrates the evaluation of our system by simulating different use cases. Finally, Section 7 concludes the paper by assessing the outcomes of our proposed system.

II. Related Work

In the literature, there exist wide varieties of emergency management services which aim to reduce the crash response time as possible. These models can be classified as Pre-crash and Post-crash systems. While pre-crash models use radar, laser sensors and cameras to detect the damages caused directly by the crash, post-crash systems focus on eliminating the chances of crash injuries or fatalities due to the secondary effects of collision, such as fire.

Based on this classification there are 5 different organizational categories related with our work. The first category consists of the solutions offered by top telecom companies such as China mobile, Vodafone[3-4], Verizon [5-6], etc. The second category is named as the solutions by municipalities. The third category is the solutions offered by corporations such as General Mobile[7], Toyota[8], Hyundai[9] etc. Fourth category includes the solutions developed by governments. Last category is the EU projects such as CVIS, AISE and HAVEit[10-12] which aims at providing comprehensive emergency management solutions. Contrary to the longevity experiences on this issue, most of the proposed systems do not offer fully-integrated practical and cost-effective solutions which can be directly applied to the vehicles in real time.

III. The System Overview

In this paper, we present an accurate and effective crash notification service based on detecting anomaly on the movement of the smartphone to identify the accident case. Fig.1 illustrates our proposed system architecture along with its components and key features. The system architecture consists of two key components: mobile client and web server application.

The mobile application has been developed on both Android and iPhone platforms to detect the accidents and gather the relative data for determining the crash severity. Once the crash has been detected, the system automatically sends accident related data to the web application. The web application processes the raw data transmitted from Android/iPhone mobile client applications and illustrates the vehicle coordinates and its associated crash video on a map along with its address information. This service is specifically designed for emergency management centers to provide instant notification and easy-to-access call mechanism to the nearest hospital, fire and police departments. The emergency management personnel can reach to user and alert concerned departments right after the crash.

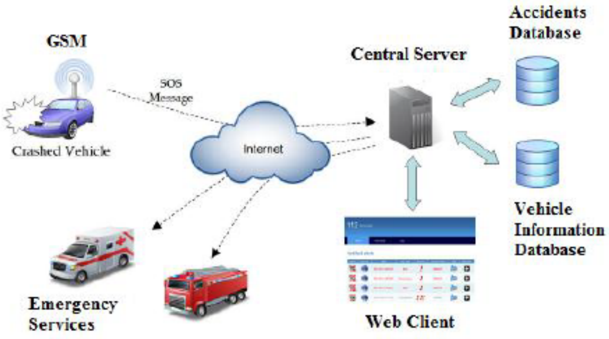


Figure 1: Overview of the system architecture

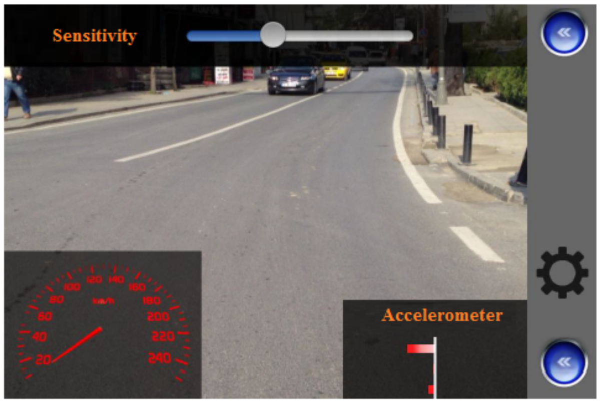


Figure 2: Overview of the system architecture

IV. The Mobile Client Application

Figure 2 illustrates the main screen of the mobile application. The application starts with the screen that shows the view of the camera. In this screen, there are some indicators regarding to accelerometer measurements. On the lower left, there is a speedometer that shows the speed of the vehicle. This data are collected using the GPS of the device. On the lower right, current G-Force of the device is shown for each axis. These values are received from the accelerometer of the device. On the top of the screen, there is a bar that holds the sensitivity slider. This sensitivity slider sets the lower limit of the sensitivity of the crash detection system. Higher slider values require more force to be applied in order to trigger the system. On the right side of the screen, there is a small control panel which consists of two control buttons. The buttons control the sensitivity bar by displaying or hiding it.

In Setting section, there are three options and their submenus to allow user to choose the communication type after the crash. The system automatically sends the accident location by either SMS or E-mail messages to concerned emergency departments depending on the choice of the user.

Usage When the application is started, crash detection system becomes activated. The camera immediately starts recording and the system tracks the G-Force measurements unless the user interrupts the operation. The maximum dura-

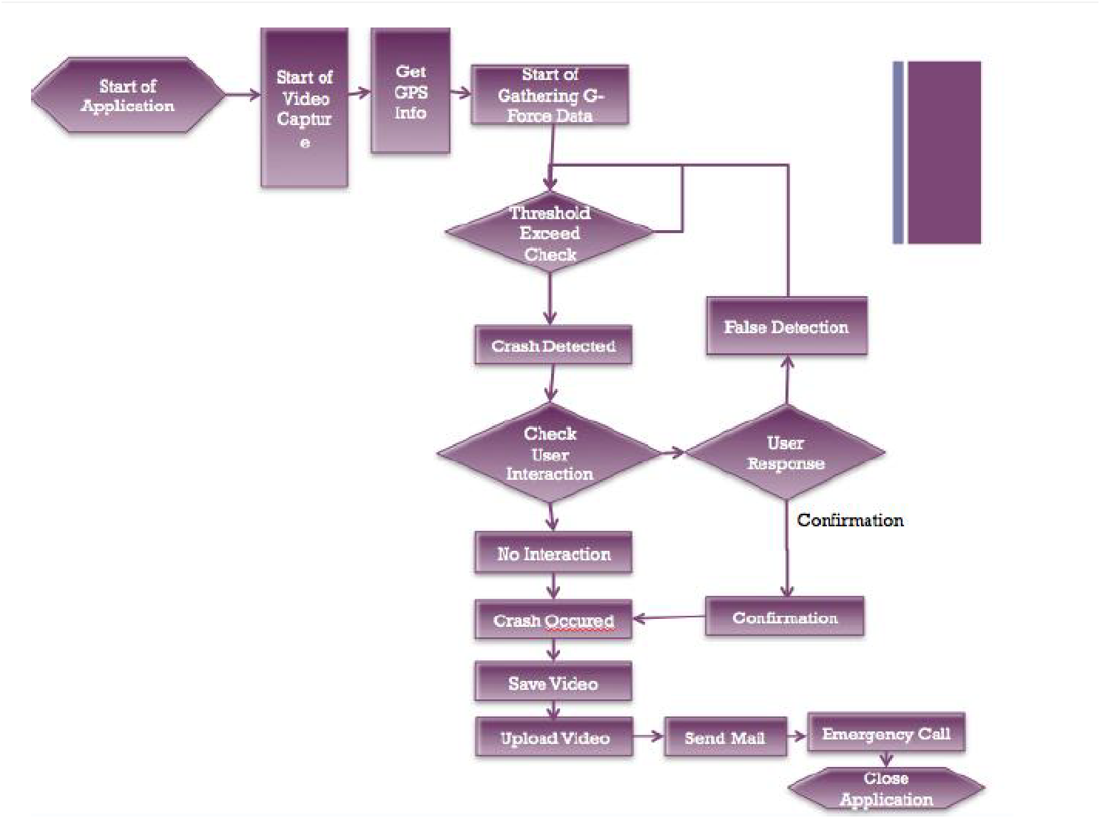


Figure 3: The flow chart of the data gathering algorithm

tion for an accident video can be easily set by the user via our user interface. After maximum duration is reached during a video record, recording stops. The video is not saved and released from the memory immediately. Instead, new record starts right after that. Once an anomaly has been detected, the system collects all of the video data along with the location information and sends the data to a central server.

A. *Crash Detection Algorithm*

The data flow chart of the algorithm is illustrated in Figure 3. As it has been stated in the previous section, the whole crash detection algorithm is based on the accelerometer measurements. Each time accelerometer value is received, the values of three axis (x,y,z) are sent to the function that checks if there is a abnormal situation. In this context, the abnormal situation refers to exceeding threshold values in at least one of the three axes.

If one of the values is over of the threshold value, the system is triggered. However, sudden changes in the accelerometer measurements is not solely depend on accidents. In order to avoid false positive data, an alert pops up to ask the users whether they want to cancel the data transmission operation or not. Including the user feedback to the application provides more reliable mechanism to avoid false positives. Figure 4 demonstrates the procedure in detail.

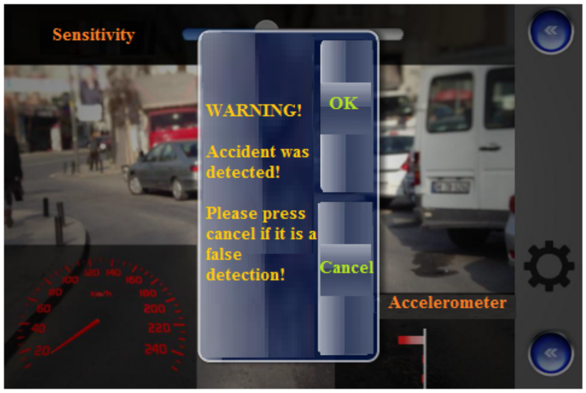
After user confirms that there is an accident and need help or unable to confirm it, post accident system is being activated. The camera stops capturing video. When video is available to take actions on it, internet connection of the device is checked. If the device has an internet connection, the taken video starts to be compressed and uploaded to the servers over an FTP connection. Before all that, emergency contacts receive an e-mail with the coordinate information and emergency message for the user. After the video is uploaded to the servers successfully, another e-mail is sent with a link of the uploaded video. If the device does not have an internet connection, the video is saved to the to device's photo album.

1) Video Uploads: The videos that have been captured during the accident are uploaded to servers through cellular network connection or Wi-Fi. This operation is completed in background. We set up this connection through FTP and the connection settings (hostname, username, password, upload location, filename) are done hard-coded. The video is converted into bytes and transferred over this connection.

2) Sending E-mail After Accident: For this application, e-mail sould be sent without user interaction. In E-Mail settings, there are two text fields: one for recievers and the other for the message. The coordinates and a short video of the accident are appended to the e-mail.

3) Sending SMS After Accident: Likewise sending E-Mail, sending SMS without user interaction is not possible. Hence, it should be handled via a web service. In the application, settings for SMS are completed for both contact number and text message.

4) Making Emergency Call: During the first time that the application has been installed, the users can set emergency



WARNING!

Accident was detected!

Please press cancel if it is a false detection!

Figure 4: Our mobile application allows users to cancel the sending of SMS message in 10 seconds

contacts which are automatically called right after the crash.

V. The Web Server Application

In this phase of the project, we developed a web application which illustrates the location of the crash along with its video retrieved from our Android/iPhone mobile client applications. Client application has an underlying service for sensing an anomay in accelerometer measurements. Once an anomaly has been detected, location information and a short video which consists of the few seconds of crash views are transmitted to server application via an e-mail service for further processing. In server side, incoming e-mails are parsed into components to reveal the location information and crash videos. Later on, the extracted data are saved to a back-end database and illustrated in a map via our web application. The web application is designed to notify emergency responders of accidents such as emergency service, police and fire departments by sending audio messages.

A. Motivation

The main reason that leads to the idea of a web application is to develop a framework to analyse and test our raw crash data retrieved from the test vehicles on the field that use our Android/iPhone application. The web application allows us to illustrate multiple crashes occurred simultaneously and alert concerned departments such as hospitals, fire and police departments by calling them through its interface. Hence we are able to find out the exact address information and determine the severity of the crash with its video uploaded to the server in real-time.

The web application is focused on achieving several aims as listed below:

1) Illustrate the map-view of the data collected by the Android/iPhone smartphone applications.
2) Generate random crash data for various setups and find out the address information associated with each coordinate.
3) Parse the crash e-mail which consists of the coordinates and a short video of the accident to extract the location information and to save the crash video into server.
4) Run the geocoder in the background to convert the raw coordinates to address information.
5) Provide an easy-to-access call mechanism to alert hospitals, fire and police departments in the case of emergency.

B. The Interface

Figure 5 depicts an accident management setup for a single crash alert. Each of the crash notifications that appear in the web application are denoted by a single marker and an information window associated with it. The marker in the figure pins the crash location. Information window is triggered when the marker is clicked and it illustrates the crash-related data to describe the crash in more detail.

On the upper-left corner, information window shows the cell phone number for calling the crash victim right after it has occurred. The crash video appears in the middle of the information window to determine the severity of the crash. Finally, a fully described street address is located on the lower-left corner of the information window.

Our web application provides a fully-featured calling system via Skype API. The emergency management personnel can alert the concerned departments about the crash by simply calling them via Skype. Lower-left corner of our web page has three buttons to activate the calling procedure. The buttons trigger an interface to notify emergency, police and fire departments respectively.

C. Use Case Scenario

The section presents a crash scenario using our web application. In our scenario, the purpose is to simulate an ordinary traffic crash and the emergency response associated with it. The scenario starts with an incoming e-mail sent by the crash victim via our mobile application.

The incoming e-mails are parsed and inserted into the database in the format illustrated in above. Each crash is associated with an automatically generated unique identifier number and latitude/longitude pair retrieved from the subject field of the incoming e-mail. Furthermore, the video file is instantly uploaded to the server computer and the file path is stored in the database under the caption of description attribute. Once the procedure is completed, the resulting crash information is illustrated in the web interface.

Our web application enables the emergency management personnel to call the concerned organizations such as emergency, police and fire departments via just clicking the buttons in the lower left corner. The buttons instantly connects with the already set Skype account and triggers a Skype interface to call the emergency personnel. Figure 6 demonstrates the triggered Skype interface right after clicking the blue button in the lower left corner.

VI. Conclusions

In this paper, we introduced a fully integrated emergency management system for traffic accidents. We then evaluated our model under various use case scenarios to determine its

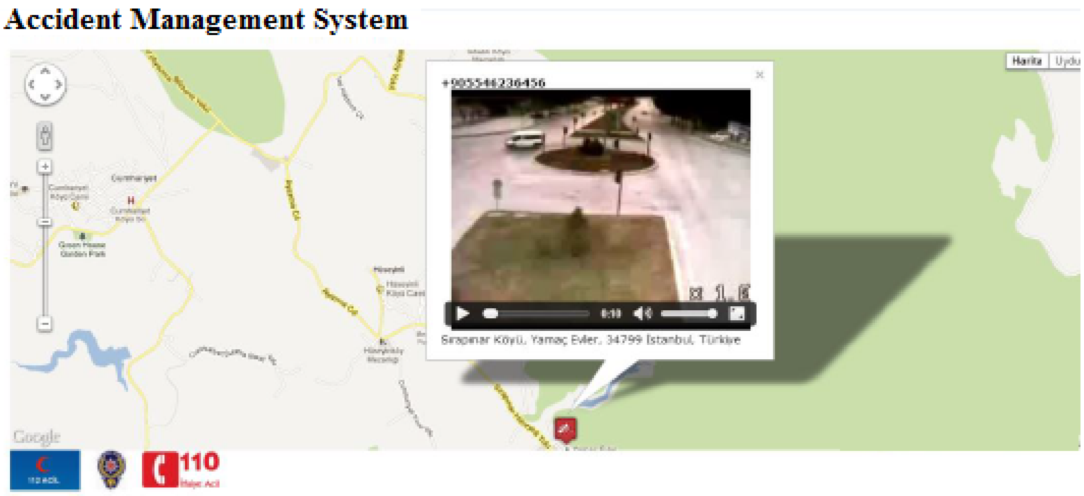

Accident Management System
+905546236456
Sırapınar Köyü, Yamaç Evler, 34799 İstanbul, Türkiye
Harita
Uydu
Google
110
Hızır Acil

Figure 5: The overview of the user interface

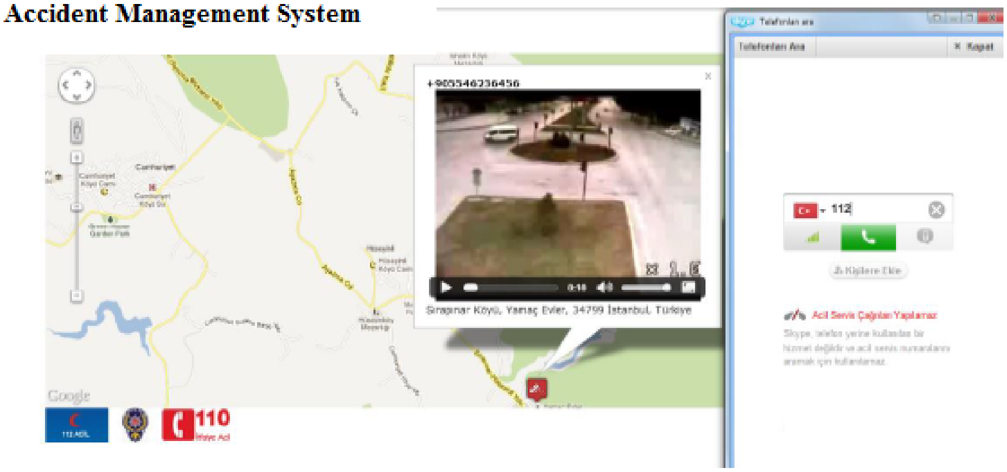

Accident Management System
+905546236456
Sırapınar Köyü, Yamaç Evler, 34799 İstanbul, Türkiye
Telefondan Ara
Kapat
112
Kişilere Ekle
Acil Servis Çağrıları Yapılamaz
Skype, telefon yerine kullanılabilir hizmet değildir ve acil servis numaralarını aramak için kullanılamaz.
Google
110
Hızır Acil

Figure 6: The triggered Skype interface for calling the emergency personnel

feasibility for near-future traffic flow prediction. The proposed system presents promising results in the early detection of accidents.

VII. Acknowledgements

This work was funded by the Turk Telekom Research & Development Center.

Project: Accident Detection and Management on Vehicular Networks

References

[1] Martinez, F.J., Chai-Keong Toh, Cano, J.-C., Calafate, C.T., Manzoni, P., "Emergency Services in Future Intelligent Transportation Systems Based on Vehicular Communication Networks," *Intelligent Transportation Systems Magazine, IEEE , vol.2, no.2, pp.6,20*, 2010

[2] "Japanese Government MLIT Solution", [Online] Available: http://www.mlit.go.jp/index_e.html, 2013

[3] "ITS Handbook Japan", [Online]. Available: http://www.mlit.go.jp/road/ITS/index/indexHBook.html

[4] "ITS US Forum Presentation", [Online]. Available: http://www.itsforum.gr.jp/Public/E4Meetings/P07/SS64%20wani.pdf, 2013

[5] 9]"AIDER Final Report", [Online]. Available: http://www.transport-research.info/Upload/Documents/201001/20100115_103651_51950 _AIDER _Final%20Report.pdf

[6] "ATLANTIC-A thematic long-term approach to networking for the telematics and ITS community" [Online]. Available: http://www.trg.soton.ac.uk/archive/its/atlantic.htm, 2013

[7] "Advanced telematics for enhancing the safety and comfort of motorcycle riders" [Online]. Available: http://www.saferider-eu.org/, 2013

[8] "Reliable application specific detection of road users with vehicle on-board sensors" [Online]. Available: http://www.adose-eu.org/, 2013

[9] "European Union ITS Project: COOPERS", [Online]. Available: http://www.cooper-eu.org, 2013

[10] H. Greenberg, "An Analysis of Traffic Flow", *International Journal on Operations Research and The Management Sciences,vol.7, pp.79-85* , 1959.

[11] "PREVENT OBU", [Online]. Available: http://www.sevecom.org/Deliverables/Sevecom_Deliverable_D2.1_v3.0.pdf, 2013

[12] R. Uzcategui, A. de Sucre, G. Acosta-Marum, "Wave: A tutorial-[topics in automotive networking]", *IEEE Communications Magazine, vol. 47, no. 5, pp. 126-133*, May 2009.

Self-Handover Optimization in NEMO Networks for LTE-Advanced Congestion Control Purposes

H. Mzoughi[1], **F. Zarai**[1], **and L. Kamoun**[1]
[1]LETI laboratory, University of Sfax, Tunisia

Abstract - *Radio system evolution is moving towards total convergence to achieve: multi-services, multi-standards, multi-bands, reconfigurable and reprogrammable systems. Such systems are based on cognitive radio technology, by giving more intelligence to radio system devices, in order to facilitate the administration and to ensure the network self-optimization. Several protocols are used such as Network Mobility, which provides a continuous connectivity for users when moving together. Handover process has become a key focal point in the network management these days. In the present work, we focus on handover optimization in NEMO networks, with congestion control consideration in 3GPP LTE-ADVANCED cells. We employ network capacity as key metric to manage the handover procedure. We deal with the target network selection phase in the handover process. The simulation results reveal that our proposed scheme ensures a low packet loss rate, as well as handover call blocking probability.*

Keywords- NEMO; MIH; LTE-A; Congestion Control; Simulation Analysis.

1 Introduction

The trend towards convergence of networks, services and terminal devices in the telecommunication area requires a high mobility between existing technologies, provided that it does not deteriorate the Quality of Service (QoS). That compromise, has led to several studies examining user's mobility in various scenarios. The LTE-Advanced network (LTE-A) [1], which presents the fourth generation network, illustrates the convergence concept. In this paper, we deal with the congestion control in the LTE_A network, and more especially in the case of Network Mobility (NEMO) [2]. NEMO protocol has been introduced by the Internet Engineering Task Force **(**IETF), as an extension of Mobile IPv6 [3] protocol, to support the mobility of moving network. In such case, a moving group of Mobile Nodes (MN) is directly connected to a mobile network router called Mobile Router (MR). In one NEMO network, there is one or more MR, and only one of them is directly connected to the serving access router (AR) in the backbone IP, which called default or primary MR (PMR). The PMR takes over the role of all MNs behind it of interfacing with the core network, chiefly when performing the mobility management process. NEMO solves many problems of Mobile IP (MIP) to support a moving group of users, and it can be used to allow the customers to stay connected to the Internet without service disruption when traveling by vehicles (airplanes, trains, ships or cars).However, it may introduce some problems in the target network when moving from one network to another with all user resources requirements. In such a case, the target network may be unable to serve all users; or by serving them, it becomes in the congestion state.

This paper is organized as follow. Section II shows existing congestion control schemes and an overview of the adopted technologies in our work. Then, we introduce in section III our proposed scheme. Finally simulation results are presented and discussed in section IV, which is followed by a brief conclusion.

2 Background

2.1 Related work

In the literature, various methods for congestion control in new generation networks have been proposed. Many of the previous related works focused on the congestion problems including these that are associated with TCP congestion detection and avoidance mechanism [4-8]. But, TCP mechanism is not suitable for wireless networks, as it consumes a lot of resources and it shows lack of flexibility to satisfy devices in heterogeneous environments like 4G-LTE-A networks .A large number of methods have been devised to tackle this issue. These methods are deployed by different layers of the OSI stack. Several techniques are presented and discussed in [9] and [10]. In [11], a congestion control scheme for LTE system is proposed, where, the authors adopt the system load to manage congestion. Moreover, in [12] they proposed a user priority-based congestion control algorithm for cross-traffic assistance in LTE networks. The network load is always used as criteria of congestion detection. Then, in order to detect the congestion state, low priority users will be disconnected in order to serve high priority users. This solution is aggressive for low priority users. In addition, in both [11] and [12], only network load has been considered. In the case of wireless systems, interference is a key parameter, which affects system capacity as well as provided QoS. In this paper, we attempt to define a new scheme of congestion control that is more responsive to the needs of LTE-A systems. We integrate the congestion control with the mobility management in case of NEMO networks.

2.2 Media Independent Handover (MIH)

Media Independent handover (MIH) [13] is an IEEE 802.21 standard, which provides intelligence to the link layer by giving more information about neighbors' networks for upper layers, in order to optimize handover between heterogeneous technologies. Several network components are considered in MIH frameworks, which are briefly described below:

- Mobile Node (MN): This is the customer's mobile equipment attached to the network.

- Point Of Attachment (PoA): It is the network end point that the MN is directly connected to, like the eNodeB in LTE-A network.
- Point Of service (PoS): This network entity provides dynamic information to the MN useful for handovers process.
- Non-PoS: This entity contains network static information like network topologies.

These entities, like any MIH framework device, have to be active MIH entities by implementing a protocol stack including the MIH Function (MIHF) sub-layer between the second layer and the upper one in the protocol stack. Thus, MIHF handles the communication exchange between all network devices, as well as communication between lower layer entities and higher layer entities in a local device. As presented in Figure1, different MIH Service Access Points (SAPs) interfaces are defined to ensure information exchange under MIH, through a large set of MIH primitives for control and collecting information.

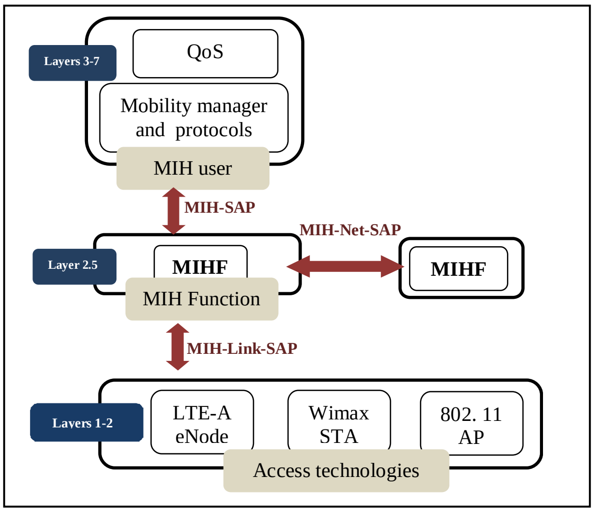


Figure 1. Media Independent Handover Protocol Stack

These primitives are classified into three categories: Event, Command and Information Services.

- Media Independent Event Service (MIES): MIH events primitives pertain to the physical and logical link layers factors. And, are also used to predict state change of these layers. MIH events are generated by lower layers or MIHF either at the MN or at a network node, and are intended for upper layer at local or remote level.
- Media Independent Command Service (MICS): MIH commands (MICs) are initiated by the high layer or by MIHF entity. And they are delivered locally for MIHF entity or lower layer. MICs are used to control lower layers, and are classified by two main categories: (1) Link commands: used by MIHF entity, on behalf of the MIH Users, for links status control operations. (2) And, MIH Commands which are sent by the higher layers to the MIHF entity.
- Media Independent Information Service (MIIS): The MIIS provides a framework of collecting information about existing networks and operators. The MIHF entity in the MN side benefits from this service when making handover decision as well as at for selecting the target network in order to make a successful handover between heterogeneous technologies. The MIHF entity in the MN exchanges MIH information with a PoS entity in the network side.

3 That proposed scheme

In our work, we focus on optimizing the target access router (AR) selection by the MR. We aim to prevent the congestion state in LTE cells, and way required QoS can be maintained. The movement of a mobile network means the movement of the PMR with all MNNs behind it, as presented in Figure 2. When performing the handover during the movement of NEMO network, MRs keeps this movement transparent to all MNNs [14, 15].

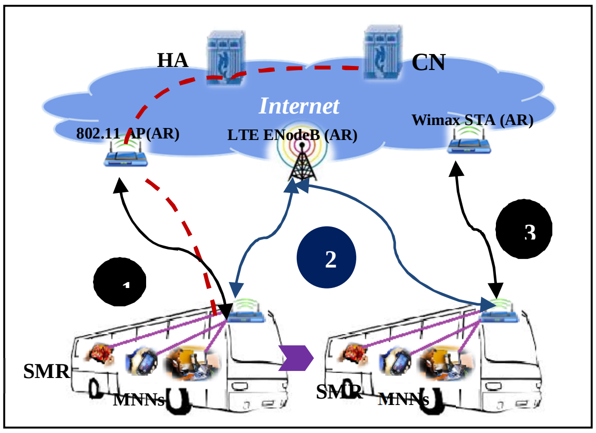


Figure 2.NEMO Architecture

When moving, MR should observe the active link state. In case of quality degradation, a handover process is triggered by another AR. First, PMR needs to discover ARs of existing technologies, and one of them will be selected as target AR. Here, target network should be able to serve all MNNs in the NEMO network. Otherwise, it risks that target network will be congested.

The proposed algorithm for target AR selection and its operation is detailed in the following subsections.

3.1 Target AR selection algorithm

When the MR detects quality degradation of the active link, it needs to change the current AR. The link degradation may occur when the MR becomes far away from its serving AR, or in case of a high level of interference. For selecting a target AR, MR begins with discovering phase to form a list of candidate ARs (CARs). Then, we adopt three levels for the target AR selection:

- Level1: This is based on the Received Signal Strength(RSS) from candidates ARs by the MR.
- Level 2: It is based on the congestion factor proposed parameter, explained below.
- Level3: This is based on the cell available capacity.

In the first step, the MR selects the AR with the highest RSS, then it verifies the cell load level based on congestion factor (C_f) parameter. Since, we work on the congestion control in 3GPP LTE network, we complete the case when the CAR is an eNodeB. Here the eNodeB should estimate its $\boldsymbol{C_f}$ as below:

$$C_f = e^{-\frac{RB_{av}}{RB_T}} \tag{1}$$

Where:RB_{av} is the number of available physical resources blocks (PRBs), and RB_T is the total number of PRBs offered by the cell.

The $\boldsymbol{C_f}$ factor indicates the occupancy rate of the cell. Once the cell is slightly occupied, this eNodeB will be selected as target one and a HO will be triggered. Otherwise, if the cell is mildly loaded, the MR asks the eNodeB for the maximum capacity that it can handle, and compares it with the average data rate required by all MNNs behind it. The available capacity ($\boldsymbol{C_{av}}$) in the LTE cell is calculated using the Shannon formula by the selected eNodeB:

$$C_{av} = B_{av} log_2(1 + SINR_{eNB-MR}) \tag{2}$$

Where: $\boldsymbol{B_{av}}$ is the bandwidth corresponding to all the available PRBs, $\boldsymbol{SINR_{eNB-MR}}$ is the SINR of the MR, which is given by:

$$SINR_{eNB-MR} = \frac{S_{eNB-MR}}{ICI+N_0} \tag{3}$$

Here, $\boldsymbol{S_{eNB\text{-}MR}}$ is the RSS from the eNodeB to the MR, *ICI* is the inter-cell interference indicator and $N_0$is the additive white Gaussian noise.
On the level of MR, the average of MNNs required data rate (D_R) for a period of time δt, is simply calculated as:

$$D_R = \frac{\sum_{active\ MNNs} T_r}{\delta t} \tag{4}$$

Where: T_r is the amount of received traffic in bit from MNNs to the MR during δ*t*.

If the $\boldsymbol{C_{av}}$is greater than D_R, this eNodeB will be finally selected as target AR and MR passes to the execution phase of the HO process. According to the $\boldsymbol{C_f}$ factor, if the LTE cell is fully loaded, this eNodeB will be removed from the list of CARs and MR goes back to reselect the AR with highest RSS in the updated CAR list. For other existing RANs, the execution of their admission control scheme gives the answer if the HO request is accepted or rejected.

The proposed target AR selection algorithm is summarized below.

Target AR selection algorithm

```
1.  If current_RSS< RSS-threshold or Cf~=1
{
      //Discover existing ARs
2.    If CAR-RSS > RSS-threshold
          Add AR to CAR list
      End
3.    While HO_dec==0 and CAR list not empty
      {
         //Select CAR with highest RSS from list
         If selected CAR=eNodeB
              //ask congestion measurements from the
               eNodeB
              If $C_f$< 1/2
                  HO-dec=1,
                  Target AR=CAR,
             Elseif $C_f$~=1
                 Remove this AR from C-AR list
             Else
                Ask eNodeB for available capacity
                If available capacity> ARR_MR
                    HO_dec=1,
                   Target AR=CAR.
               End
             End
         Else
           Target AR=CAR,
           HO-dec= RAN-Admission Control,
         End
      }
4.  If  HO_dec==0
      NEMO reconfiguration to multi-homed NEMO
            with single home agent.
      For each MR
            Go to step 3
       End for
    End if
}
```

Step 4 in the algorithm resolves the case when the PMR cannot find a target AR which can handle all MNNs required resources. In such a case, NEMO will be reconfigured to obtain a multi-homed NEMO with two PMR and single home agent, in order to reduce the requirement capacity. In such a configuration, each PMR will have only one route to the Internet through its own bi-directional tunnel.

Here, old PMR should select the best MR to perform this role. For that some factors are considered such as:

- Load balancing between MRs.
- MRs Localization.
- Devices' capacities.

Then PMR advertises both the selected MR and the current AR in order to establish the link between them. Each one of the two PMR will now select different target AR and will execute the HO independently. The old PMR gathers the two parts as soon as possible and when it is current cell is able to support the whole NEMO network. So, the PMR is kept informed about cell capacity state. This scheme can be applied also for nested NEMO.

In next sub-section, we detailed the exchanged messages between different network devices under MIH environment in order to perform target AR selection algorithm.

3.2 Proposed Target AR selection process

Using MIH protocol, the information exchange between the network devices in heterogeneous environment becomes significantly easier. The 4G network provides high mobility and dynamicity. It is, therefore necessary to make Target AR selection with the most recent gathered criteria. Hence, MR asks for the RAN measurements directly from the discovered CARs. The used MIH services primitives are listed in Table 1 with their corresponding parameters.

Table 1. Used MIH Services in the Proposed Scheme

Primitive	Service	Parameters
MIH_link_get_parameter	MICS	MR Id or SAR Id, RSS, Bit Error Rate (BER).
MIH_linkgowing down	MIES	MR Id, SAR Id, Time Interval, Link-Going DownReason .
MIH_link_detect	MIES	CAR Id, detected link information
MIH_MN_congest_measurment *(NEW)*	MICS	CAR Id, congestion factor; available capacity.
MIH__link_measurment_report	MIES	MR Id, local and remote collected information

The operating entities in the target AR selection process exchange required information by messages and MIH primitives. The diagram presented in Figure 3 shows the progress of the target AR selection procedure.

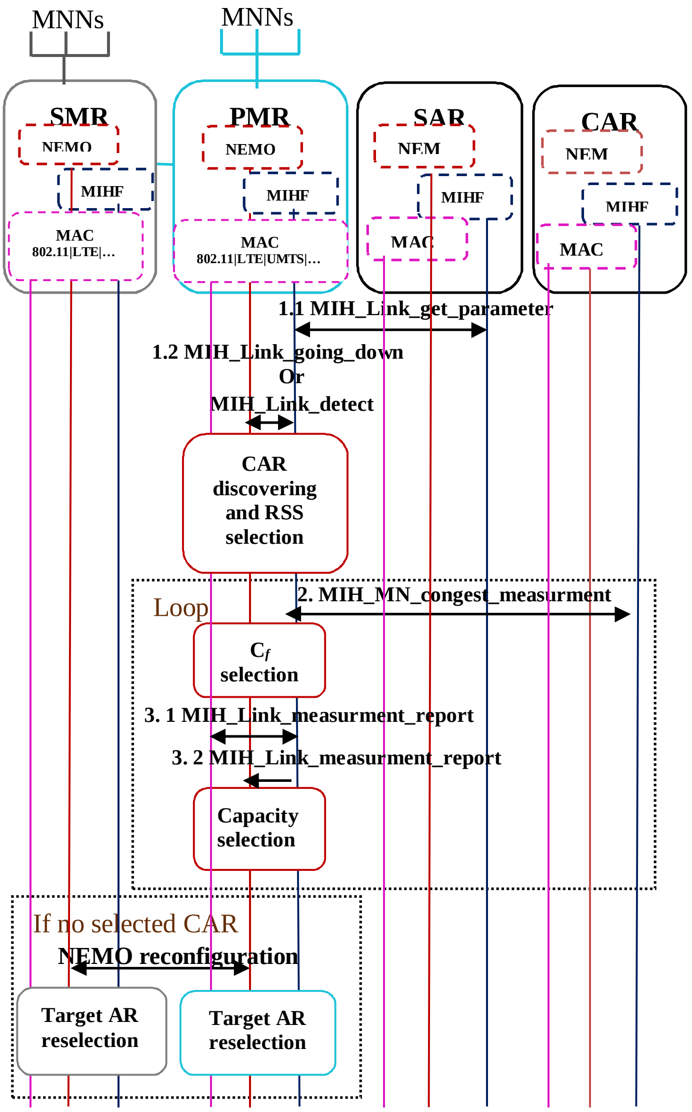


Figure 3. Proposed Target AR Selection Process

4 Performance evaluation

This section will describe the simulation model developed for analyzing the performance of our proposed scheme. The implementation is performed using MATLAB, since existing simulators do not provide appropriate support to simulate new generations of wireless networks, whilst considering the heterogeneity of such environment. We adopted the Mont-Carlo method for the simulation proceeding [16-17].

4.1 Simulation parameters

In order to evaluate the performance of our proposed scheme, we implemented the lower layers of 3GPP LTE-A network, as well as the NEMO network. We also considered the heterogeneity of 4G environment by adding WiMAX and WiFi access points. Like described above, each time an eNodeB is preselected as target AR based on RSS parameter, our scheme will be executed.

Parameters used in our simulation model are shown in Table 2.

Table 2. Simulation Parameter Setting

Parameters	Values
LTE-A bandwidth	20 MHz (50 RB et 180KHz per RB)
Cell radius	1500m
Distance-dependent path loss	PL=k/d$^{\alpha}$; k=1.4×10^{-4} and α=3.5
Heigh of eNB	32m
Heigh of UE	1.5m
Shadow fading	Log-normal 9db standard deviation
UE distribution	Uniform randomly
Total eNodeBTx Power	46 dBm
UE power	24 dBm
Thermal noise density	-174 dBm/Hz
Link adaptation	ACM Modulation: 64QAM
Bandwidth requirement	VoLTE : 64kbps Data :512kbps
number of users per cell	[100,600]
number of users in the NEMO networks	[50,350]
Vehicular velocity	20 m/s

Call arrivals are modeled by a Poisson process. Simulation experiments are run for various number of MN in the LTE cell. A MN can take the role of a MR, so in such case it hides behind it a various number of MNNs to form a NEMO network within.

4.2 Simulation results

In this sub-section, we present the simulation results of our proposed schemes. The purpose, of this work is firstly to give MR the ability to prevent congestion in the LTE target cell. Hence, we chose the following performance evaluation metrics:

- Rate of packet loss due to congestion state in LTE cell, which is calculated as follow:
Rate of packet loss= Lost packets/ total number of packets.

- Rate of blocked handover: This is the rate of blocked handover when the PMR can't find a target AR which can accept the whole mobile network:
Rate of handover blocking= Blocked handover requests/ Total number of vertical handover requests.

- Rate of LTE used resources: This rate shows the level of resources allocation in the LTE-A cell and is equal to:
Rate of LTE used resources = Used resources/Total network resources.

Figure 4 shows the variation of packet loss rate caused by a congestion state in LTE-A cell, according to a various number of mobile nodes (MN) initially located in the 3GPP LTE cell. The rate of packet loss is between 0.25% and 0.48% which is considered a low value. The effect of the increase of the number of users, either in the LTE-A cell or the mobile network, on the lost packet rate is relatively small even it follow this increase. In fact, the purpose of our scheme is to prevent as much as possible the congestion state in the LTE-A cell. According to this result, we can confirm that the loss of packets due to congestion state is resolved by our congestion control mechanism, and so network QoS is enhanced.

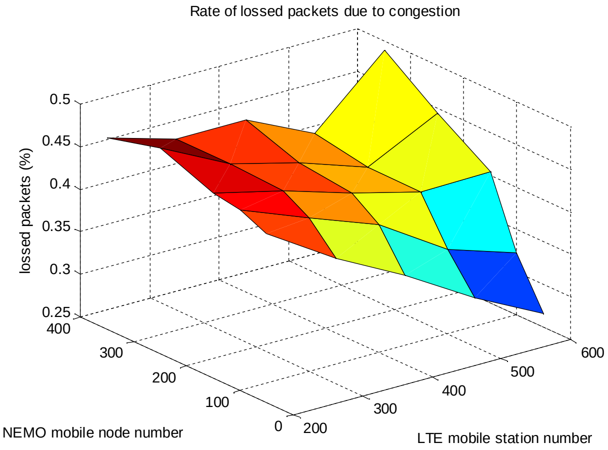


Figure 4. Rate of Packet Loss in LTE-A

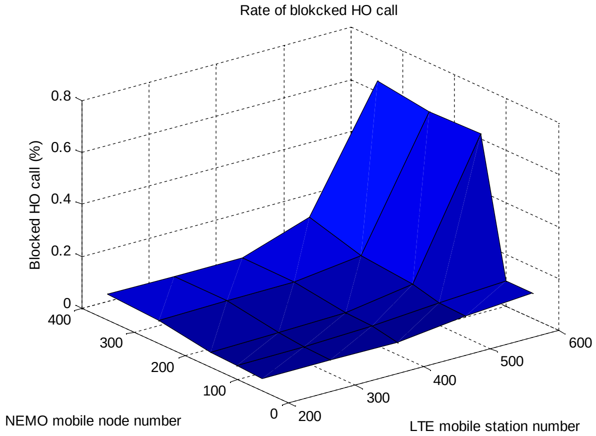


Figure 5.Rate of Handover Call Blocked

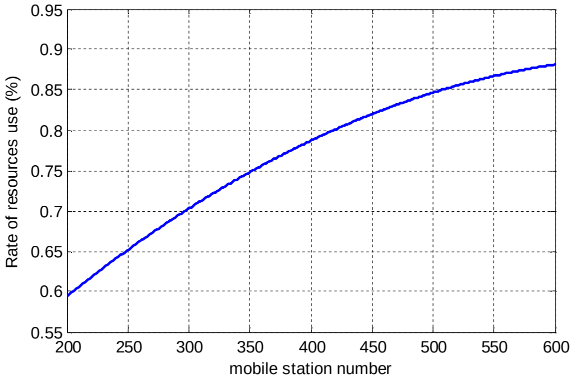


Figure 6.Rate of Resources' Utilization in LTE-A cell

Moreover, we observe the rate of blocked handover calls to LTE selected cells, when executing our proposed solution. When congestion is detected in the cell, or if the selected target cell for a handover cannot satisfy visitor user's requirements, current handover request will be rejected. The handover block rate is plotted in Figure5 for various number of mobile nodes in the LTE-A cells.

As seen in Figure5, the rate of blocked handover is significantly reduced. Until 600 MNs in the LTE-cell and 350 MNNs in the NEMO network the rate of blocked HO does not exceed 0.65%.So, it can be seen that our proposed solution improves the decision making for a handover. It also optimizes the target AR selection process to avoid handover interruption.

Another challenging network performance issue is the rate of resources utilization, which is illustrated in Figure 6.
The rate of resources utilization in LTE-A cell is plotted according to various number of MNs in the cell and 250 MNNs in the NEMO network. Figures 6 shows a high level of utilization rate, achieving 83% for 600 MNs in the cell. Indeed, the proposed congestion control method considers the network capacity and application requirements, for monitoring the process of regulating the total amount of data entering the network in order to keep traffic levels at an acceptable value according to the provided bandwidth.

5 Conclusion

The congestion problem in communications networks including 3GPP LTE networks has been around for a long time and always will. A large number of solutions has been proposed in literature. In this paper, some existing solutions are discussed. An effective congestion control scheme coupled with handover process is proposed. Our proposed solution defines a congestion prevention scheme in 4G-LTE systems. We adopt the case of NEMO networks. The mechanism is executed by both the MR in NEMO networks and the eNodeB. When the MR detects that current cell is approaching a congestion state, it makes a handover decision in order to resolve congestion. And, when an eNodeB is preselected as target AR, the MR should check whether this cell is able to satisfy the whole NEMO requirements or not. We used the LTE-A system capacity as congestion management criteria. Simulation results show that our proposed scheme reduces the rate of packet loss, as well as the rate of handover blocked calls. Moreover, MR is now able to maintain required QoS by their MNNs, via managing congestion state in current and target network. Although we adopt a specific case; our scheme can be generalized; thanks to the use of MIH protocol.

References

[1] 3GPP, RP-090665, “Revised SID on LTE-Advanced,” May 2009.

[2] V. Devarapalli, R Wakikawa, A. Petrescu, and P. Thubert, “Network Mobility Basic Support Protocol”, IETF RFC3963, January 2005.

[3] D. Jhonson, C. Perkins and J. Arkko, "Mobility Support in IPV6", IETF RFC3775, June 2004.

[4] S. Vangala, M.A. Labrator,“The TCP SACK-aware snoop protocol for TCP over Wireless networks”, IEEE 58th Vechicular Technology Conference, pp. 2624 - 2628. October 2003.

[5] A. Khurshid, M.H Kabir and M.A.T Prodhan,“An improved TCP congestion control algorithm for wireless networks”, IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim), pp. 382-387. August 2011.

[6] H. Zhang, X. Zhang, B. Fan and L. Shao, “Adaptive FAST TCP”, Second International Conference on Future Network, pp. 114-118. January 2010.

[7] H. Jamal, and K. Sultan, “Performance Analysis of TCP Congestion Control Algorithms”. International journal of computers and communications, Volume 2, No.1, 2008.

[8] J. Wang, J. Wen, J. Zhang and Y. Han, “TCP-FIT — A novel TCP congestion control algorithm for wireless networks”, IEEE GLOBECOM Workshops (GC Wkshps), pp. 2065-2069, December 2010.

[9] V. Kavitha and M. Muthuselvi, “A survey on congestion control techniques in Wireless Sensor Networks”, International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT), pp. 1146-1149. March 2011.

[10] J. Zhao,L. Wang, S. Li, X. Liu, Z. Yuan and Z. Gao, “A Survey of Congestion Control Mechanisms in Wireless Sensor Networks”, 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), pp. 719-722.October 2010.

[11] R. Kwan, R. Arnott, R. Trivisonno and M. Kubota, “On Pre-Emption and Congestion Control for LTE Systems”, IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall), pp. 1-5. September 2010.

[12] T. Lung-Chih, L. You and G. Mario, “Priority-Based Congestion Control Algorithm for Cross-Traffic Assistance on LTE Networks”,the 78th IEEE Vehicular Technology Conference (VTC Fall), pp. 1-5. September 2013.

[13] IEEE Standard for Local and metropolitan area networks- Part 21: “Media Independent Handover”. January 2009

[14] T. Ernst, "Network Mobility Support Goals and Requirements", IETF RFC 4886, July 2007.

[15] K. Leung, G. Dommety, V. Narayanan and A. Petrescu,”Network Mobility (NEMO) Extensions for Mobile IPv4”, IFTF RFC5177. April 2008.

[16] M.Guizani, A. Rayes, B. Khan and A. Al-Fuqaha, “Network Modeling and Simulation: A Practical Perspective”, John Wiley & Sons, January 2010.

[17] M. S. Obaidat and N. Boudriga,” Fundamentals of Performance Evaluation of Computer and Telecommunications Systems,” Wiley, 2010.

Application of Autonomous Mobile Agents to Provide Security in Wireless Networks

[1]Odhiambo MO, [2]Aderemi Lawal
Department of Electrical and Mining Engineering, University of South Africa, Pretoria, South Africa
[1]ohangmo@unisa.ac.za, [2]remilaw@gmail.com

Abstract - *The designs of wireless networks are constantly challenged by the ever changing network configurations due to node mobility in addition to propagation impairments and interference in wireless channels. In order to meet these challenges, networking solutions should incorporate distributed intelligence that enables network nodes to autonomously adapt to changes in networking environments and network configurations. By propagating mobile agents (software codes) to mobile nodes for execution and allowing them to spawn new agents for propagation to other nodes, mobile agents can provide an effective solution for these challenges. The existing wireless network management schemes follow a centralized approach and the process of data gathering and analysis usually involves huge transfer of management data. Consequently, this generates congestion in the area around management stations and it causes lack of scalability. There is the need to have a model for distributed and intelligent network management. Additionally, wireless connections can be lost or degraded by mobility. The mobile agent technology offers very promising solution to this problem. Mobile agents can analyze data and make decisions in order to preserve the reliability and quality of service in the wireless network.*

Keywords: mobile and intelligent agents, mobile networks, load balancing, congestion.

1 Introduction

Mobile agents are computer program or software components characterized by autonomy (ability to act on their own), reactiveness (to process external events), pro-activeness (to reach goals), co-operation (to efficiently and effectively solve tasks), adaptation (to learn by experience) and mobility (migration to new places). This paper describes the applications of autonomous mobile agents in distributed network management system for improving reliability and quality of service (QoS). This research proposes to evaluate autonomous mobile agents as one enabling technology for active wireless/mobile network management, we highlight some key properties of mobile agents which could be important for the design of such networks, particularly with regard to the anticipated heterogeneity in terms of radio technologies and applications and discuss their potential advantages for a wireless network scenario. The proposed autonomous mobile agent technology is yet to be implemented. However, it is work-in-progress. And, from their characteristics, we are certain that agents be used to manage security issues in a distributed network.

Mobile agents are computer programs which are autonomous and have the ability to learn. They move from one node to another and interact with each other, sharing information to carry out the desired goals. Mobile agents spread intelligence across the network [1], while they are traversing the network. The mobility of mobile agents allows them to be created, deployed and terminated without disrupting the network configuration [2]. Mobile agents are capable of migrating across the network and performing application-specific tasks like sensor readings, make decisions in order to reduce the network traffic and distribute the load across the network by using load balancing schemes.

Telecommunications in today’s world follow certain trends which can be structured into different levels, including solution areas such as electronic business and mobile business, content and services, multi-media and entertainment, internet, TV and local area networks where the user is always connected to services and content, such as in GPRS and 3G communication. These trends are characterized by an increased mobile devices, software and humans in the global society. However, in order to realize the potential created by the technology innovations, some important requirements for the next generation of networks need to be satisfied, which include:

- Large bandwidth access to content and services to exploit the potential of multimedia.
- Provide users with added-value and ease of navigation through the diverse content and service offerings.
- Need to be support the user with intelligent interfaces over multiple devices, like mobile phones, next generation Personal Digital Assistants (PDAs), web pads and PCs, etc.
- Building more security mechanisms into intelligent communications
- Need for network management systems that support scalability as networks are become more complex and dynamic.

The agent technology in telecommunications promises to be a key vehicle for achieving higher level of communication and enabling more intelligence in service provision and network management by integrating different services and to support improved quality of service.

In order to fulfill these requirements, the agents need to communicate to discover peers, negotiate and co-operate in open environments. Most agent systems need to build on an interface with a variety of existing and upcoming developments and standards at the underlying network systems level. Some applications, especially those related to multimedia, require guaranteed transmission of data with a certain degree of reliability and quality of service in a communication network [3].

One fundamental problem in telecommunication network management is load balancing, devoid of overloading and traffic congestion in networks, even when most of the network nodes are not used to full capacity [4]. The mechanism of the mobile agent is based on the load balancing method in which no router is left idle or overloaded and allows a network to utilize its capacity to the maximum. Quality of service method allows for better use of the existing network infrastructure, improves service to the end users and reduces the cost of providing these services. The main purpose of quality of service is to dedicate bandwidth and such like parameters to meet latency within the network required by some of the real time applications and for recovering degradation in network quality of service [5]. This paper is structured as follows: section 1, introduces the paper and the application of mobile agents in providing security in wireless networks, section 2 provides a brief overview of the related work on the application of mobile agent technology in network management. In section 3, we discuss mobile agent technology and some of its key characteristics and potential applications in the telecommunications industry. Section 4 highlights some advantages of mobile agents. Section 5 concludes this paper.

2 Related work

Extensive research work has been done over the past years, on mobile agent implementation for network management. There are several threads of research that have used mobile agents in telecommunications networks to manage connectivity and load balancing; some of these are highlighted below.

- The project Mobile Intelligent Agents for the Management of the Information infrastructure (MIAMI) [6] has the objective of examining the applicability of mobile agents to a network and has defined a case study and associated environment, which will allow co-operating customers to dynamically form virtual enterprises for providing services to end-users. The virtual enterprise makes use of services offered by an active virtual pipe provider, a business role similar to telecommunication management network value added service provider or telecommunication information networking architecture retailer.
- Decentralizing control and intelligence in network management [7] is a research in which network management is seen as capable of scripting and delegating agents to remote sites where they are incorporated into local network management programs and are used for intelligent tasks such as managing information based filtering. This application brings mobile agents into network monitoring and network control. Indeed, facilitating the migration or traversal of mobile agents in a telecommunication network and allowing asynchronous and cooperative processing of tasks and specialization of services.
- Current network management systems such as Simple Network Management Protocol (SNMP) for data networks [8], and Common Management Information Protocol (CMIP) for telecommunication network [9], are typically designed according to a centralized model, which are characterized by lack of distributed operation, low degree of flexibility, re-configurability, efficiency, scalability and fault tolerance [10]. They also require network administrator to make real-time decisions and find solutions for the series of problems in the network. These network managements deal only with data gathering and reporting methods, which in general involves substantial transmission of management data. This consumes a high bandwidth in the system, causing considerable strain in the network [11].

These management activities are limited since they cannot do intelligent processing such as judgment, forecasting, decision making, analyzing data and make positive efforts to maintain quality of service. Therefore, there is need for distributed of management intelligence by using mobile agent to overcome the limitations of centralized management.

3 Intelligent agents

An intelligent agent is a computer system, residing in some environment that is capable of flexible, autonomous action in order to meet its design objectives [12]. A multi-agent system is a dynamic federation of software agents that are coupled through shared environments, goals or plans that cooperates and coordinates their actions [13]. It is this ability to communicate, coordinate and cooperate that makes agents and multi agent systems a worthwhile concept in computing and attractive when it comes to tackling some of the requirements in next-generation telecommunications systems.

Mobile agents or intelligent agents are software agents that are capable of migrating between multiple hosts to carry out computations on different hosts, following an itinerary. Mobile agents can travel in a network following their itinerary carrying logic and data to perform a set of management tasks at each of the visited nodes in order to meet their designed objectives. Mobile agents allow the transformation of current networks into remotely programmable platforms. Mobile agents are powerful software interaction model that let a program to be moved between hosts for remote execution. They are solutions for managing distributed networks [14].

The primary goal of using mobile agents in management of telecommunication network is to reduce the network traffic by using load balancing and building scalable and reliable distributed network management system

Intelligent agents are software components characterized by autonomy (ability to act on their own), reactiveness (to process events), proactiveness (to reach goals), cooperation (to efficiently and effectively solve tasks), adaptation (to learn by experience) and mobility (migration to new places). The real strength of an agent is based on the community of multi agent system and the negotiating mechanisms and coordinating facilities. A multi agent system is a dynamic group of software agents, coupled by common environments, goals or plans, which cooperate with each other to coordinate their actions [14]. Applications requiring distributed computing are better supported by multi agent systems, since agents can be designed as fine-grained autonomous components acting in parallel. However, to support multi agent systems, an appropriate environment has to be supported, an infrastructure has to be established specifying communication and interaction protocols, which is open and not centralized, and contains agents which are autonomous, adaptive and cooperative.

Some specific properties of agents include the following:

Agent Mobility: is the capability of transporting objects which include code and state to a network element. Software agents can be static or mobile. Stationary agents remain resident on a single system during their execution lifetime; mobile agents on the other hand, have the ability to migrate from one system in a network to another during their execution lifetime. Mobile agents are able to suspend execution on one platform and move to another and resume execution.

Agent autonomy and intelligence: this is the capability of autonomous decisions particularly as a reaction to events in the network. In the context of networking, autonomy and intelligence are interconnected. Giving an agent a great degree of autonomy and intelligence may also increase the risk of damage in the event of a malfunction or when such an agent is used for an attack.

Mobile agent Implementation: The implementation of an agent is the code it uses to execute. It is important that an agent's implementation should be both executable at the destination host and also be safe for the host to execute in such a way that ensures, the confidentiality, integrity and availability of the host, its data and services. In a realistic scenario, the prototype of a mobile agent could be implemented using the Java Development Kit, including the platform JADE (Java Agent Development Framework) 3.7 which is a software framework that simplifies the implementation of multi-agent systems [15]. The agent platform can be distributed by moving agents from one host to another. Once implemented, mobile agents operate directly at the hosts at which actions are taken, this makes their reaction time faster than a system where actions are taken by a central controller. In other words, mobile agents can be dispatched from a central controller to monitor the network activities of interest, making them to respond to changes in the execution environment at the destination host in real-time. The distributive nature of the mobile agent network enables the network administrator to effectively monitor the trend of traffic flow in the network to assess network performance and identify unusual conditions. The analysis of data can be achieved from the management information base, which preserves various data objects for network management.

Potential Wireless Applications of Mobile Agents: Several mobile agent researchers have proposed to use mobile agents in various wireless applications.

The following applications can be potentially deployed more efficiently using mobile agents.

- Information Retrieval: The most prominent application of the mobile agents is in distributed information retrieval. The information available on Internet is exponentially growing. In addition, the information is widely distributed. The information that can be retrieved using a search engine has its own limitations. For example, the use of traditional methods of communication can result in significant overhead both in terms of wireless bandwidth consumption and latency. The deployment of mobile agent technology can significantly improve the applications performance because of the ability of mobile agents to roam autonomously in the network until the information is gathered and hence no need for intermittent communication through the wireless channel [16].
- Filtering Data Streams: The second prominent application of mobile agents is in filtering the bulk of resulting data to return only what is relevant to the mobile user. Clearly, the agent avoids the transmission of unnecessary data, but does require the transmission of agent code from client to server. If the agent code is precisely sent, a great saving of bandwidth and time can be attained.
- Distributed management of mobile networks: Mobile agents have become an interesting method to exploit synergies between current research on network management and agent-related research. While network management looks for the new ways to overcome the limitations of current client-server technology, mobile agent and peers computing provide technologies and architectures to enable decentralized, peer-to-peer communication. Existing network management systems basically use the client/server method for their functionalities; these systems suffer from poor scalability due to the increase in the amount of communication and generate too much traffic in the network and the number of failures in nodes and channels. For proper network management, network administrators need to locally observe and control components on multiple nodes in the

system. The traditional network management architecture is inefficient, expensive and difficult to change. Hence, the need to increase the level of the automation for improving the effectiveness of management operations and reducing the cost [17]. Thus, there is a need to employ mobile agents as an autonomous entity in the network management and to transfer the administration tasks to the agents. In this situation, the network management tasks and computational load are distributed instead of being centralized towards and on the manager host. One important goal of the network management is to have a balanced and reliable network loading such that connections in the networks can be established quickly without noise, or several trails. An important function in the area of network management is performance measurement, which involves gathering statistical information about network traffic, methods to reduce and present data. The use of decentralization in this kind of applications potentially solves most of the problems associated with centralized client/server solutions. Hence applications can be more scalable, more robust, can be easily upgraded or customized and they reduce the traffic in the network.

Mobile agents allow the transformation of current networks into remotely programmable platforms. Mobile agents are a powerful software interaction model that let a program be moved between hosts for remote execution. They are basically solutions for managing distributed networks. The concept of remote programming using mobile agents is considered as an alternative to the traditional client-server programming based on the remote procedure call or the static distributed object paradigm [18]. In a distributed network, the network operator monitors the trend of network flow to assess network performance and identify unusual conditions. The analysis of data can be achieved from the management information base. Management information base preserves various data objects for network management; the information in management information base is ordered in clusters and maintained in a tree-like structure managing the complex network tasks a distributed network environment [19]. Network management also aims to organize the network to work more efficiently, successfully adjust to changes and react to problems such as traffic patterns with ease.

4 Advantages of using mobile agents in wireless networks

Key advantages of using mobile agent technology [1], [4] in wireless communication networks include:

- Distribution of Management Code: Mobile agents distribute management code to the Simple Network Management Protocol (SNMP) agents by reducing bandwidth in wireless network, which is of particular importance considering the scarcity of the bandwidth resource in wireless communication.
- Decentralization: Mobile agents effectively decentralize network management functions. Mobile agents can autonomously, proactively carry out administration tasks such as installing and upgrading software, or periodically monitoring the network and they also reduce traffic needed for management.
- Dynamically changing network policies: Mobile agents can change the rules underlying network management from time to time. In current network management systems, this is done following a complete, rewrite, compile, run cycle, using agents, these adaptations can be done dynamically and incrementally, by replacing agent or agent functions one at a time.
- Network Monitoring: Mobile agents are useful for supervision for SNMP variables and long-term controlling of network elements, especially in wireless network as the configuration might change over time.
- Data collection: Mobile agents deal with huge amount of data which they can search, collect, and filter. They can be used to process data-intensive requests from network element. Here, the agent acts as a “smart query” that visits the data and performs the necessary computation locally, instead of passing large chunks of data over the network.
- Robustness: Agents can carry out their tasks at least to a degree, even if parts of the network are not reachable temporarily. This is important in mobile computing, where links are expensive and unstable.
- Re-activeness: Agents can react promptly to local events, such as the breakdown of a link.

As a research, we propose the implementation of autonomous mobile agents using the Java Development Kit, including the platform JADE (Java Agent Development Framework) 3.7 which is a software framework that simplifies the implementation of multi-agent systems [15]. The agent platform can be distributed by moving agents from one machine to another. Once implemented, mobile agents operate directly at the hosts, where actions are taken; this makes their reaction time faster than a system where actions are taken by a central controller to monitor the network activities of interest, making them to respond to changes in the execution environment at the destination in real-time. The distributive nature of the mobile agent network enables the network administrator to effectively monitor the trend of traffic flow in the network to assess network performance and identify unusual conditions. The analysis of data can be achieved from the management information base, which preserves various data objects for network management.

5 Conclusion

Network management systems based on the Simple Network Management Protocol use more bandwidth and create network traffic. They cannot satisfy the various requirements of heterogeneous networks, maintain an essential level of quality of service and reliability for the end user and multimedia applications. Therefore, mobile agents offer a solution of flexibility in management of today's telecommunication networks. Agents are autonomous entities and their usage in network monitoring and management can over the shortcomings of SNMP by decentralizing network monitoring and management. Therefore, in this research, *network management* and *agent mobility* have been identified as two major areas where the deployment of mobile agents could be particularly beneficial in wireless communication.

This paper outlined what is essentially a work-in-progress report however; it is our conviction that taking into account the characteristics and the advantages of using autonomous mobile agents, improved network management can be achieved. The next phase of the research is to implement and test the proposed solution based on autonomous mobile agents.

6 Referencing

[1] C. Tsatsoulis, L. K. Soh, "Intelligent Agents in Telecommunication Networks", Computational Intelligence in Telecommunications Networks, W. Pedrycz and A. V. Vasilakos (Eds.), CRC Press, 2000.

[2] M. Ghanbari, D. Gavalas, D. Greenwood, M. O'Mahony, " Advanced network monitoring applications based on mobile/intelligent agent technology", Computer Communications Journal, April 2000

[3] Sahai, C. Morin, "Towards Distributed and Dynamic Network Management'', in the Proceedings of IFIP/ IEEE Network Operations and Management Symposium (NOMS), New Orleans, U.S.A., Feb, 1998.

[4] S. Lipperts, B. Kreller, "Mobile agents in Telecommunications Networks a Simulative Approach to Load Balancing", Proc. 5 thIntl. Conf. Information Systems, Analysis and Synthesis (ISAS'99), 1999

[5] Bieszczad, B. Pagurek, T. White, "Mobile Agents for Network Management", IEEE Communications Surveys, 1998.

[6] Mobile Intelligent Agents for the Management of the Information Infrastructure (MIAMI) ACTS project, <http://www.ee.surrey.ac.uk/CCSR/news/publicity/CCSR -1997>

[7] K. Meyer, M. Erlinger, J. Betzer, C. Sunshine, G. Goldszmith, Y. Yemimi, "Decentralizing Control and Intelligence in Network Managements", in Proceedings of the Fourth International Symposium on Integrated Network Management, Santa Barbara, California

[8] W. Stallings, "SNMP and SNMPv2: The Infrastructure for Network Management", IEEE Communications Magazine, vol.36, no.3, pp. 37-43, March 1998.

[9] Y. Yemini, "The OSI Network Management Model", IEEE Communications Magazine, vol.31, no.5, pp.20-29. May 1993

[10] M. Ghanbari, D. Gavalas, D. Greenwood, M. O'Mahony, " Advanced network monitoring applications based on mobile/intelligent agent technology", Computer Communications Journal, 23(8), pp. 720-730, April 2000.

[11] T. White, B. Pagurek, A. Bieszczad, "Network Modeling for Management Applications Using Intelligent Mobile Agents", Journal of Network and Systems Management, vol. 7, no.3, 1999.

[12] Jennings, N.R., Sycara, K., and Wooldridge, M.J. A Roadmap of Agent Research and Development. Autonomous Agents and Multi-Agent Systems, 1998.

[13] Huhns, M.N. Multiagent Systems. Tutorial at the European Agent Systems Summer School (EASSS'99), 1999.

[14] Satoh, "Building Reusable Mobile Agents for Network management", IEEE, 2003.

[15] http://jade.tilab.com

[16] D. Kotz, G. Cybenko, R. Gray, G. Jiang, R. Peterson. Performance Analysis of Mobile Agents for filtering Data streams on Wireless Networks. *Mobile Networks and Applications.* 2002. pp 163-174

[17] Bieszczad, B. Pagurek, T. White, "Mobile Agents for Network Management", IEEE Communications Surveys, 1998

[18] T.C. Du, E.Y. Li, A.P. Chang, "Mobile Agents in Distributed Network Management", Communications of the ACM, vol. 46, 2003

[19] C. Bohoris, G. Pavlou, H. Cruickshank, "Using Mobile Agents for Network Performance Management", in proceedings of the IFIP/ IEEE, Network Operations and Management Symposium (NOMS'00), Hawaii, U.S.A., April 2000

Fixed cluster head based routing for heterogeneous mobile wireless sensor network

Jaideep Lakhotia[1], **Rajeev Kumar**[2]
[1,2]Computer Science and Engineering Department, NIT Hamirpur, Hamirpur, Himachal Pradesh, India

Abstract - *The Wireless Sensor Network is one of the most research focused area in today's ongoing research work. It catches researchers' attention due to its capability of deploying anywhere for sensing purpose such as a military area, power plant etc. One of the research issues in Wireless Sensor Network is the mobility of sensor nodes as some of the applications require mobile nodes such as health monitory and machine performance monitoring etc. So mobile node based routing brought up new challenges in Wireless Sensor Network. Some Researchers have proposed the routing protocols based on mobile nodes but some have its own assumption that is not feasible for real world implementation as it incurs more cost. Others have proposed the protocol by keeping extra TDMA time slot, GPS installed on a sensor node etc. So we have proposed the routing protocol with fixed cluster heads for mobile sensor network. This protocol increases the lifetime of mobile nodes and reduce the time mobile sensor nodes take to join the new cluster head.***

Keywords: *Mobile wireless sensor network, Routing, Mobile Nodes, Cluster, Static Node*

1 Introduction

Initially the routing protocols designed for sensor network are developed considering static sensor nodes as in LEACH [1]. But as of now, some applications require mobile nodes in wireless sensor network. So researchers are working to develop routing protocols considering mobile nodes in the network. Some of the routing protocols developed so far that considered mobile nodes are LEACH-M[2], LEACH-ME[3], CBR-M[4], ECBR-MWSN[5], E2R2[6], 2L-LEACH-M[7], FTCP-MWSN[8], LFCP-MWSN[9]. All these protocols work on a cluster based wireless sensor network with different cluster head selection criteria and different ways to deal with the mobile nodes present in the network. The main focus of these protocols is to use the limited energy, and the packet loss should be less due to mobility, less end to end delay, high throughput and so forth. So in this paper we have introduced a cluster based routing protocols in which the cluster is fixed and cluster head is also fixed as the cluster head is a static node with unlimited energy supply and there is mobile nodes which joins these cluster heads to transmit their data, so the main focus of this paper is to introduce a routing protocol which is having a less delay and less amount of packet loss as compared to existing protocols. This paper is organized as follows: In section 2, different routing protocols that work for Mobile Wireless Sensor Network has been described. And section 3 describes the network model for mobile node based Wireless Sensor Network. In Section 4, proposed algorithm for mobile sensor network is discussed. And we define conclusion and future work in section 5.

2 Related Work

In LEACH[1], it considered all the nodes as static so does not work efficiently in case mobile nodes being introduced in sensor network, so the authors proposed M-LEACH[2] protocol which selects cluster head based on mobility, location, residual energy of node. It also distributes cluster head uniformly by dividing sensing area into subareas and it is assumed that nodes are being installed with GPS so that node's location can be known. Mobile Node detects that it goes out of cluster when it does not receive request from cluster head for two consecutive frames of data transmission phase and then it sends join request message for registering with new cluster head but it considers only mobile sensor node and base station is fixed.

LEACH-ME [3] is another advancement to LEACH [1] protocol and M-LEACH [2] protocol in which cluster head is selected on the basis of no movement of a node or it has less relative movement in the cluster. It also includes extra time slot in TDMA schedule for mobility calculation based on the number of times node moves from one cluster to another, this extra time slot is added in TDMA schedule on the basis of slot frequency which depend on node movement in the cluster. Here also the node can detect that it goes out of range of current cluster head if it does not receive data requests from cluster head in two consecutive frames of data transmission phase and Cluster Head remove the TDMA time slot for a mobile node if it does not receive data from a mobile node in two consecutive time slots.

CBR-M [4] is the cluster based routing for mobile nodes in wireless sensor network. In this protocol the work is done to reduce packet loss occur in [2]. Cluster Head removes the TDMA time slot for a mobile node if it does not receive data from a mobile node in its allocated time slot. So cluster head rebroadcasts the updated TDMA schedule to the cluster members in case node is moved from the cluster. The main idea in this protocol is that there should be one cluster free to receive packets from the mobile node that goes out of the cluster and cannot receive data request message from its cluster head. The nodes in a sensor network wake up one time slot before the TDMA schedule for it and for the rest of the time it is in sleep mode that saves energy of these nodes. ECBR-MWSN [5] is advancement in CBR-M [4] protocol in terms of energy efficient cluster head selection. It selects cluster head based on primary criteria i.e. low mobility, high residual energy and secondary criteria i.e. distance between node to a base station which is used in case of a tie occur after checking primary criteria.

Another protocol is 2L-LEACH-M [6] which is the advancement in LEACH [1] protocol to support mobility and result shows that it improves data transfer rate in comparison to LEACH [1] protocol in the mobile node scenario. In this protocol node are divided into two levels, Level1 is cluster head level and Level0 is member level. In this cluster member belongs to cluster head nearest to it. Nodes are location aware of itself. If the node moves it send the data and the new cluster head has no TDMA schedule for this node then after finishing current frame cluster head checks whether it receives sensed data from any node for which it has no time schedule, if it receive data from such a node then it include that node in TDMA time schedule.

E2R2[7] routing protocol considers link failure that can occur due to the movement of mobile nodes. In this Base Station selects two deputy cluster head and one cluster head in each cluster. These deputy cluster heads help in cluster management in case of link failure. A node that cannot send a data to the cluster head due to link failure can send it to deputy cluster head and deputy cluster head forwards it to the Base Station. Cluster head also keeps TDMA slot for these deputy cluster heads which is not being used usually it is not only used for sensing purpose but also for cluster management. In case of link failure between cluster heads, the cluster head can transmit data through its deputy cluster head to base station. It has been analyzed that it performs better than leach protocol in terms of lifetime and throughput of the network is considered.

FTCP-MWSN [8] is energy efficient and fault tolerant routing protocol for Mobile wireless sensor network. By this protocol we can actually determine that a node is mobile or it fails. It considers that all nodes are mobile and there is a high probability of a node getting out a cluster and another node coming into the cluster. It does not keep extra time slot for calculating mobility of sensor node rather the node sends its mobility information in the TDMA schedule itself. If the node moves out of the cluster at x time interval and same node sends join request in x+1 time interval to another cluster head then it is declared mobile otherwise node get failed and this is calculated by the base station as cluster heads inform the Base Station about node ID when it leaves the cluster or it enters the cluster. In this protocol sensor node sends a special packet in case it doesn't have data to send and saving the energy in this way. So there is no false detection of the mobile node as the failed node.

LFCP-MWSN[9] is similar to FTCP-MWSN[8] in terms of fault tolerance but it also introduces anchor nodes for calculation of location rather than any GPS installation on sensor nodes. Anchor nodes send its location to mobile nodes so that it can calculate its current position. It reduces energy consumption and end to end delay as compared to M-LEACH[2] and LEACH-ME[3].

2.1 Important Factors in Mobile Wireless Sensor Network

A. *Energy-Efficiency*: Protocols should not have the high computation overhead and always consume less energy to increase the network lifetime.

B. *Less end to end delay*: There should be less delay in passing the data information forward to the Base Station.

C. *Less packet loss:* Packet loss occurs due to the mobile node should be minimized by the approach.

D. *Link management*: In case of link failure due to mobility of node the alternate route is to be selected dynamically.

E. *Cluster Head Selection*: Appropriate cluster head selection is important such that nodes in a cluster do not consume more energy to send data to cluster head.

F. *Location Aware:* Base Station cannot communicate with a sensor node if it does not know the location of the node. So location awareness is necessary in case of mobile nodes.

3 Network Model

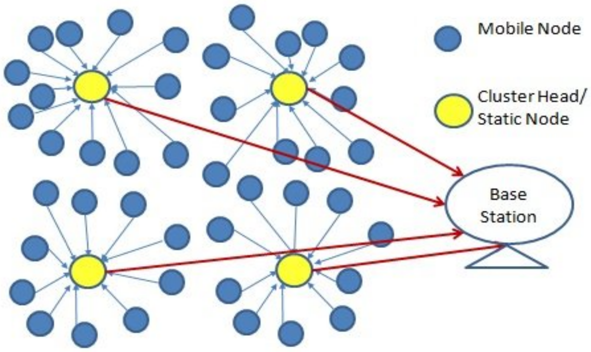


Fig1: Network Model of Mobile Sensor Network

In the network model described in figure 1, I have considered static and mobile nodes in the network and static nodes for the purpose of becoming a cluster head which is fixed because these nodes's batteries can be recharged if needed. Mobile nodes with homogeneous mobility in the network are for the purpose of data transmission based on their TDMA schedule to the cluster head. So here we have considered the routing in mobile sensor networks where the cluster head is fixed and it is suitable for the application where the supply of

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

electricity is possible like in industries, houses and workplaces.

4 Proposed Scheme

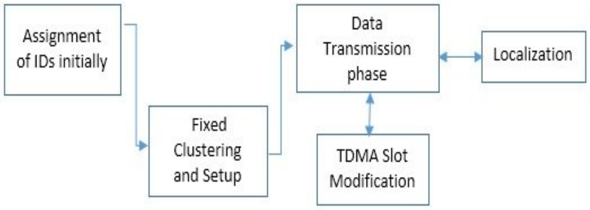


Fig 2: Working Model of MWSN

After deployment of the sensor nodes accordingly then setup phase takes place to make clusters and Cluster Heads so that data transmission can be done. In this approach there is a fixed cluster and fixed cluster heads.

4.1 Algorithms:

4.1.1 For Fixed Clustering and Set Up

```
1. BaseStation broadcasts --> { Nodes_ IDs,Initial Energy,
   CH location}
2. BS --> (CH, Cluster Area)
3. for i<-1 to No. of Nodes
4.   LocalizationProc(Node[i])
5.   CH Broadcast RegistrationReq
6.   for each cluster j do
7.       if (Position[Node[i]] belongs to Cluster
         Area[ClusterID[j]]andRecvMsg[Node(i)]) then
8.         Node[i] <- ClusterID[j]
9.       endif
10.  end for
11. for j<-1 to No. of Cluster Heads
12.   for i<-1 to No. of Mobile nodes sent registration
      Message
13.       CHj make TDMA schedule for the mobile nodes$_i$
14.   end for
15. end for
16.Cluster Head broadcasts --> (TDMA Schedule,Mobile
   Node)
```

Modification in assigning TDMA slot

Ist Frame: Node will be assigned TDMA schedule based on the id distance. Node with Higher Distance will be assigned slot first. Because of homogeneous mobile node, chances of nodes near the boundary of cluster moving out are higher.

Rest of Frame: *Node Distance + Mobility is being checked*

We modify the TDMA slot according to the mobility of nodes.

1. Nodes are arranged in decreasing order of their distance with cluster head.

2. After arranging them, check whether node mobility is higher than the average mobility of all nodes then assigned the time slot first and rest of the node which have less mobility will be assigned in the rest of the time slot with their decreasing distance.

4.1.2 For TDMA Slot Assignment

```
1. For each node i in cluster j
     Sort nodes in descending order
2. If (frame == 1)
     Assign node in timeslot in this decreasing order
   Else
       Calculate Average mobility= Total sum of mobility/n
       where n is no. of nodes in cluster.
   for i<-1 to n
     if(node$_i$[mobility]>Average mobility)
       Assign Time slot
3. Assigning timeslot according to descending order.
4. CH broadcasts TDMA schedule for cluster's node.
```

Alive Message: When Node does not receive data requests from Cluster head then Node sends an Alive Message to it which contains the location and node id of that node. The node also sends Alive Message in reply to the Status message.

Status Message: When Cluster Head does not receive data from sensor node then it sends Status Message containing its ID, location and node ID of sensor nodes.

Linked Message: This is the message sent by the Cluster head to the sensor node in reply to Alive Message.

During Steady State phase in which data transmission has to be done following cases arises:

Case 1: If any mobile sensor node does not receive data requests in its respective time slot, after waiting for half of a time slot then that node send alive message to cluster head. In reply to it Cluster head has to reply Linked message. If the sensor node does not receive Linked message then it assumes that it moves out of the cluster and start sending a join request.

Case 2: If Cluster Head does not receive data or the alive message up to half of time slot then after sending data request, it sends a status message for getting information of sensor node if it is there. Otherwise if Cluster Head does not get any reply from sensor node then it assume that the node moves out of the cluster.

Case 3: If any mobile sensor node does not receive ACK after sending data. Then node just remains active up to next time slot to check whether it is in cluster if it receive a data request by the cluster head to another node then it recognizes that it is still present in cluster so no need to send join request.

Case 4: In case the node does not have any data to send it sends the special packet to its cluster head so that cluster head able to know that the node does not have any data to send and in that phase cluster head can send time slot for the new node. For Data transmission phase the algorithm proposed in [9] has been modified.

4.1.3 For Data Transmission Phase

```
1. for each frame f do
2. for each node j of a cluster k <-t(a timeslot) do
3. if(node[k][j] has an sensed event and received data request
     then
      node[k][j] sends data to CH
      Evaluate DataEnergyConsumption[k][j], ReceivedEnergy
      of CH[k]
      CH sends ACK to node [k][j]
   else
      node[k][j] sends a special packet to CH
       Evaluate Special Energy Consumption[k][j], Recive
       Energy of CH[k]
       CH sends ACK to node[k][j]
4. if node[k][j] does not receive data req from  CH[k]
       Then Send Alive Message to CH[k] before the half time
         of slot; if no reply
           ++CountMobFactor[node[k][j]]
            Localize[CountLoc[k]++] <- ID[node[k][j]]
            Start broadcasting Join Request
5.  if CH not receive data/special from node[k][j] or Alive
     Message even at half time then
        Send status message to node[k][j] and  if no reply
          delete node[k][j] and timeslot of node[k][j]
          notify BS about node[k][j] sending MOVED-Node(j)
6. if node[k][j] not receive ACK from CH then
        Remain Wake up for next slot
7. if receive data request of another node
        wakeup during assigned time slot in next data   frame.
   else
        ++CountMobFactor[node[k][j]]
        Localize[CountLoc[k]++] <- ID[node[k][j]]
        Broadcast JOIN-REQUEST
        CH within shortest communication range replies with
          ACK-JOIN and notify BS the ID   of node[k][j] .
    if BS receives ID  for node[k][j] in two frames  then
         mark node[k][j] as a moved-node
   else if BS receives not receives ID in two frames
         mark node[k][j] as failed //fault tolerance
   end for //each node
8. CH[k]aggregates Data and Sends to BS
10. for i<-1 to countLoc[k] do
         LocalizationProc(Localize[i])
    end for
   end for frame
11. Calculate CHEnergyConsumption[k]
12. if CH not receive data from the node where
     CountMobFactor<Average MobFactor
        then TDMA slot modification and  broadcasts.
   else
       Start another frame.
```

Localization Procedure is done using Mobile Anchor Nodes as in [9].

5 Conclusions

So in this paper we have proposed the routing protocols that supported the wireless sensor network having mobile nodes that can move out of the cluster. The work has been carried out for the improvement of the existing protocol. Delay is the importance factor of data transmission, so in this work we have considered delay as a prime factor and proposed an algorithm to reduce the delay and packet loss. So in the future work we will add mobility of cluster head in the proposed protocol with simulation results.

6 References

[1] Heinzelman W.R. , A.P. Chandrakasan and H. Balakrishnan.: ‘*Energy-Efficient Communication Protocol for Wireless Microsensor Networks*’. Proc. of the 33rd IEEE Int. Conf. on System Sciences, Honolulu, USA, Jan. 2000, pp. 1–10.

[2] Kim D.S., Chung Y.J.: ‘*Self-organization routing protocol supporting mobile nodes for wireless sensor network*’. First Int. Multi-Symp. On computer and Computational Sciences (IMSCCS’06), 2006.

[3] Kumar, G.S., Vinu, M.V., Athithan, P.G., Jacob, K.P.: *’Routing Protocol enhancement for handling node mobility in wireless sensor networks’*. TENCON 2008, 2008 IEEE Region 10 Conf., 2008, pp. 1-6.

[4] Awwad, S.A.B., Ng, C.K., Noordin, N.K., Rasid, M.F.A.: ‘*Cluster based routing protocol for mobile nodes in wireless sensor network*’. Int. Symp. On Collaborative Technologies and Systems, CTS’ 09, 2009, pp. 233-241.

[5] R.U.Anitha, P.Kamalakkannan.: ‘*Enhanced Cluster Based Routing Protocol for Mobile nodes in Wireless Sensor Network*’.Proc. of the 2013 IEEE Int. Conf. on Pattern Recognition, Informatics and Mobile Engineering, Feb, 2013, pp. 187-193.

[6] C.Zhang, J.Liu, H.Li, X.Jiang.: ‘*Two-Level Routing Protocol for Mobile Sensor Network Based on LEACH Algorithm*’. IEEE 2010, pp. 950-953.

[7] Hiran Kumar Deva Sarma, Avijit Kar, Rajib Mall.: ‘ *Energy Efficient and Reliable Routing for Mobile Wireless Sensor Networks*’. IEEE 2010, pp. 1-6.

[8] L. Karim, N. Nasser.: ‘*Energy Efficient and Fault Tolerant Routing Protocol for Mobile Sensor Network*’.IEEE Int. Communications Conf. (ICC’11), Ontario, Canada, 2011.

[9] L. Karim, N. Nasser: ‘*Reliable location-aware routing Protocol for Mobile Wireless Sensor Network*’.IET Commun., 2012, Vol. 6, Iss. 14, pp. 2149-2158.

SESSION

WIRELESS NETWORK APPLICATIONS, AD-HOC NETWORKS, VANET, NOVEL ALGORITHMS, PROTOCOLS AND RELATED TECHNOLOGIES

Chair(s)

TBA

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

Joint of Feedback and Feedforward Controller on Application to Wireless Tele-Control System Considering Uncertainty

[1]Faramarz Alsharif,[2]Shiro Tamaki, [3]Katsumi Yamashita,[4]Tustomu Nagado,[5]Mohammad Reza Alsharif and [6] Heung Gyoon Ryu

[1, 2, 4, 5] Graduate School of Engineering and Science
University of the Ryukyus
Nishihara, Japan
[3]Osaka Prefecture University
Osaka, Japan
[6] 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]asharif@ie.u-ryukyu.ac.jp
[6]ecomm@chungbuk.ac.kr

***Abstract*— In this paper, we aim to design a Tele-Control system for the closed-loop control system. Tele-Control system consists of control plant, controller, feedforward channel and feedback channel in the closed-loop control system. Feedforward channel is located between controller and control plant and feedback channel, respectively. Basically, channels would be a multipath channel due to the propagated wave reflections. Therefore, undesired signal will be observed due to multipath channel. Thus in Wireless Tele-Control system, in order to design a suitable controller for the closed-loop system, we have to consider not only the control plant but also the feedforward channel and feedback channel. In other words we design the controller corresponding to the open-loop system. Additionally, if there is an uncertainty in plant, there would be a huge error in system or even would be unstable, since feedback controller designed corresponding to plant. Thus, we implement feedforward controller to overcome the uncertainty issues. Moreover, the effects of channels are reduced after equalization by FIR(**Finite Impulse Response)**filter. The stability and performance of the closed-loop system can be evaluated by step response. The control plant is set to be an unstable single input and single output system. Eventually, in conclusion we discussed about the performance and stability of Tele-Control system considering uncertainty that joint controller of feedforward and feedback could enhanced the performance and maintain stability.**

Keywords: Tele-Control System, Multipath Channel, Equalization, PID Controller, Time Lag

1 Introduction

The utilization of Tele-Control system is one of the significant issues in the servo systems. Especially, when system requires control in distant. The advantage of Tele-Control System is that we can realize servo systems to be managed and observed it's behaviors from distant and for maintenance of controller since controller is located in observation center. Let us clarify the Tele-Control system. Basically, in Tele-Control system they are always two channel. 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. So, these channels are disadvantages of utilization of 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 we have 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 in receiver side we will have distracted signal by the multipath channel. 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 channel of the unknown channel can be done by FIR adaptive filter. However, the composition of the replica channel is not sufficient. We have to realize the inverse system of replica Channel. Therefore, the inverse channel is realized after the receiving the distracted signal. This has role of equalizing the received signal. These kinds of process should be implemented in two different stages. One is in the feedforward side of the receiver and the other one in 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, practically in plant there is uncertainty constantly. Thus, feedforward controller is implemented in wireless Tele-Control system. The approached method feedforward controller is model matching method. By Model matching method free parameters is tuned to stabilize the closed loop system and enhance the performance, simultaneously. In the next chapter more details of wirelss Tele-Control system and joint controller of feedback and feedforward are explained.

2 Design of Feedback Controller for an Unstable System

In this chapter we introduce briefly design of an unstable system in order to stabilize the closed loop system. First of all let us consider a plant which is that $G = \frac{1}{s-p}$ for *a* >0 . A basic feedback controller can be considered as $K = \frac{as+b}{cs}$ that set (a, b, c) are tuned to make the closed loop system internally stable.

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: $cs^2 + (a - pc)s + b = 0$.

To satisfy condition (1) *a* should be greater than *pc. That maked a/c >P which c>0 and b>0* as well.

Then the solution (pole of the closed loop) is

$$s = \frac{-(a-pc) \pm \sqrt{(a-pc)^2 - 4bc}}{2c}$$

Then we would have three cases.

Case1. When $(a\text{-}pc)^2\text{-}4bc>0$

$$p_1 = \frac{-(a-pc) + \sqrt{(a-pc)^2 - 4bc}}{2c} < 0$$

$$p_2 = \frac{-(a-pc) - \sqrt{(a-pc)^2 - 4bc}}{2c} < 0.$$

Case2. When $(a\text{-}pc)^2\text{-}4bc=0$ *(Duplicated solution)*

$$p_{1,2} = -(a-pc) < 0.$$

Case3. When $(a\text{-}pc)^2\text{-}4bc<0$

$$\mathrm{Re}(p_1) = \mathrm{Re}\left(\frac{-(a-pc) + j\sqrt{(a-pc)^2 - 4bc}}{2c} \right) < 0$$

$$\mathrm{Re}(p_2) = \mathrm{Re}\left(\frac{-(a-pc) - j\sqrt{(a-pc)^2 - 4bc}}{2c} \right) < 0.$$

(2)- Frequency response of Sesitivity function *S(jω)* should contain following specification.

ω_b : Band width frequency

$$For\ \omega < \omega_b \quad | S(j\omega) | << 0\ [dB]$$
$$For\ \omega > \omega_b \quad | S(j\omega) | \simeq 0\ [dB]$$

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 trackoing error.

(3)- Frequency response of Complementary sensitivity function *T(jω)* should contain following specification.

$$For\ \omega < \omega_b \quad | T(j\omega) | \simeq 0\ [dB]$$
$$For\ \omega > \omega_b \quad | T(j\omega) | << 0\ [dB]$$

The above condition described that when for Complementary sensitivy in the low frequencies it nearly equal to 0 [dB] that means for reference signal that contains direct current componenet, it would become almost same in the output of plant that is desiable. For high frequencies it is desirable to drop to small gain since it can reduce the affect of feedback noise and plant uncertainty.

(4)- Frequency response of open loop function *G(jω)K(jω)* should contain following specification.

ω_c: Cross frequency

$$For\ \omega < \omega_c \quad | G(j\omega)K(j\omega) \quad | >> 0\ [dB]$$
$$For\ \omega > \omega_c \quad | G(j\omega)K(j\omega) | << 0\ [dB]$$

The above condition describes that when open loop in low frequency domain contain big gain that it can enhance the performances and when it is in high ferqnecies it should be small gain to reduce the uncertainty of plant and disturbances.Here is an example of the described conditions.

$$plant : G(s) = \frac{1}{s-2}$$

$$Controller : K(s) = \frac{10s+1}{s}$$

$$Open\ \ Loop : G_O(s) = G(s)K(s)$$

$$Sensitivity : S(s) = \frac{1}{1+G_O(s)}$$

$$Complementary\ \ Sensitivity : T(s) = \frac{G_O(s)}{1+G_O(s)}$$

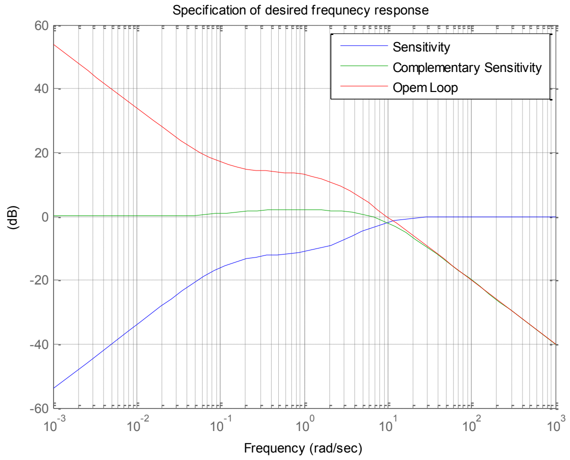


Fig.1 Specification of desired frequency response

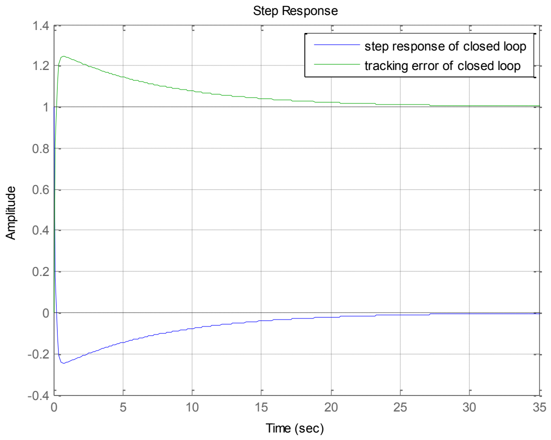


Fig.2 Specification of desired step response

As there are shown in the above figures, Fig.1 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.

$$when \quad s = j\omega \quad and \quad 0 < \varepsilon < 1$$

$$\|G_o(j\omega)\| << \varepsilon \quad for \quad \omega > \omega_c$$

$$\|G_o(j\omega)\| >> \frac{1}{\varepsilon} \quad for \quad \omega < \omega_c$$

$$Through \quad Fig.1$$

$$\omega_b \approx \omega_c$$

$$then$$

$$for \quad \omega < \omega_c$$

$$Sensitivity : \|S(j\omega)\| = \left\| \frac{1}{1+G_O(j\omega)} \right\| = \frac{1}{\|1+G_O(j\omega)\|}$$

$$= \frac{1}{\|1+G_O(j\omega)\|} \approx \frac{1}{\|G_O(j\omega)\|} << \varepsilon.$$

$$for \quad \omega > \omega_c$$

$$Sensitivity : \|S(j\omega)\| = \left\| \frac{1}{1+G_O(j\omega)} \right\| = \frac{1}{\|1+G_O(j\omega)\|}$$

$$\approx \frac{1}{1+\|G_O(j\omega)\|} = 1.$$

$$for \quad \omega < \omega_c$$

$$Complementary \quad Sensitivity : \|T(j\omega)\| = \left\| \frac{G_O(j\omega)}{1+G_O(j\omega)} \right\|$$

$$= \frac{\|G_O(j\omega)\|}{\|1+G_O(j\omega)\|} \approx \frac{\|G_O(j\omega)\|}{\|G_O(j\omega)\|} = 1.$$

$$for \quad \omega > \omega_c$$

$$Complementary \quad Sensitivity : \|T(j\omega)\| = \left\| \frac{G_O(j\omega)}{1+G_O(j\omega)} \right\|$$

$$= \frac{\|G_O(j\omega)\|}{\|1+G_O(j\omega)\|} << \varepsilon \quad .$$

As it is obviously 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 Design of Feedforward Controller based on Model Matching method.

In the most of practical cases actual plant differs from the theoretical modeled plant due to environmental and physical condition. As well there would be a difference between actual model and theoretical model plant by linearization. These kind of uncalculated part, linearized part and perturbation of parameters considered as uncertainty. Uncertainty can affected the stability of closed loop system severely. Even though closed loop system is internally stable, however, uncertainty may degrade the performance of system. To overcome these problems, as we have discussed in the previous chapter, by decreasing the gain of open loop in high frequency domain, it can cover the uncertainty affects. However, to satisfy the necessity when uncertainty is larger than a definite value then feedback is not sufficient. Therefore, in order to sufficiently satisfy the open loop desired frequency response, model matching is implemented as feedforward controller. In the below figure model matching controller is implemented as feedforward controller.

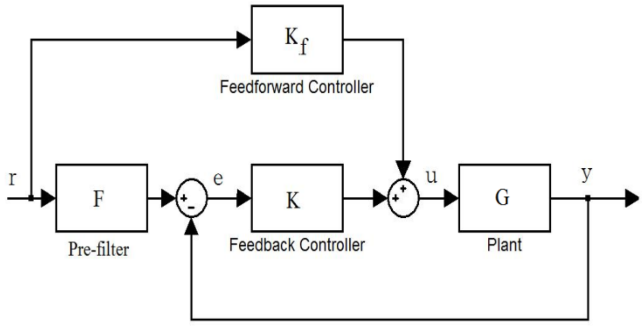


Fig. 3 Joint system of feedback and feedforwad controller

From the above figure, we obtain the sensitivity, complementary sensitivity and open loop transfer function, respectively.

$$\text{Sensitivity}: S_f(s) = \frac{1}{1+G(s)K(s)}$$

$$\text{Complementary Sensitivity}: T_f(s) = \frac{F(s)K(s)+K_f}{1+G(s)K(s)}G(s)$$

$$\text{Open Loop}: G_{fo}(s) = G(s)(F(s)K(s)+K_f(s))$$

Feedforward controller can be designed based on inverted estimated plant and propering filter in casced process. The objective of implementation of propering filter *Q(s)* is to make feedforward controller proper since inverted estimated plant may be an improper system. Thus, *Q(s)* considered as a free parameterization of feedforward controleller.

$$\textit{Estimated plant}: \tilde{G}(s)$$

$$\textit{Feedforward controller}: K_f(s) = \tilde{G}^{-1}(s)Q(s)$$

It is significant that performance of system can be improved by tuning pre-filter when small gain theorem is satisfied.

$$\left\|(K_f(s)+K(s)F(s))W(s)S_f(s)\right\|_\infty < 1.$$

Now let us verfy the previous example by implementing feedforward controller.

$$\textit{plant}: G(s) = \frac{1}{s-2} + W(s)$$

$$\textit{Additive uncertainity } W(s) = \frac{1}{s^2+2^2} + \frac{1}{s^2+10^2} + \frac{1}{s^2+100^2}$$

$$\textit{Estimatedplant}: \tilde{G}(s) = \frac{1}{s-2}$$

$$\textit{Feedback Controller}: K(s) = \frac{10s+1}{s}$$

$$Q(s) = \frac{1}{s^2+7s+1}$$

$$\text{Pre}-\textit{filter } F(s) = Q(s)$$

$$\textit{Feedforward Controller}: K_f(s) = Q(s)\tilde{G}^{-1}(s)$$

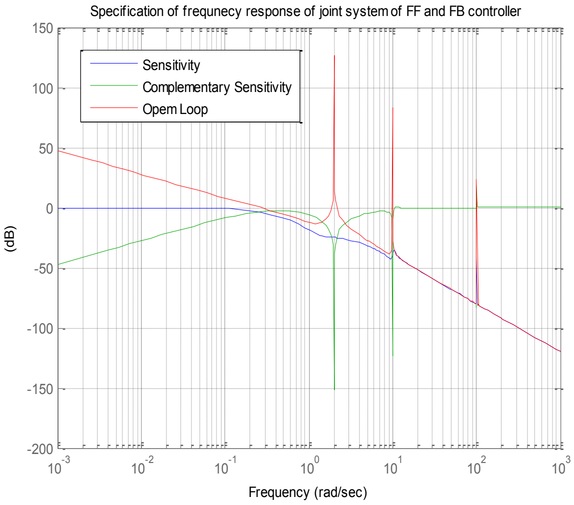


Fig.4 Specification of frequency response of joint system of FF and FB controller

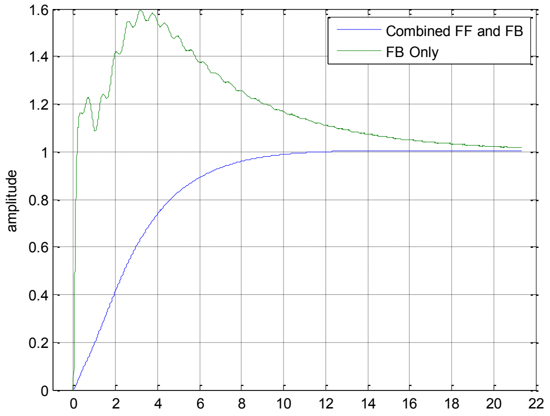


Fig.5 Comparison of step response with joint system of feedback and feedforward controller with feedback controller

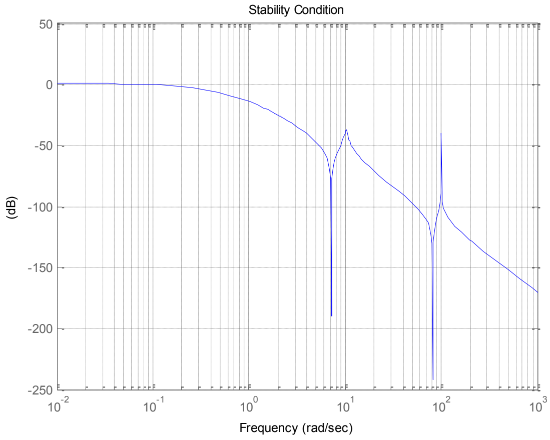


Fig.6 Stability condition of system containing uncertainty

4 Tele-Control System

So far we have discussed about the process of design the controller for combined system of feedforward and feedback controller. Next let us define Tele-Control System as follows. Following figure shows the structure of Tele-Control system.

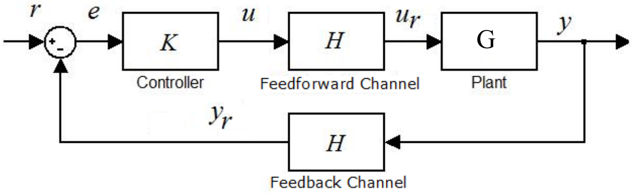


Fig. 7 Tele-Control System

Here H is Multipath channeland u_r,y_r are received input and received output signal, respectively. Through Fig.6, we can get the closed-loop system's transfer function according to following equations.

$$y = PHKe \qquad (1)$$

$$e = r - Hy \qquad (2)$$

Afterward we get the transfer function between r and y which iscomplementary sensitivitytransfer function as follows.

$$y = \Delta_H PHKr \qquad (3)$$

Where, $\Delta_H = \dfrac{1}{1+PKH^2}$ 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. Our proposed method is to reduce the effect of the channel in the sensitivity function. The proposed method has shown in following Fig. 7.

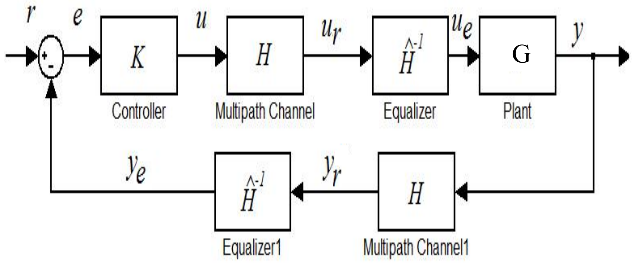


Fig. 8 Configuration of the proposed method

Here, y_e, u_e and $\hat{H}^{-1}$ stand for the equalized input signal, equalized output signal and Equalizer, respectively. $\hat{H}$ itself 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.

$$1 + PH^2 = 0 \qquad (4).$$

Here Multipath channel's model can be express as follows.

$H(s) = \sum_i \alpha_i e^{-L_i s}$. Where α is the attenuation factor and L is Time-Delay and i is the number of taps. In next chapter Multipath channel will be introduced in details.

The polar Expression of Multipath channel is:

$$H(j\omega) = \sum_i \alpha_i e^{-jL_i\omega} \Rightarrow H(j\omega) = |H(j\omega)| e^{j\angle H(j\omega)}.$$

Where, $|H(j\omega)| = \sqrt{\left(\sum_i \alpha_i \cos(L_i\omega)\right)^2 + \left(\sum_i \alpha_i \sin(L_i\omega)\right)^2}$

$$\angle H(j\omega) = \tan^{-1}\left[\frac{\sum_i \alpha_i \sin(L_i\omega)}{\sum_i \alpha_i \cos(L_i\omega)}\right].$$

As it is clear in the above characteristic equation, the stability of system can be determined by solution of equation (4). Thus let us see, what if we have multipath channel and a stable pole.

$$Plant \quad P(s) = \frac{1}{s+1}$$

$$H(s) = \sum_{i=1}^{5} \alpha_i e^{-\tau_i Ls}$$

$$\alpha = [1 \;\; 0.9 \;\; 0.7 \;\; 0.5 \;\; 0.2 \;\; 0.01],$$

$$\tau = [1 \;\; -1.2 \;\; -1.5 \;\; -2.0 \;\; -2.5 \;\; -4.0]$$

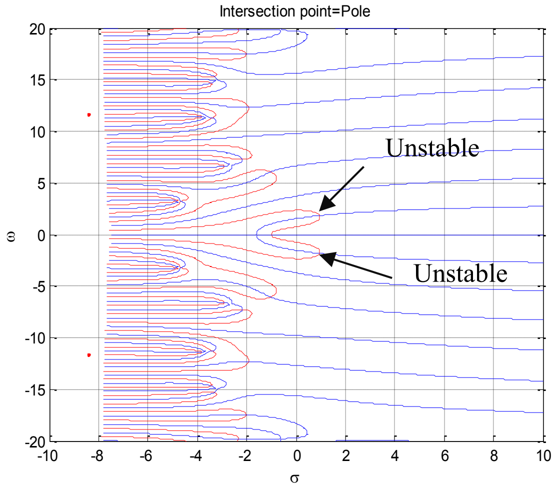


Fig.9 solution of characteristic equation with stable pole considering round trip multipath channel

Figure 9 shows the plotted curve of real and imaginary part of $1+PH^2$*(for* $s=\sigma+j\omega$*)*, respectively. Nodes between red and blue curves are the solutions of characteristic equation. As it is clear even though plant is stable, multipath channel can shift the stable pole to unstable region (Right half plane). Thus, multipath channel not only degrade the performances but also it may cause the instability of system. Next chapter will introduced the reduction of multipath channel effects on control system.

5 Reduction of Multipath Channel Effects in Closed-Loop System

As we have discussed previously, existence of the multipath channel make the system unstable. 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. Assume that $H_{ff}=H_{fb}=H$ then we have

$$y = KH\hat{H}^{-1}Pe \tag{5}$$

$$e = r - H\hat{H}^{-1}y \tag{6}$$

Afterward, we have as follows.

$$y = \frac{PH\hat{H}^{-1}K}{1+PK\left(H\hat{H}^{-1}\right)^{2}}\, r \tag{7}$$

That $\dfrac{1}{1+PK\left(H\hat{H}^{-1}\right)^{2}}$ is the sensitivity function of the proposed method. Here, if and only if when $\hat{H}=H$, then we would obtain equation (12) that is identical to the conventional feedback control system. However, this does not happen since replica channel cannot realize the precise characteristic of multipath channel. Nevertheless, we can reduce the effects of the multipath channel. So this makes the closed-loop system stable. However, performance is going to be degraded anyhow.

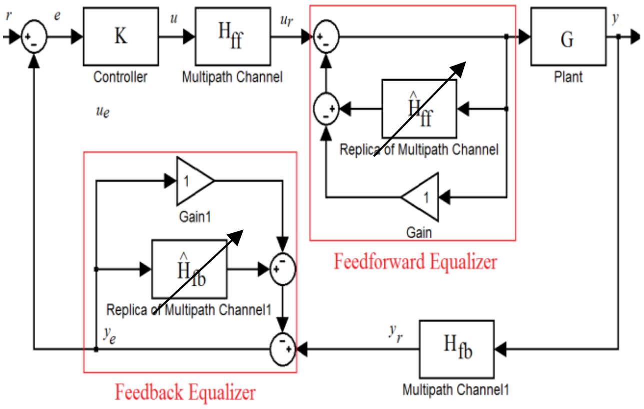


Fig. 10. Configuration of the proposed method in detail

6 Combination of Feedforward and Feedback Controller with Channel Equalizer

As it is known feedback controller is utilized for compensation of error signal which is to modify the deviated output signal with respect to commanded signal (reference signal) and enhancement of performances of the closed-loop system. Moreover, it is required to guarantee the internal stability. So, it has been used for several purposes to realize a suitable and desirable control system. However, for the plant that is aimed to be controlled, there is uncertainty. Uncertainty caused by several factors such as the parameter fluctuation of plant's physical model, plant's linearization, un-modeled part of plant and so on. If uncertainty of the plant is omitted or it is not considered during the controller design, then we would have a degraded performance or even instability circumstances in the result. Thus, controller should be designed corresponding to the actual plant which includes uncertainty. One of the suggested schemes to overcome the uncertainty issue is the joint of feedforward controller with feedback controller. Feedforward controller has the roll to enhance the performances of the closed-loop system considering the plant uncertainty. Following block diagram shows the joint of feedforward and feedback controller.

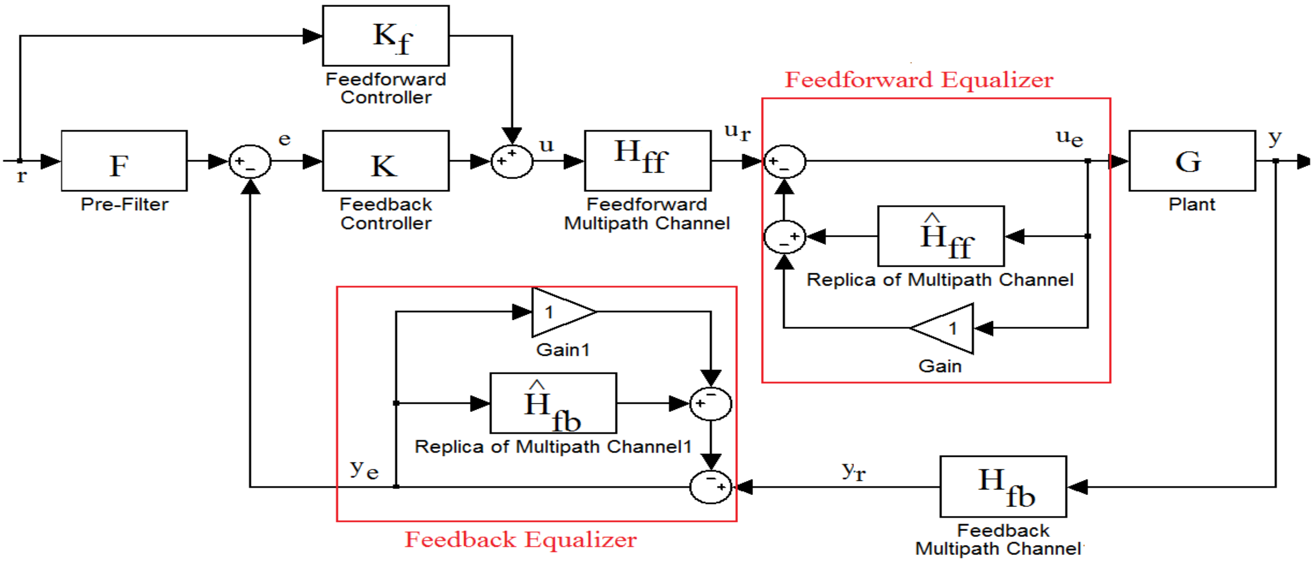


Fig.11 Combination of Feedforward and Feedback Controller in Tele-Control System with Equalizer

7 Adaptive Filters

An adaptive algorithm [6]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. The Least- Mean- Square (LMS) algorithm is the most widley used among various adaptive algorithm. 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). Before getting start the calculation, let us define the error signal in the adaptive filter e_f(n).

$$e_f(n) = h(n) * u(n) - \hat{h}(n) * u_e(n) \qquad (8)$$

According to stochastic gradient algorithm, we would have as follows. Here *n* and *i* are iteration and filter's tap number, respectively.

$$\hat{h}_i(n+1) = \hat{h}_i(n) - \mu \frac{\partial e_f^2(n)}{\partial \hat{h}}$$

Eventually, we obtain the following equation.

$$\hat{h}_i(n+1) = \hat{h}_i(n) - 2\mu e_f(n) u_e^T(n-i) \qquad (9)$$

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 [7], the LMS algorithm normalizes the step size with respect to the input signal power.

8 Simulation and Results

In order to evaluate the perforamnce and stability of the proposed method, we have simulated based on follwing condition.

$$\textit{Estimated} \quad \textit{plant} : \tilde{G}(s) = \frac{1}{s-2} \quad \textit{plant}: G(s) = \frac{1}{s-2} + W(s)$$

$$\textit{Additive uncertainity} \;\; W(s) = \frac{1}{s^2+2^2} + \frac{1}{s^2+10^2} + \frac{1}{s^2+100^2}$$

$$\textit{Feedback} \quad \textit{Controller} : K(s) = \frac{10s+1}{s}$$

$$Q(s) = \frac{1}{s^2+7s+1}, \quad \mathrm{Pr}e-\textit{filter} \quad F(s) = Q(s)$$

$$\textit{Feedforward} \quad \textit{Controller} : K_f(s) = Q(s)\tilde{G}^{-1}(s)$$

$$\cdot \ \text{ChannelSpecification (5 taps, M = 5)} \; H(s) = \sum_{i=1}^{M} \alpha_i e^{-sL_i}$$

attenuated and time delay factor are indicated as follows.

$\alpha_i = rand(i)e^{\frac{-0.1i}{M}}$, $\tau_i = 0.5i \times sort(|randn(i)|)$, *(i=1 to 5).*

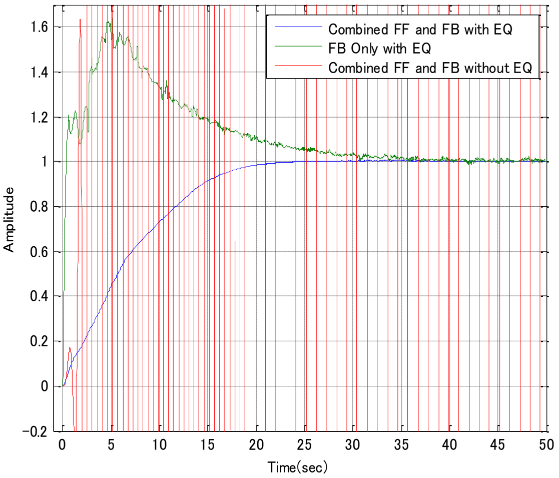


Fig.12 Comparison of proposed method (combined FF and FB with EQ) and without EQ

As it is clear in fig.12 system without equalizer becomes unstable due to channels existence. Thus, implementation of equalizer is implemented to stabilize the c,losed loop system. Moreover, combined feeforward and feedback controller enhance the performance of closed loop system even system contains uncertainty in plant.

9 Conclusion

In this paper we have proposed to implement equalizer in Tele-control system in order to get rid of instability and performance degdration of the closed-loop system which cuased by multipath channel. As a result the closed-loop system is asysmptoticaly stable. Moreover, combination of feedforward control and feedback control is propsed to enhance the performanmce when plant contains uncertainty. As a future wrok, in order to enhance the perofmance, more fast adaptive filter should be considered. Thus, improvement of adaptive filter is one of the significant future works.

REFERENCES

[1] R.C.Dorf, R. H .Bishop "Modern d ControlSystem",Prentice Hall2002

[2] WitoldPedrycz, "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., SilvaAniruddhaDatta, 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.

Performance Analysis of Broadcasting Real-Time Data Placement over Wireless Multi-Channels

Ding-Jung Chiang[1], Ching-Sheng Wang[2], Chien-Liang Chen[2], and Wen-Jay Lo[3]
[1]Department of Digital Multimedia Design, Taipei Chengshih University of Science and Technology
Taipei, Taiwan, R.O.C
[2]Department of Computer Science and Information Engineering, Aletheia University
New Taipei, Taiwan, R.O.C
[3]Exploration and Development Research Institute, CPC Corporation
Miaoli, Taiwan, R.O.C
Email:dingjung.chiang@gmail.com;cswang@mail.au.edu.tw;clyde@cobra.ee.ntu.edu.tw;155276@gmail.com

Abstract—*In this paper composite mobile environments, modeled by overlap service areas are taken for performance analysis of a wireless multiple channel link. The accurately approximated Gamma distributions is used to obtain analytical expressions for performance metrics such as average real-time miss rate and outage probability. The equivalent probability distribution function of client request to channel hopping is approximated. In the analysis multiple channel model based on currently transmitting is considered over independent and not identically distributed wireless channels. In addition the performance of currently transmitting is compared with the mobile user mobility. The numerical results are validated by Monte-Carlo simulations. It is shown that for broadcasting real-time data placement, this contribution is very useful and efficient for exact performance analysis and design of wireless multi-channel links.*

Keywords: Wireless Multi-Channel, Real-time Miss Rate, Broadcast Real-Time Data Placement,Monte-Carlo simulation

1. Introduction

Wireless transmission in wireless networks have been researched in the recent period due to its advantages over traditional networks in terms of deployment and scalability. Broadcast delivery [1] has been proposed and proven to be an efficient way of disseminating data to the mobile client population. This is due to the asymmetric nature of wireless communication, i.e., the downlink bandwidth is much higher than the uplink bandwidth. Associated with broadcast delivery is the problem of how to schedule the broadcasting of the requests to minimize the wait time of the clients. The wait time is also referred to as mean data access time, which is the average amount of time from the arrival of a request, to the time that the requested page is broadcast. With broadcasting, the server can satisfy all pending requests on a data item simultaneously, thus, eliminating the potentially very large overhead of data requests, and saving both the wireless bandwidth and a mobile client's battery energy. Another feature is that it greatly increases the scalability of the broadcast system by keeping the server from being swamped with data requests.

With the rapid growth of time-constraint information services and business-oriented applications, there is an increasing demand to support quality of service (QoS) in mobile environments. In many situations, user requests are associated with time constraints as a measure of QoS. These constraints can be imposed either by the users or the applications. For example, the timing of buying/selling stocks for a stock holder is very crucial. If the stock information cannot reach a stock holder in time, the information might become useless. For another example, the information about traffic congestion that is caused by a traffic control should also reach a mobile client heading toward this direction timely. If a client receives such information early enough, the client is able to react accordingly to avoid the traffic jam. The value of the information would degrade significantly when the client gets closer to the spot of the control. With data broadcasting approach a broadcast server can serve many mobile clients simultaneously. Therefore, data broadcasting is usually adopted for disseminating data in mobile computing environments.

Most of the related current research focuses on a data broadcasting approach, where the transmission of data is done without considering the data items with time constraints. In this study, we present an on-line scheduling algorithm to maximize the total number of satisfied users in asymmetric communication environments with time requirements. This is achieved by means of dynamic adaptation of the broadcast program to the needs of the users, taking into account the bandwidth constraints inherent in asymmetric communication environments and the deadline requirements of the user requests. The goal of our research is to study broadcast scheduling strategies on the multichannel systems for data broadcast with timing constraints. In such a broadcasting environment, the goal is to determine how well the scheduling algorithms ensure that the database server does not miss deadlines, instead of minimizing wait time. There are several scheduling algorithms for multichannel systems in mobile environments [2]. However, we demonstrate that traditional well-known algorithms do not always perform the best in a mobile environment, such as greedy and dynamic programming, when they are applied with time constraints on the multichannel systems in a mobile environment. We propose a model of a multichannel broadcast system for a simulation analysis and also propose an efficient scheduling algorithm called dynamic adjustment with time constraint (DATC).

The remainder of paper is organized as follows: Section 2 briefly describes the preliminary and related work. In Section 3 expressions for problem formulation and scheduling algorithm for real-time data placement. Section 4 provides performance results and discussion. Finally, in Section 5, we summarize the conclusions of our studies.

2. Preliminary and Related Work

Scheduling of transactions for real-time databases in a non-mobile environment is studied extensively in [1]. A real-time client/server model is considered in which the server assigns priorities to transactions based on several strategies, including Earliest Deadline First (EDF) and Least Slack (LS) first. As its name implies, for EDF the transaction with the earliest deadline is given the highest priority. For LS, the slack time is defined as: $d-(t+E-P)$, where d is the deadline, t is the current time, E is the execution time and P is the processor time used thus far. If the slack time is ≥ 0, it means that the transaction can meet its deadline if it executes without interference. The slack time indicates how long a transaction can be delayed and still meet its deadline. The Least Slack LS differs from EDF because the priority of a transaction depends on the service time it has received. If a transaction is restarted, its priority will change. Simulation results show that the EDF is the best overall policy for real-time database systems in a non-mobile environment. However, when system loads are high, the LS and EDF strategies lose their advantage, even over FCFS, as most transactions are likely to miss their deadlines.

For push-based systems, the longest wait first LWF algorithm has been shown to outperform all other strategies at minimizing wait time [3]. In LWF, the sum of the total time that all pending requests for a data item have been waiting is calculated, and the data item with the largest total wait time is chosen to broadcast next. However, LWF has been recognized as expensive to implement. In [4] a strategy, called requests times wait (RxW), is presented for push-based systems that makes scheduling decisions based on the current state of a queue (instead of access probabilities). The RxW algorithm provides an estimate of the LWF algorithm by multiplying the number of pending requests for a data item times the longest request wait time. In general, the performance of the approximate algorithms has been shown to be close to LWF.

There has been some research work to consider broadcasting for mobile real-time systems [5]. A push-based protocol for organizing broadcast disks for real-time applications, called Adaptive Information Dispersal Algorithm (AIDA). In this work, the data must be broadcast periodically to satisfy the timing constraints. The AIDA protocol considers fault-tolerance and the data items are allocated to the broadcast disks to minimize the impact of intermittent failures by utilizing redundancy. AIDA guarantees a lower bound on the probability of meeting timing constraints. Similar work addressing fault-tolerant real-time broadcast disks appears in [2]. In this work, the authors show that designing strategies for real-time broadcast disks is related to pinwheel scheduling. The authors derived a pinwheel algebra, which utilizes rules that can be used to construct fault-tolerant real-time broadcast disks. Their work differs from our work because we assume that we schedule all data items with time constraints using adaptive algorithms under limited bandwidth to minimize miss rate. In the multichannel broadcast disks model, the server periodically repeats a computed broadcast program, based on user access patterns. A broadcast cycle is defined as one transmission of the periodic broadcast program. Deadline constraints have been integrated into the broadcast model in [6]. In order to minimize the total number of deadlines missed by making the most effective use of the available bandwidth, scheduling approach has to focus on critical factors such as access frequency, time constraint, and bandwidth requirements. In [7], scheduling mechanisms for broadcasting data that are to minimize the delay incurred by insufficient channels, but it is reasonable that all clients are satisfied with an expected time to optimize average access time.

3. Problem Formulation and Proposed Method

We now describe a framework to support the push-based broadcast scheduling problem with time constraints. In this section, the real-time scheduling problem, system architecture and solving mechanism are introduced.

3.1 System Architecture

Server side: We assume that there are K channels in a broadcast area, each channel denoted $C_i, 1 \leq i \leq K$. A database is made up of N unit-sized items, denoted $d_j, 1 \leq j \leq N$. Each item is broadcast on one of these channels, so channel C_i broadcasts N_i items, $1 \leq i \leq K, \sum_{i=1}^{K} N_i = N$. Each channel cyclically broadcasts its items. Time is slotted into units called ticks. The size of data item is fixed and equal to one tick. Each data item is denoted $d_i\left(id_i, t_i, p_i\right)$ by the following parameters[2]:

- id: identifier of data item.
- t_i: relative deadline, i.e. the maximum acceptable delay for its processing.
- p_i: access probability for d_i.

Requests are for single item and assumed to be exponentially distributed.

Client side: Each client can require one data item per request associated with a time constraint. When a client needs a data item, it first tunes in the broadcast channels to retrieve the contents of channel. By the information of channels, the client can determine whether he can get the data item from the broadcast channels. If the needed data item is in the broadcast data set, the client tunes in the broadcast channels and retrieves the desired data item. Otherwise, the client sends a request to the server via the uplink channel, and listens to the broadcast channels to retrieve the data pages.

3.2 Problem Formulation

We formulate our problem to make it a resolvable problem as follows. Given a number of data items N to be broadcast in multiple broadcast channels K. Each data item is associated with a time constraint. Every access of a client is only one data item. Expected delay, w_i,is the expected number of ticks a client must wait for the broadcast of data item d_i. Average expected delay is the number of ticks a client must wait for an average request and is computed as the sum of all expected delays, multiplied by their access probabilities:

$$Average\ Expected\ Delay(W) = \sum_{i=1}^{N} w_i p_i \qquad (1)$$

,where w_i is expected delay[8] and p_i is access probability for data item d_i respectively. With time constraints, a request for data item d_i has missed its deadline when timing fault(expected delay for data item d_i exceeds its time constraint $t_i < W$) occurred at some time slot. The miss rate of all data items is defined as follows:

$$Miss\ Rate = \sum_{j=1}^{K} \sum p_i \qquad (2)$$

Our goal is to broadcast all data items with time constraints on multiple broadcast channels that minimizes the miss rate.

3.3 Design of Algorithm DATC

We provide an algorithm to generate a valid broadcast program so as to minimize miss-rate. If miss-rate is zero, let average access time minimized. We formulate our problem and make appropriate assumptions to make it a resolvable problem as follows.

Let each item contains two attributes: access probability and time constraint. Given a database D with its size $|D| = N$ and the number of channels = K, we aim to allocate each item in D into K channels, such that N items are cyclically broadcast on multi-channel, the miss rate can be written as: $\sum_{i=1}^{K}\left(\sum_{d_i \in c_i} p_i\right)$. Given an example using above the assumptions, we make a comparison between greed algorithm [9] and our algorithm DATC.

Algorithm 1 DATC(int N, int K, float P, int T)

```
  {$N$: number of items, $K$: number of channels}
  {$P$: access probabilities, $T$: time constraints}
Require: $N$ unit sized items ordered by popularity.
Ensure: $K$ partitions to minimize miss-rate.
  Measure the priority of data items by definition 3;
  Partition_number = 1;
  while Partition_number $< K$ do
    {$C^d_{ij}$ is computed as the expected delay of a data item
    in a channel of size $j - i + 1$}
    for each partition $k$ with data items $i$ to $j$ do
      for $(s = i; s \leq j; s = s + 1)$ do
        if $((s = i)$ or (Local_change $> C^d_{ij}$)) then
          Local_S = $s$;
          Local_change = $C^d_{ij}$;
        end if
        if ((k=1)or(Global_change>Local_change)) then
          Global_change = Local_change;
          Global_S = Local_S;
          Best_part = $k$;
        end if
        Split partition Best_part at point Global_P;
        Partition_number = Partition_number + 1;
        for data items on each channel do
          Minimize miss rate on each channel;
        end for
      end for
    end for
  end while
```

Item	d_1	d_2	d_3	d_4	d_5
AP	0.190	0.149	0.114	0.099	0.089
T	10	3	3	8	6

Item	d_6	d_7	d_8	d_9	d_{10}
AP	0.085	0.071	0.076	0.065	0.062
T	2	9	7	8	10

AP: Access Probability
T: Deadline

Given K = 3 broadcast channels, consider a set of N = 10 data items, $\{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}\}$, with the following skewed access probabilities and time constraints. Figure 1 illustrates the broadcasting program using the above example on multichannel, and Figure 2 shows the adjustment results after refining by the proposed algorithm.

4. Experimental Result and Performance Analysis

4.1 Simulation Environment

In our real-time broadcast simulation model, bandwidth is not explicitly modeled. Instead, similar to previous work [5], we use broadcast ticks as a measure of time. The greatest advantage of this approach is that the results are not limited to any particular bandwidth and/or data item size. Rather, it aims to capture the fundamental characteristics of the systems. The model simulates a one-hop wireless network. All data items are stored in a data server in a fixed location. Mobile clients need to send requests to the server via an uplink back-channel before the requested page can be broadcast. The arrival of requests generated by mobile clients follows a Poisson process and the inter-arrival time is exponential with mean λ. Each request has a request id, arrival time and deadline. For each page, a queue is maintained to store the information about requests on the object. We assume the results produced after a deadline are useless (firm deadlines), so all requests that have missed their deadlines are discarded. Mobile clients are responsible for re-sending requests when link errors occur. We also assume a deadline can capture the mobility of clients who are no longer able to receive the broadcast. In our model, since newly generated data requests are sent to the server immediately, the request generating time is equal to the time the server receives it (assuming network delay is ignored). We also ignore the overheads of request processing at the server, because the main purpose of the model is to compare the scheduling power of various strategies. We assume requests generated by mobile clients are read only, and no update request is allowed. Concurrency control issues are not our main concern, and thus, not considered. At each tick of the simulation clock, the following occurs. A simulated request generator generates requests with exponential inter-arrival time. The information about each request id, arrival-time and deadline is recorded. The request is then inserted to the corresponding queue. The server checks the deadlines of all the arrived requests, and discards those requests that have missed their deadlines. Then the server selects a page to broadcast by applying a scheduling strategy and starts to broadcast the selected page. All requests requesting the page are satisfied when the broadcast is finished. A client can request multiple pages and a page can be requested by multiple mobile clients at a time. We assume that data demand probabilities p_i follow the Zipf [10] distribution in which:

$$p_i = \frac{(1/i)^\theta}{\sum\limits_{i=1}^{M} (1/i)^\theta}, \quad (i = 1, 2, 3, \ldots, M)$$

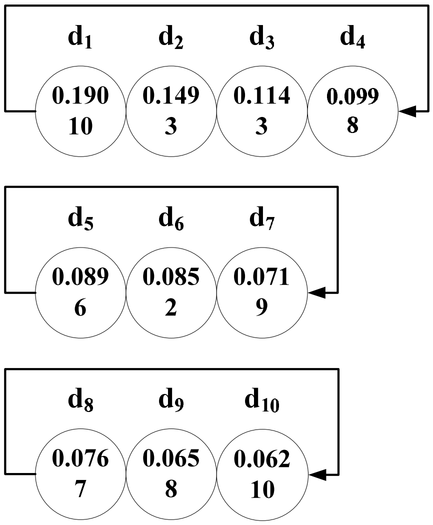
where p_i represents the i'th most popular page. The Zipf distribution allows the pages requested to be skewed. Figure 3 shows the results of our simulation comparing the DACT strategy to the strategy EDF for uniformly distributed deadlines.

4.2 Simulation Parameters

We compare the DATC approach with the algorithms described in Section 2: EDF, LSF, RxW, and LWF. Figure 3 illustrates the distribution of miss rate. We only choose the push-based algorithms to compare the results since we believe these push-based algorithms better adapt to the dynamic changes of the intensity and distribution of system workloads. The push-based (access probabilities, broadcast histories, etc.) off-line algorithms are not considered due to the fact that they are mainly for fairly stable systems. We implement the simulation model described in the previous section using C++. In each experiment, we run the simulation for 5000 time units, and we use an average of 20 runs of each simulation as the final result. The Parameters used in this simulation are summarized in Table 1. The default total number of data pages stored in the server, referred to as DBSIZE, is 100 pages. Client requests reach the system with exponential inter-arrival time with mean λ, and λ is varied in our simulation from 2 - 60. It is assumed that each data request requires 1 broadcast tick to broadcast. An open system model is used to simulate the system for extremely large, highly dynamic populations. Data access follows a skewed Zipf distribution with parameter Θ to control the skewness. The minimum slack time is 10, with the maximum slack time ranging from 20 to 300. This variation in maximum slack time allows us to vary the tightness in the deadlines. In addition to a uniform distribution of deadlines, an exponential distribution is utilized with lambda ranging from 10 to 300. After doing a large number of experiments with various factors that affect the performance, we come up with an overall performance comparison between the previous algorithms and our scheduling algorithm DATC in Table 2. We grade the level of performance from 1 to 5. The higher the degree is, the better the performance is. On the contrary, overhead with higher degree shows that the algorithm gets more cost in Table 2.

5. Conclusion

Wireless broadcast is a promising data dissemination method to improve system scalability and deal with dynamic data access pattern. In this paper, we have investigated a number of previous scheduling algorithms in real-time data broadcast environment. A new heuristic



For program 1, broadcast cycle = 4

Time constraint of d_2 = 3 and d_3 = 3 < broadcast cycle = 4

Miss rate = 0.149 + 0.114 = 0.263

For program 2, broadcast cycle = 3

Time constraint of d_6 = 2 < broadcast cycle = 3

Miss rate = 0.085

For program 3, broadcast cycle = 3

Time constraint of d_8 , d_9 and d_{10} > broadcast cycle = 3

No miss rate

Total miss rate = 0.263 + 0.085 = 0.348

Fig. 1: Broadcasting 10 data items with time constraints to partition 3 disjoint subsets and the miss rate = 0.348.

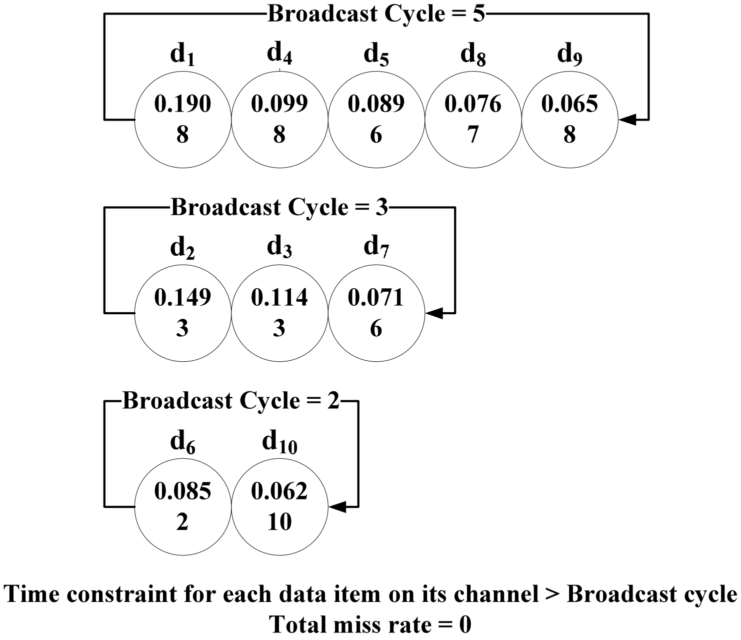


Fig. 29 Using DATC algorithm to partition 3 disjoint subsets and the miss rate = 0.

Table 1: Simulation Parameters

Symbol	Description	Default	Range	Unit
DBSIZE	Total number of data pages stored in server	100	100-10000	pages
λ	Mean request arrival rate (exponential)	-	2-60	requests/tick
θ	Request skewness (Zipf)	1.0	0.0-1.0	-
$MinSlack$	Minimum slack time	1.0	-	ticks
$MaxSlack$	Maximum slack time	-	20-300	ticks
$\lambda_{deadline}$	Parameter of exponential deadline distribution	-	10-300	ticks

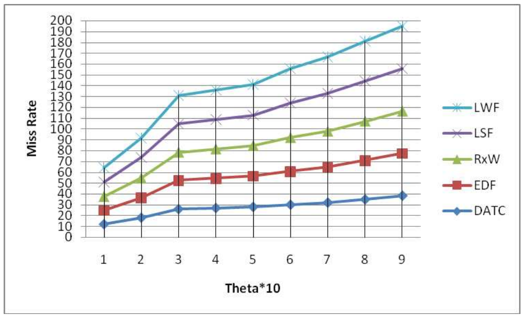


Fig. 3: Deadlines missed

Table 2: Performance comparison of different scheduling algorithms

	DMR	ART	AS	Overhead
LWF	1	2	3	5
LSF	1	1	2	5
RxW	3	5	4	4
EDF	3	3	1	4
DATC	5	5	4	3

DMR = Deadline Miss Rate
ART = Average Response Time
AS = Average Stretch

online scheduling algorithm is proposed for real-time broadcast in heterogeneous settings. Our approach is developed by taking various factors critical to performance into account, including channel number, deadline and frequency of requests. We have conducted a series of simulation experiments to evaluate the performance of our approach. The results demonstrate that our algorithm substantially outperforms other algorithms in terms of deadline, response time and stretch. The results also show that the overhead of our algorithms is low compared with the other scheduling algorithm, while a balanced performance can be maintained. In the further study, since scheduling and client cache management affect each other, we will investigate client cache management schemes to improve data availability to the clients. Prefetching techniques can be combined with the cache policy to reduce the request response time. Furthermore, we will investigate the scheduling problem when clients request more than one data items at a time

Acknowledgements. We are grateful to the support of National Science Council, R.O.C. under contract number 102-2221-E-149-006.

References

[1] S. Bodas, S. Shakkottai, L. Ying, and R. Srikant, "Scheduling in multi-channel wireless networks: Rate function optimality in the small-buffer regime," *IEEE TRANSACTIONS ON INFORMATION THEORY*, vol. 60, no. 2, pp. 1101–1124, February 2014.

[2] F. Adelstein, S. K. Gupta, G. G. R. III, and L. Schwiebert, *Fundamentals of Mobile and Pervasive Computing*. McGraw-Hill, 2005.

[3] T. AktasÂÿ, A. O. zguÂÍr YÄślmaz, and E. Aktas, "Practical methods for wireless network coding with multiple unicast transmissions," *IEEE TRANSACTIONS ON COMMUNICATIONS*, vol. 61, no. 3, pp. 1123–1133, March 2013.

[4] A. Boukerche and H. O. II, "Media synchronization and qos packet scheduling algorithms for wireless systems," *ACM Mobile Networks and Applications*, vol. 10, no. 1-2, pp. 233–249, February 2005.

[5] D.-J. Chiang, T. K. Shih, and C.-L. Chen, "Disseminating data with time constraint based on multichannel over ubiquitous computing environments," *WORLD WIDE WEB-INTERNET AND WEB INFORMATION SYSTEMS*, vol. 14, pp. 223–241, 2011.

[6] L. M. D. and M. A. C., "Quality of service management in ip networks through dynamic service rate reconfiguration," *Journal of Internet Technology*, vol. 7, no. 1, pp. 45–57, January 2006.

[7] Y. Dhungana and C. Tellambura, "Uniform approximations for wireless performance in fading channels," *IEEE TRANSACTIONS ON COMMUNICATIONS*, vol. 61, no. 11, pp. 4768–4779, November 2013.

[8] R. E. KALMAN, "A new approach to linear filtering and prediction problems," *Transactions of the ASMEâĂŞJournal of Basic Engineering*, vol. D, no. 82, pp. 35–45, 1960.

[9] T. H. Cormen, C. E. Leiserson, and R. L. Rivest., *Introduction to Algorithms*. The MIT, 1992.

[10] G.K.Zipf., *Human Behaviour and the Principle of the Least Effort.* Addison-Wesley, 1949.

A Transmission Parameter Optimization Method for GA-based Cognitive Engines

Keunhong Chae[1], Huaping Liu[2], and Seokho Yoon[1,◇]
[1]College of Information and Communication Engineering, Sungkyunkwan University, Suwon 440-746, Korea
[2]School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331 USA
◇Corresponding author

Abstract— *This paper deals with a transmission parameter optimization method for the cognitive engine based on the genetic algorithm (GA). The performance of the secondary user (SU) system would be improved if a frequency band with a good channel status is used, and thus, we consider the frequency band selection as a new transmission parameter to be optimized. Then, four single objective fitness functions are designed with a transmission set including the new transmission parameter, and finally, a multiple objective fitness function is proposed via a weighted sum of the four single objective fitness functions. Numerical results demonstrate that we can obtain transmission parameter sets optimized for given transmission scenarios with the GA-based cognitive engine incorporating the proposed objective fitness function.*

Keywords: Transmission parameter; Genetic algorithm; Fitness function

1. Introduction

Due to the advent of various wide-band wireless communication systems such as long term evolution (LTE), IEEE 802.16, and digital video broadcasting (DVB), frequency spectrum has become one of the scarcest resources. As a spectrum-efficient technology, the dynamic spectrum access (DSA) has been proposed by virtue of the software defined radio (SDR) capable of tuning its transmission parameters [1]-[3], where underused frequency bands are utilized opportunistically.

In DSA systems, a secondary user (SU) adjusts its transmission parameters (e.g., transmit power, modulation index, and transmission bandwidth) by observing the surrounding environments. Specifically, the SU first determines if a primary user (PU) is utilizing the spectrum bands of interest via the spectrum sensing [4]. Then, the transmission parameters of the SU are optimized by an intelligent signal processing unit referred to as a cognitive engine. The implementation of the cognitive engine has been studied mainly based on the artificial intelligence (AI) techniques such as the genetic algorithm (GA), expert systems, neural networks, and case-based reasoning [5]. Especially, the GA-based cognitive engine has attracted much interest since it is capable of self-evolving (similarly to the human cognition process) unlike other AI-based cognitive engines [6].

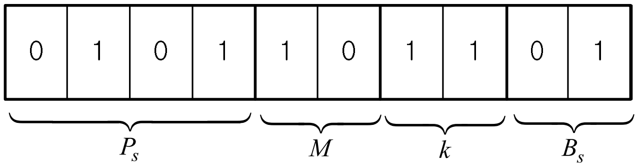


Fig. 1: An example of a chromosome structure representing the transmission parameters.

Although several GA-based cognitive engines [6]-[10] have been developed, the conventional engines do not consider any optimization in choosing one out of multiple vacant frequency bands. The channel status might be good or poor depending on the frequency band, and thus, it could improve the performance of the SU system to use a vacant frequency band with a good channel status. Thus, in this paper, including the frequency band choice problem in the transmission parameter optimization, we design a multiple objective fitness function to be used in the GA-based cognitive engine.

The rest of this paper is organized as follows. In Section II, we describe the GA-based cognitive engine and the associated objective fitness function. Section III proposes a multiple objective fitness function with a new transmission parameter set. Section IV shows that the GA-based cognitive engine with the proposed multiple objective fitness function can offer optimized transmission parameter sets for given transmission scenarios. Finally, conclusion is drawn in Section V.

2. System model

Transmission parameters are variables to be optimized based on the information from ambient parameters including the noise density and the test statistic used in the spectrum sensing. In this paper, we consider the following transmission parameters: The transmit power P_s of SU, the modulation index M, the bandwidth B_s of the SU signal, and the frequency band index k to be selected out of multiple vacant frequency bands.

To optimize the transmission parameters based on the GA, we first construct a chromosome structure with a bit stream representing the values of the transmission parameters. For example, the chromosome structure shown in Fig. 1 is composed of a bit stream with a length of 10 bits, where 4, 2, 2, and 2 bits are used to represent the values of P_s, M, k, and B_s, respectively, and thus, P_s, M, k, and B_s have 2^4, 2^2, 2^2, and 2^2 candidate values, respectively. Then, a multiple objective fitness function f is defined as

$$f = a_1 f_1 + a_2 f_2 + \cdots + a_L f_L \qquad (1)$$

to select the most suitable transmission parameters for a transmission scenario of interest, where $\{f_l\}_{l=1}^{L}$ are L single objective fitness functions and each of which represents a performance measure with the transmission parameters to be optimized, and a_l denotes a weight value for f_l with $\sum_{l=1}^{L} a_l = 1$ [9]. The weight value represents the priority of a single objective fitness function in a given transmission scenario, and thus, a higher (lower) weight value is assigned to the single objective fitness function with a higher (lower) priority. Through some operations such as the selection, crossover, and mutation, finally, the GA-based cognitive engine finds an optimum transmission parameter set maximizing the multiple objective fitness function [6].

3. Proposed fitness function

In this section, we design four single object fitness functions making up the multiple object fitness function, each of which corresponds to the bit error rate (BER), throughput, interference reduction, and frequency band selection, respectively, and has the transmission parameters (P_s, M, k, and B_s) to be optimized as its factors.

3.1 The single objective fitness function for the BER

A single objective fitness function f_{BER} is designed as

$$f_{\text{BER}} = \frac{\log_{10}(0.5) - \log_{10}(P_b)}{\log_{10}(0.5) - \log_{10}(P_{b,\min})}, \qquad (2)$$

where $P_{b,\min}$ is the minimum value of the SU BER $P_b = \frac{2P_s}{B_s \times \log_2(M) \times N_0}$ with N_0 denoting the noise density. From (2), we can easily see that f_{BER} has the maximum (minimum) value 1 (0) when $P_b = P_{b,\min}$ ($P_b = 0.5$): It should be noted that $P_b = 0.5$ is the worst case.

3.2 The single objective fitness function for the throughput

The throughput of the data transmission is proportional to the modulation index M, thus, a single objective fitness function $f_{\text{throughput}}$ can be expressed as

$$f_{\text{throughput}} = \frac{\log_2(M) - \log_2(M_{\min})}{\log_2(M_{\max}) - \log_2(M_{\min})}, \qquad (3)$$

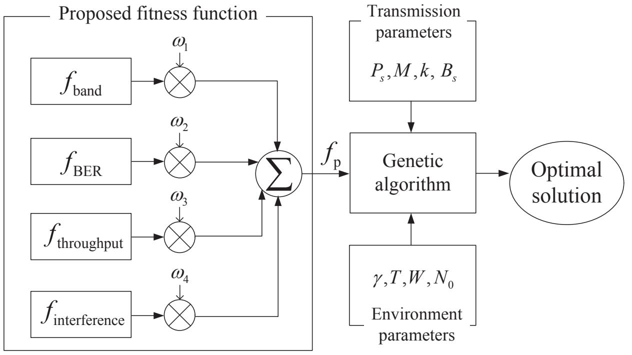


Fig. 2: The overall process of the proposed transmission parameter optimization method.

where $M_{\max}$ and $M_{\min}$ are the maximum and minimum values of M, respectively. The function is maximized (minimized) when $M = M_{\max}$ ($M = M_{\min}$).

3.3 The single objective fitness function for the interference reduction

It is also desired to reduce interference to the PU signal, which depends on the power P_s and bandwidth B_s of the SU signal. Thus, we design a single fitness function $f_{\text{interference}}$ as

$$\begin{aligned} f_{\text{interference}} =& 1 - \frac{1}{2}\left(\frac{P_s - P_{s,\min}}{P_{s,\max} - P_{s,\min}}\right) \\ &- \frac{1}{2}\left(\frac{B_s - B_{s,\min}}{W(k) - B_{s,\min}}\right), \end{aligned} \qquad (4)$$

where $P_{s,\max}$ and $P_{s,\min}$ are the maximum and minimum values of P_s, respectively, $W(k)$ is the bandwidth of the kth band assigned to the PU, and $B_{s,\min} \leq W(k)$ is the minimum value of B_s.

3.4 The single objective fitness function for the frequency band selection

Since the spectrum sensing is performed before the operation of the cognitive engine, it is natural to assume that the test statistic value $T(k)$ and the threshold $\gamma(k)$ for the spectrum sensing of the kth band and the bandwidth $W(k)$ of the kth band are known. Although most of the candidate bands are detected as a vacant band by the spectrum sensing process, some of the bands may be occupied by the PU signal due to the missed detection of the spectrum sensing, and the frequency band is more likely to be vacant when the difference between $\gamma(k)$ and $T(k)$ is a larger value. Reflecting the observations, thus, we design a term $\left(\frac{D(k) - D_{\min}}{D_{\max} - D_{\min}}\right)$, where $D(k) = \gamma(k) - T(k)$ and $D_{\max}$ and $D_{\min}$ are the maximum and minimum values of $D(k)$, respectively. It is noteworthy that $D(k) > 0$ since the bands of interest are already detected as a vacant band (i.e., $\gamma(k) > T(k)$) by the spectrum sensing process. In addition, to choose a wide

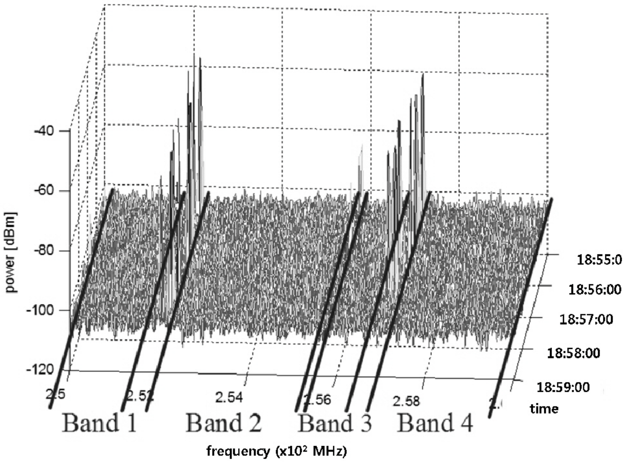


Fig. 3: The frequency spectrum of [250 MHz, 260 MHz] bands.

frequency band and to fully use the selected frequency band, we also design two additional terms $\left(\frac{W(k)-W_{\min}}{W_{\max}-W_{\min}}\right)$ and $\left(\frac{B_s-B_{s,\min}}{W(k)-B_{s,\min}}\right)$, where $W_{\max}$ and $W_{\min}$ are the maximum and minimum values of $W(k)$, respectively, and $B_{s,\max}$ is the maximum value of the bandwidth candidates B_s of the SU. Normalizing and combining the designed terms, we propose a single objective fitness function f_{band} as

$$f_{\text{band}} = \frac{1}{3}\left(\frac{D(k)-D_{\min}}{D_{\max}-D_{\min}}\right)+\frac{1}{3}\left(\frac{W(k)-W_{\min}}{W_{\max}-W_{\min}}\right) + \frac{1}{3}\left(\frac{B_s-B_{s,\min}}{W(k)-B_{s,\min}}\right). \tag{5}$$

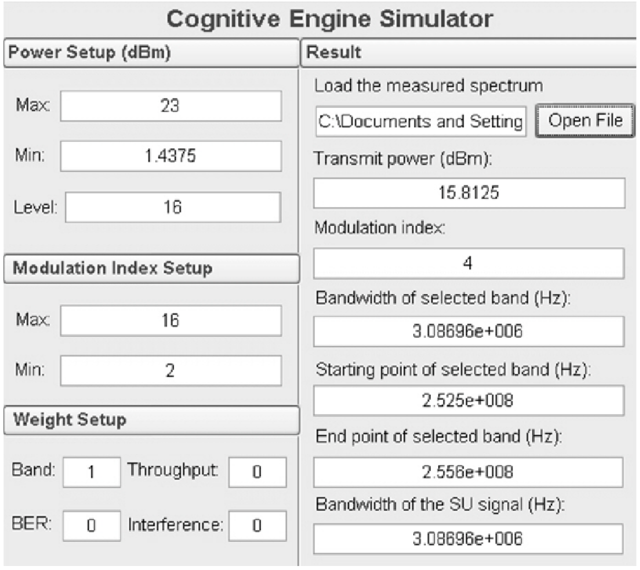
In summary, the function f_{band} is designed (i) to maximize the probability that the chosen band is vacant, (ii) to choose a band with a larger bandwidth, and (iii) to transmit the SU signal with a larger bandwidth.

3.5 The multiple objective fitness function

With (2), (3), (4), and (5), finally, we propose a multiple objective fitness function f_{p} as

$$f_{\text{p}} = w_1 f_{\text{band}} + w_2 f_{\text{BER}} + w_3 f_{\text{throughput}} + w_4 f_{\text{interference}}, \tag{6}$$

where $\{w_l\}_{l=1}^4$ are the weight values for fitness functions f_{band}, f_{BER}, $f_{\text{throughput}}$, and $f_{\text{interference}}$, respectively, and $\sum_{l=1}^4 w_l = 1$. The overall optimization process incorporating the proposed multiple objective fitness function is depicted in Fig. 2, where the proposed multiple objective fitness function is initially evaluated with a candidate set of the transmission parameters, and then, the GA continues to evaluate the remaining candidates by using the selection, crossover, and mutation operations until it finds the optimum candidate set maximizing the fitness function.



Cognitive Engine Simulator

Power Setup (dBm)

Max: 23

Min: 1.4375

Level: 16

Modulation Index Setup

Max: 16

Min: 2

Weight Setup

Band: 1 Throughput: 0

BER: 0 Interference: 0

Result

Load the measured spectrum

C:\Documents and Setting

Transmit power (dBm): 15.8125

Modulation index: 4

Bandwidth of selected band (Hz): 3.08696e+006

Starting point of selected band (Hz): 2.525e+008

End point of selected band (Hz): 2.556e+008

Bandwidth of the SU signal (Hz): 3.08696e+006

Fig. 4: The cognitive engine GUI simulator.

4. Numerical results

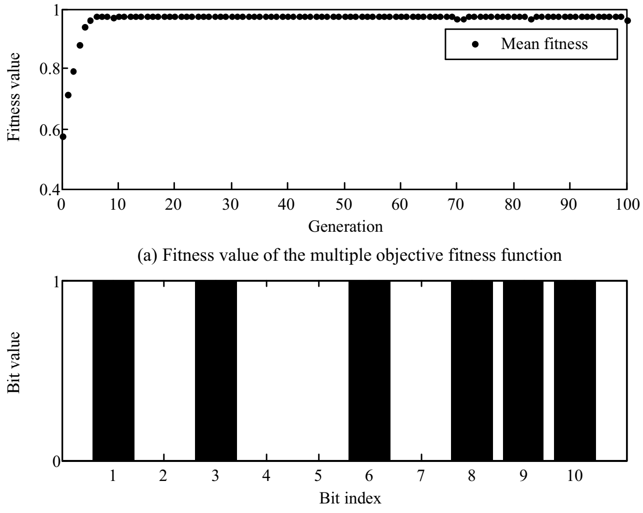
In this section, we demonstrate a cognitive engine simulator incorporating the proposed optimization method. We firstly measured a frequency spectrum of [250 MHz, 260 MHz] bands at the top of a mountain in Yongin city, Korea, and then, used the measured data as the input of the simulator. The measured spectrum is shown in Fig. 3, where four spectrum bands (Band 1 $\sim$ Band 4) are detected as the vacant bands, and the simulator is implemented using MATLAB graphic user interface (GUI) programming and its main screen is shown in Fig. 4.

For simulations, we use a chromosome structure with a length of 10 bits, where 4, 2, 2, and 2 bits are used to represent the values of P_s, M, k, and B_s, respectively, and the candidates for P_s, M, k, and B_s are given by $\{\frac{23}{16}, \frac{2\times23}{16}, ..., 23\}$ dBm, $\{2\,(\text{BPSK}), 4\,(\text{QPSK}), 8\,(\text{8PSK}), 16\,(\text{16QAM})\}$, {Band 1, Band 2, Band 3, Band 4}, and $\{10, 100, 500, W(k)\}$ kHz, respectively, and each candidate of a transmission parameter has its own bit representation as shown in Table 1. The noise density N_0 is set to the power spectral density of a frequency band with the lowest power over the spectrum range of [250 MHz, 260 MHz], and the threshold for the spectrum sensing performed via the energy detector [11] is determined to satisfy the false alarm probability of 0.01.

For transmission scenarios, we first consider the simplest case of $\bar{w} = [w_1, w_2, w_3, w_4] = [1, 0, 0, 0]$ to verify the simulator, and subsequently, we demonstrate the results for a scenario with the weight vector $[0.6, 0.2, 0.1, 0.1]$ as an example. Fig. 5 shows (a) the fitness function value and (b) the solution result of the simulator when $\bar{w} = [1, 0, 0, 0]$. From the figure, we can observe that the fitness value is converged as the number of the generation increases,

Table 1: Candidates of transmission parameters.

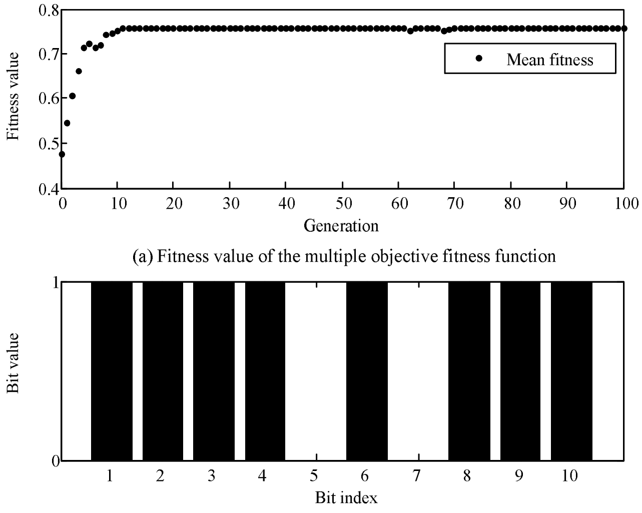
Transmission parameter	Value (bits)			
P_s [dBm]	$\frac{23}{16}$ (0000)	$\frac{2\times23}{16}$ (0001)	$\frac{3\times23}{16}$ (0010)	$\frac{4\times23}{16}$ (0011)
	$\frac{5\times23}{16}$ (0100)	$\frac{6\times23}{16}$ (0101)	$\frac{7\times23}{16}$ (0110)	$\frac{8\times23}{16}$ (0111)
	$\frac{9\times23}{16}$ (1000)	$\frac{10\times23}{16}$ (1001)	$\frac{11\times23}{16}$ (1010)	$\frac{12\times23}{16}$ (1011)
	$\frac{13\times23}{16}$ (1100)	$\frac{14\times23}{16}$ (1101)	$\frac{15\times23}{16}$ (1110)	$\frac{16\times23}{16}$ (1111)
M	2, BPSK (00)	4, QPSK (01)	8, 8PSK (10)	16, 16QAM (11)
k	Band1 (00)	Band2 (01)	Band3 (10)	Band4 (11)
B_s	10 kHz (00)	100 kHz (01)	500 kHz (10)	$W(k)$ Hz (11)



(a) Fitness value of the multiple objective fitness function

(b) The solution result in a form of a 10 bit stream

Fig. 5: Simulation results when $\bar{w}$=[1, 0, 0, 0].



(a) Fitness value of the multiple objective fitness function

(b) The solution result in a form of a 10 bit stream

Fig. 6: Simulation results when $\bar{w}$=[0.6, 0.2, 0.1, 0.1].

and also, from the fact that the 7th and 8th bits of the chromosome are '01' and the 9th and 10th bits are '11', we can see that Band 2 (the widest vacant band of Fig. 3) is chosen as the frequency band and the SU uses the whole spectrum of Band 2. It should be noted that the fitness function f_{band} is not a function of P_s and M, and thus, the transmit power (the 1st-4th bits) and the modulation index (the 5th and 6th bits) are randomly selected by the GA.

Fig. 6 shows (a) the fitness function value and (b) the solution result of the simulator when $\bar{w} = [0.6, 0.2, 0.1, 0.1]$. When all weight values are non-zero and the weight value w_1 is the largest, the scenario can be regarded as 'multimedia service mode' where the SU needs to transmit data using a wide vacant band. Thus, again, Band 2 is chosen as the frequency band and the SU uses the whole spectrum of Band 2 as in the case of $\bar{w} = [1, 0, 0, 0]$; however, for the transmit power P_s and the modulation index M, the maximum power of 23 dBm and QPSK modulation is selected, respectively, considering the fact that the weight value for f_{BER} is the second largest.

5. Conclusion

In this paper, we have proposed a novel transmission parameter optimization method for the GA-based cognitive engine. Considering that the performance of the SU system can be changed depending on the channel status of the selected frequency band, we have first designed four single objective fitness functions with a new transmission parameter set including the frequency band index, and then, constructed a multiple objective fitness function using the weighted sum of the designed single objective fitness functions. From numerical results, it has been confirmed that the GA-based cognitive engine incorporating the proposed objective fitness function provides optimized sets of the transmission parameters for given transmission scenarios.

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. Song, C. Xin, Y. Zhao, and X. Cheng, “Dynamic spectrum access: from cognitive radio to network radio," *Int. J. Next-Generation Networks*, vol. 19, no. 1, pp. 23-29, Feb. 2012.

[2] P. Yadav, S. Chatterjee, and P. P. Bhattacharya, “A survey on dynamic spectrum access techniques in cognitive radio," *Int. J. Next-Generation Networks*, vol. 4, no. 4, pp. 27-46, Dec. 2012.

[3] S. S. Nair, S. Schellenberg, J. Seitz, and M. Chatterjee, “Hybrid spectrum sharing in dynamic spectrum access networks," in *Proc. International Conference on Information Networking (ICOIN)*, pp. 324-329, Bangkok, Thailand, Jan. 2013.

[4] T. Yüech and H. Arslan, “A survey of specturm sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys & Tutorials*, vol. 11, no. 1, pp. 116-130, Mar. 2009.

[5] X. Dong, Y. Li, and S. Q. Wei, “Design and implementation of a cognitive engine functional architecture," *Chinese Science Bulletin*, vol. 57, no. 28-29, pp. 3698-3704, Oct. 2012.

[6] C. J. Rieser, “Biologically inspired cognitive radio engine model utilizing distributed genetic algorithms for secure and robust wireless communications and networking," Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, 2004.

[7] T. W. Rondeau, B. Le, C. J. Rieser, and C. W. Bostian, “Cognitive radio with genetic algorithms: Intelligent control of software defined radios," in *Proc. Software Defined Radio Forum Tech. Conf.*, pp. C3-C8, Phoenix, AZ, Nov. 2004.

[8] D. Maldonado, B. Le, A. Hugine, T. W. Rondeau, and C. W. Bostian, “Cognitive radio applications to dynamic spectrum allocation," in *Proc. IEEE Int. Symp. New Frontiers in Dynamic Spectrum Access Networks*, pp. 597-600, Baltimore, MD, Nov. 2005.

[9] T. R. Newman, B. A. Barker, A. M. Wyglinski, A. Agah, and J. B. Evans, “Cognitive engine implementation for wireless multicarrier transceivers," *Wirel. Commun. Mob. Comput.*, vol. 7, no. 9, pp. 1129-1142, Nov. 2007.

[10] L.-C. Wang, C.-W. Wang, and F. Adachi, “Load-balancing spectrum decision for cognitive radio networks," *IEEE J. Selected Areas in Commun.*, vol. 29, no. 4, pp. 757-769, Apr. 2011.

[11] H. Urkowitz, “Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523-531, Apr. 1967.

Ultra High Video Data Compression for Android Devices Using OpenCV and other Open-Source Tools

Ronald Yu
School of Information and Computer Sciences
University of California, Irvine
ronaly1@uci.edu

Tong Lai Yu
School of Computer Science and Engineering
California State University, San Bernardino
tyu@csusb.edu

WORLDCOMP'14:ICWN14, July 21-24, 2014, Las Vegas, Nevada, USA.

Keywords: Open-CV, Speech Recognition, 3D Graphics, Android TTS

Abstract

We describe in this paper how to use open-source resources, in particular OpenCV, to design and implement an Android application that achieves ultra-high video compression for special videos, which consist of mainly a human face and speech, such as the scene of a news announcement or a teleconference. Google Voice Recognition[30], which is a free and open Android tool, is utilized to convert the speech of the video to text. Human face images are classified by OpenCV (Open Source Computer Vision) [32] into a predefined number of common face features. Rather than saving the audio and image data of the video directly, we save the class of the image as metadata along with the speech text, which are compressed losslessly and transmitted to the receiver.

The receiver decompresses the encoded data to recover the speech text and the image metadata. The text is converted to speech by the Android Text-To-Speech (TTS) engine[1]. It renders a three-dimensional model of a human face, which is composed of polygon meshes[24] to animate the lip movements of human speech from the input text. Blender[8, 36], a popular open-source graphics suite, is employed to create 3D models and to save their mesh data in the COLLAborative Design Activity (COLLADA) format[26], which is also an open graphics format. The image metadata are used to determine which 3D model will be loaded for animation and instruct the renderer to switch to another model when the emotion of the original image changes.

Java language is used to develop a parser[29, 49] to extract coordinates of polygons from a COLLADA file and organize the data into a format that can be rendered effectively by OpenGL ES, the graphics rendering library used by Android. The producer-consumer paradigm is employed to synchronize the animated lip movements and the speech generated by the TTS[40, 43, 46]. Semaphores[40, 43] are used to ensure that the right thread of the image model is running.

1. Introduction

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[6] and the phenomenal 3D movie Avatar[5] are based on open-source software. It is a significant task to explore the usage of available open-source tools to develop software applications for research or for commercial use[47, 48]. The Android application reported in this paper is developed with free software resources, which are mainly open-source.

Mobile devices have become ubiquitous and in the last couple of 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[16]. In recent years the number of mobile applications has been growing with tremendous speed.

Video compression has been an ongoing research topic and has an unaccountable number of applications. Traditional method of video compression[23, 34] uses a domain transform technique such as Discrete Cosine Transform (DCT) to express an image in the frequency domain. The transformed coefficients are then quantized, reordered, run-length encoded and entropy encoded. Motion estimation (ME) and motion compensation (MC) techniques are used to reduce redundancies in the video data. In recent years, graphics techniques have been used to achieve very high compression ratio for special videos whose scenes are fairly static and mainly composed of human features; human speech is animated using 3D graphics models[38]. The video compression standard MPEG-4 also has specifications on facial animation for synthesized speech[33, 14].

Though speech simulation is still an ongoing research topic, it already has numerous commercial applications including game development and customer service[7], and it contributes to both the developments of acoustic and visual applications[13, 10].

Our work reported in this paper develops and merges speech recognition, image recognition, and audio and visual technologies into one application that can achieve ultra-high video compression by making use of open-source technology. Like the MPEG-4 facial animation, our application only works for videos consisting of solely facial images making speech such as news announcement.

In the encoding process, the speech of a video is converted to text by a speech recognizer. The image of each frame is classified into a limited number of types by OpenCV[32] and represented by a special string of letters, which is combined with the speech text (see below for more detailed explanations). The text data are then compressed losslessly by the Android compress package *java.util.zip*[2] that provides the *zip* and *gzip* functionalities for compression and decompression. The resulted bit stream is transmitted to the receiver or saved in a file. Figure 1 is a block diagram showing this encoding process.

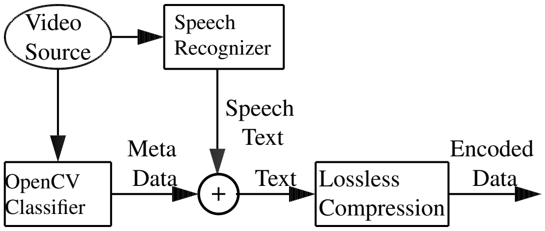


Figure 1. Encoder of Video Data

The decoder first decompresses the encoded stream into text. When it reads a word of image metadata, it loads the corresponding 3D graphics model for animation, otherwise the text is converted to speech by the Android Text To Speech (TTS) utility, which drives the animation of the facial image. The main tool we use for rendering graphics is OpenGL for embedded systems (ES). The graphics library OpenGL[39] is the industry standard for developing 2D and 3D graphics applications[3, 9], and OpenGL ES[27, 4, 31] is OpenGL modified for embedded systems. There is a major difference between OpenGL ES 1.X and OpenGL ES 2.X. While the 1.X version shares the same functionality and syntax of the traditional OpenGL APIs and, like early OpenGL, has a fixed pipeline and operates as a state machine, the 2.X version has adopted a programmable pipeline architecture that allows users to program vertex and fragment shaders[28, 37], the equivalent of OpenGL Shading Language (GLSL)[20]. The vertex shader is responsible for processing geometry. The fragment shader works at the pixel level, processing incoming fragments to produce colors including transparency. Figure 2 is a block diagram showing this decoding process.

The encoder of this application uses Google Voice Recognition (GVR)[30], which is based on neural network algorithms to convert human audio speech to text. GVR works for a number of major languages but we have only considered English in our application. 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.

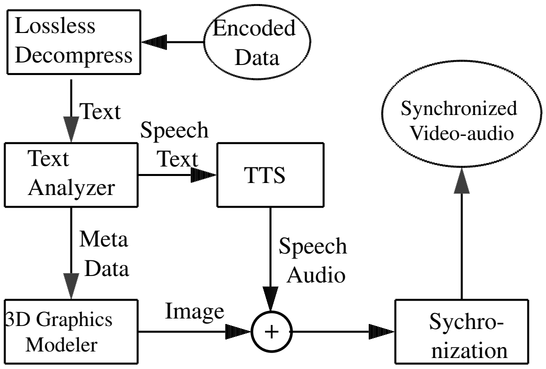


Figure 2. Decoder of Video Data

In this application, GVR uses the Internet to access 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.

In parallel to converting speech signal to English text, the encoder makes use of OpenCV[32], an open-source BSD-licensed library that includes several hundreds of computer vision algorithms, to classify the image of a video frame. The information is presented as text using special character strings. OpenCV not only supports desktop platforms such as MS Windows and Linux but also Android OS for running mobile devices.

The decoder of the application has to render graphics in a mobile device, which is characterized by a small display size[35], limited memory capacity and limited computing power. All of these aspects affect the graphic animation experience of the mobile user. These limitations make the design and implementation of a TTS animation application in a mobile device very different from that of an application running on a desktop PC.

Another problem of the decoder one must address is the audio-video synchronization. For traditional video compression of natural scenes, MPEG standard uses timestamps to synchronize audio and video streams[23]. MPEG-4 also addresses coding of digital hybrids of natural and synthetic, aural and visual information[33, 38]. Doenges et al. mentioned in their paper[14] that special attention must be paid to the synchronization of acoustic speech information with coherent visible articulatory movements of the speakers mouth in MPEG-4 synthetic/natural hybrid coding (SNHC) for animated mixed media delivery. However, they did not present the details of synchronization in the paper. Our synchronization problem of video and audio is different from that of MPEG-4 as the animation is driven by the text content. Therefore, we do not use timestamps to synchronize audio and video. Instead, the synchronization is done using the producer-consumer paradigm[41], which works effectively in this situation.

The application is developed for Android-based mobile devices. Android provides a Text-to-Speech (TTS) engine (PICO) with limited APIs[1]. After lossless decompression, the main thread of the decoder presents the text to the speech simulator that plays the sound using the Android TTS APIs and renders the corresponding visemes while performing a lip-synchronization action, keeping the audio and video synchronized. Visemes, which can be considered as visual counterpart of phonemes in audio, are visually distinct mouth, teeth, and tongue articulations for a language.

Besides the main thread, the application has a few other threads. One of them is responsible for voice synthesis and speech simulation by making use of the Android Text-to-Speech(TTS) engine[1]. Another thread controls the 3D rendering and animation of a human head. This thread implements the OpenGL ES function calls and has to decide which object to render based on the input data. The third thread is the input text thread that handles the insertion of the data into a text buffer. This thread implements the producer in the Producer-Consumer problem.

Each 3D Graphics model, which corresponds to an image emotion and gender type is controlled by a thread. If there are 16 image types, there will be 16 such threads running concurrently; however, only one of them is active and others are in the sleeping state. If the active thread detects an image meta word, it wakes up other threads. A waken thread will check the image meta word to determine whether it is its turn to work. If yes, it loads the new base 3D graphics model for animation (e.g. switching from *male* model to a *female* model). If not, it goes to sleep again. The model loading activities are coordinated by a semaphore. (Java language does not provide any semaphore; it uses high level block-based monitors to do synchronization. However, one can easily implement a semaphore from a monitor[40].)

2. OpenCV Classification

The latest OpenCV, version 2.4.x, comes with the new *FaceRecognizer* class for face recognition. It provides three algorithms for users to perform face recognition: 1. Eigenfaces (*createEigenFaceRecognizer()*), 2. Fisherfaces (*createFisherFaceRecognizer()*), and 3. Local Binary Patterns Histograms (*createLBPHFaceRecognizer()*).

Hubel and Wiesel had studied visions of animals and found that the brain of an animal has specialized nerve cells responding to specific local features of a scene, such as lines, curves or movements [21, 22]. A brain does not see the world as isolated pieces but as a whole scene composed of related objects. The visual cortex combines different various information into useful patterns. Recognizing a face is to extract meaningful features from an image and combine them into a meaningful representation that can be classified into a specific type.

The *Eigenfaces* algorithm makes use of Principal Component Analysis (PCA)[25, 12, 11], to find a linear combination of features that maximizes the total variance in data. While PCA is an effective way to represent data, it does not consider any special features of the data; it throws away information blindly and may lose a significant amount of discriminative information when throwing minor components away.

In our work here, we mainly use the *Fisherfaces* algorithm to classify faces. The algorithm, first introduced by S. R. Fisher, uses Linear Discriminant Analysis to reduce the dimensions of class-specific data[15]. The algorithm performs very well in classifying images but may not do well in reconstructing an image. In our application, we do not have to reconstruct the original image. All we need to know is what the image class is and we use a graphics model to build a model for it.

The image is first classified into one of the two gender types: *male* or *female*. Within each gender type, the facial image is classified into one of the six emotional types: *sad, happy, angry, calm, nervous, confident*. The image type is saved as meta text data and combined with the speech text data. To distinguish the metadata from the speech data, we use the special word *$@$*, which does not occur in any human speech, to signify the image metadata; a number following this word denotes the image type. For example, we use *$@$01* to represent a happy male face and *$@$11* to represent a happy female face. However, we assume that the facial emotion stays fairly constant and the classifica-

Figure 3 Model From Google 3DWarehouse

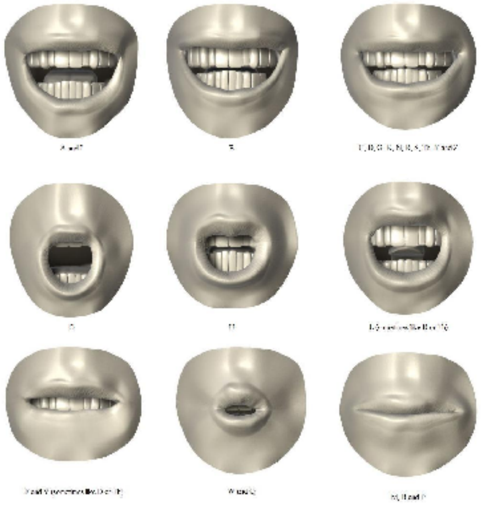


Figure 4 Preston Blair Phoneme Series

tion is done once about every 100 frames and if there is no change in the image type, no image metadata will be generated. In an extreme case, only one word of image metadata is sent for the whole video.

The databases provided by the links in the OpenCV official website are used for the training of the classification.

3. Graphics 3D Model

The 3D model is initially imported from Google SketchUp 3DWarehouse[18] and is shown in Figure 3. We use the free version of Sketchup[42], a 3D drawing tool, to convert it to a COLLADA file, which can be then imported by Blender[8, 36], a free 3D graphic suite for creating, rendering and animating graphics models[36]. Blender supports a variety of formats such as COLLADA(.dae), Wavefront(.obj), 3D Studio(.3ds), and others.

To generate a new facial expression, the model is modified by deforming the mesh, and a different copy is created and passed to the COLLADA parser to create a metafile. The viseme, or the shape of the mouth that corresponds to each phoneme, is based on the lip-sync phonetic-based animation[44, 17] used in animation movies. Figure 4 shows the lip shapes for phonemes that we have adopted.

In addition to the mouth shapes, other facial expressions are created to help simulate a more human-like agent. These facial expressions include eye blinking, eyebrow movements and yawning. These expressions are presented to keep the user entertained when the application is idle[46].

Java is employed to develop a COLLADA parser[29], which parses a COLLADA file and extracts the necessary information for rendering and animating the graphics models. The faces of the model are meshes of polygons of three or more edges. Because OpenGL ES 1.0 can only render triangles, the parser has to extract the indices of every polygon, convert them into triangles, and recalculate the normal vector for each triangle by performing a cross product of the vectors along two of the triangle's edges. Since the COLLADA file is essentially an XML document, the parser needs to make use of an XML library to carry out the parsing. Java APIs provide wide support for XML parsing and a variety of libraries to choose from such as JAXP, JDOM and SAX. Most of these libraries support the XML Path Language (XPath) [19]. While the Document Object Model (DOM) [45] is a more complete tree structure representation of the document, XPath is a straightforward language that allows the selection of a subset of nodes based on their location in the document [19]. The parsing program described here makes use of the Java package *javax.xml.xpath* to extract the necessary nodes from the COLLADA document.

The parser parses the data of a COLLADA file into a meta-file containing a set of vertices coordinates, their normal vectors, the indices of the triangle and normalized color codes[46].

Every facial expression requires a separate graphics file that has to be loaded by the Android application. In order to reduce the amount of data, if the meta-file is a variation of the base model, the parser will compare it to the base model and export only the differences. This helps to reduce the start-up time of the Android application, as it does not need to create a different graphics object for every variation. The application can duplicate the original model and apply the changes in coordinates. Every class of model (e.g. happy male or calm female) has its own base model and facial expression data, though the base models may be similar to each other. Each model is controlled by a separate thread and the loading of models is coordinated by a semaphore.

As mentioned earlier, OpenGL ES is the industry standard for embedded 3D graphics applications. This project makes use of OpenGL 1.0, which is supported by most of the commercial devices with an Android operating system. The minimum version of Android required is the Gingerbread, Android 2.3.3 API 10. One of the limitations of OpenGL ES 1.0 is that it only renders triangles. To overcome this issue, the COLLADA parser transforms a generic polygon into triangles and recalculates the normal vectors.

There are two ways to render a 3D object (or 2D for that matter) with OpenGL ES 1.0. One is array-based, and the other is element-based. To render the model with the array-based method, the vertices have to be inserted in the right order, so that OpenGL can render them in that sequence. The element-based approach is more flexible, as it does not require changes to the vertices buffer. A pointer to the indices buffer can be manipulated to render certain portions of the model at a time. This allows the program to apply certain attributes, such as color codes, to specific faces of the model without the need to load a complete color buffer with redundant information[46].

The application starts by loading the meta-files data of the 3D model previously prepared by the COLLADA parsing program into memory. The 3D model is composed of two parts. One is the upper head, and the other is composed of the mouth and jaws. Combined, they constitute a complete 3D model of a human head. While the application is loading the base model of each part to represent the resting position, a parallel thread is created to load the rest of the meta-files for different expressions. That reduces the startup time to half of what it would be if all the models are loaded in sequence. Once the meta-files are loaded into appropriate arrays and buffers, they are cached using a key-map structure for efficient access.

When there is no input, the application assumes itself to be in an idle situation and starts a timer. The application will periodically monitor the timer, and will randomly replace the resting models with animated ones, creating a frame-based movement effect. As soon as input text is detected, the application switches to the speech simulation mode, starts the TTS activity, and synchronizes the mouth animation to create the visual speech effect.

4. Lips-Audio Synchronization

The producer-consumer paradigm[23, 40, 43], a well-studied synchronization problem in Computer Science, is employed to synchronize lip movements with the speech. 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 characters are entered at the tail and are read at the head. Physically, the queue is a circular queue[23]. Logically, one can imagine it to be a linear infinite queue. 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 *head* and *tail* is equal to the length of the buffer, the queue is full, and the producer must wait.

In the application, the problem is slightly modified. It has multi-stages of producing and consuming. A consumer may take data from a queue and become a producer, putting data in another queue for other consumers to process the data. In the last stage, it has one producer and two consumers, each of which has its own head pointer. The producer thread puts the input text in a queue, and the consumers are the Android TTS engine and the animation thread that read data from it. The producer thread controls the $tail$ of the buffer and waits. Every time a new word is entered, $tail$ is incremented. The TTS thread and the animation thread read and process the data while each of them is incrementing its own $head$. When the distance between the tail and one of the heads is larger than or equal to a certain empirical constant C, the producer stops and waits for the heads to catch up. When both heads reach for the tail, the producer starts inserting new data into the buffer. To further improve the algorithm, the TTS $head$ and the animation $head$ wait for each other, which forces the speech and the animation to be more in sync as shown in Figure 5.

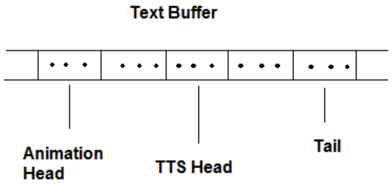


Figure 5 Producer-Consumer Data Buffer

5. Results and Discussions

We carried out some experiments using an LG Optimus P769 smart phone which runs on Android 4.1.2 (Jelly Bean) with a 1.0 Ghz Dual-Core processor. Videos were taken by the phone for a speaker who was a man or a woman and saved as MPEG-4 files. Text was generated from a speaker's speech by Google Voice and the video images were classified by OpenCV to generate image meta text, which was combined with the speech text and compressed by a program losslessly utilizing the *java.util.zip* package. The compressed text was sent to an Android receiver and uncompressed. The image metadata direct the device to load the appropriate model to animate the speech text using Android TTS. Table 1 below compares the MPEG video file size (which is already compressed) and the compressed text file size. In the table, *M1* is a video of a man reading a *Shakespere* play, *M2*, the same man reading the *Gettysburg Address*, *F1*, a woman, who is a non-native speaker, reading the *Declaration of Independence*, *F2*, the same woman reading the *Gettysburg Address*. The *Compressed Text* consists of both the image metadata and the speech text. The videos were taken at frame rates between 25 and 30 fps, which implies that MPEG had compressed the video by a factor of 12.5 to 15.5. *Period* is the time to play the video in seconds and *Bitrate* is the transmission rate, given by the size of the compressed text in bits divided by *Period* in seconds. The bottom row of the table shows the averages of the quantities. The resolution of the video is 640×480 pixels.

One can see from the table that on top of the MPEG compression, one can achieve an average additional compression factor of 91500! In other words, it has achieved a compression ratio of about 1 million for raw videos! This could save a huge amount of transmission bandwidth for the videos.

File	MPEG Size (MB)	Compress Text Size (Bytes)	Period (s)	Bitrate (bps)	Improved Compress Ratio
M1	53.8	730	85	8.6	73700
M2	55.1	738	84	8.8	74700
F1	79.0	739	119	6.2	106900
F2	81.7	738	127	5.8	110700
Av	67.1	736	104	7.1	91500

Table 1 Enhanced Compression

Our work here is more of a demonstration of the concept of video compression using contemporary techniques of speech recognition, image classification, text-to-speech synthesis, and 3D graphics modeling than developing a commercial application. The image classification and 3D graphics models are very brief. One can greatly improve the application by constructing significantly more 3D models corresponding to more human facial emotions and features such as age, facial type, race, skin color and hair type and use OpenCV with a wider database to train the classifications. However, this would require a huge amount of resources and human power, which could be only accomplished by a large corporation.

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] E. Angel, *Interactive Computer Graphics: A Top-Down Approach Using OpenGL*, Fourth Edition, Addison-Wesley, 2005.

[4] D. Astle and D. Durnil, *OpenGL ES Game Development*, Thomson Course Technology, 2004.

[5] Jun Auza, *The Technology Behind Avatar (Movie)*, *http://www.junauza.com/2010/01/technology-behind-avatar-movie.html*

[6] Charles Babcock, *Watson's Jeopardy Win A Victory For Mankind*, Information Week, Feb 2011.

[7] Koray Balc, *Xface: Open source toolkit for creating 3d faces of an embodied conversational agent*, pp. 263-266, Smart Graphics, 2005.

[8] Blender Foundation, *http://www.blender.org/*, 2013.

[9] S. R. Buss, *3-D Computer Graphics: A Mathematical Introduction with OpenGL*, Cambridge University Press, 2003.

[10] C. Bregler, M. Covell, and M. Slaney, *Video Rewrite: Driving Visual Speech with Audio*, p.353-360, SIGGRAPH'97 Proceedings, ACM Press, 1997.

[11] T.F. Cootes, G. J. Edwards, and C. J. Taylor, *Active appearance models*, p. 484-498, ECCV, 2, 1998.

[12] T.F. Cootes, C.J. Taylor, D.H. Cooper and J. Graham *Active Shape Models - Their Training and Application*, Computer Vision and Image Understanding, p. 38-59, Vol. 61, No. 1, Jan. 1995.

[13] E. Cosatto, H.P. Graf, and J. Schroeter, *Coarticulation method for audio-visual text-to-speech synthesis*, US Patent 8,078,466, Dec 2011.

[14] P.K. Doenges et al., *MPEG-4: Audio/video and synthetic graphics/audio ifor mixed media*, p.433-463, Signal Processing: Image Communication, ELSEVIER, 9, 1997.

[15] R. A. Fisher, *The use of multiple measurements in taxonomic problems*, p.179-188, Annals Eugen. (7) 1936.

[16] Forbes Magazine, *Android Solidifies Smartphone Market Share*, *http://www.forbes.com/*, Jan., 2013.

[17] Gary C. Martin, *Preston Blair phoneme series*, *http://www.garycmartin.com/mouth_shapes.html*, 2006.

[18] Google Inc. Trimble Navigation Limited. 3D Warehouse. *http://sketchup. google.com/3dwarehouse/*, 2013.

[19] E.R. Harold. *Processing XML with Java: a guide to SAX, DOM, JDOM, JAXP, and TrAX*, Addison-Wesley Professional, 2003.

[20] S. Hill, M. Robart, and E. Tanguy, *Implementing Opengl ES 1.1 over OpenGL ES 2.0*, Consumer Electronics, 2008, ICCE 2008, Digest of Technical Papers, International Conference, IEEE, 2008.

[21] 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.

[22] 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.

[23] F. June, *An Introduction to Video Compression in C/C++*, Createspace, 2010.

[24] F. June, *An Introduction to 3D Computer Graphics, Stereoscopic Image, and Animation in OpenGL and C/C++*, Createspace, 2011.

[25] F. June, *An Introduction to Video Data Compression in Java*, Createspace, 2011.

[26] The Khronos Group Inc.,*https://collada.org/*, 2011.

[27] The Khronos Group Inc., *OpenGL ES The Standard for Embedded Accelerated 3D Graphics*, *http://www.khronos.org/opengles/*, 2013.

[28] The Khronos Group Inc., *OpenGL Shading Language*, *http://www.opengl. org/documentation/glsl/*, 2013.

[29] M. Milivojevic, I. Antolovic, and D. Rancic, *Evaluation and Visualization of 3D Models Using Collada Parser and Webgl Technology*, p. 153-158, Proceedings of the 2011 International Conference on Computers and Computing, World Scientific and Engineering Academy and Society (WSEAS), 2011.

[30] S. Mlot, *Google Adds Speech Recognition to Chrome Beta*, *http://www.pcmag.com/article2/0,2817,2414277,00.asp* , PC Magazine, Jan. 2013.

[31] A. Munshi et al., *OpenGL ES 2.0 Programming Guide*, Addison-Wesley Professional, 2008.

[32] Open Source Computer Vision, *http://opencv.org/*

[33] I.S. Pandzic and R. Forchheimer, *MPEG-4 Facial Animation:The Standard, Implementation and Applications*, John Wiley & Sons, 2002.

[34] Iain E. Richardson, *The H.264 Advanced Video Compression Standard*, Wiley, 2010.

[35] Thomas Rist, and Patrick Brandmeier, *Customizing Graphics for Tiny Displays of Mobile Devices*, p.260-268, Personal and Ubiquitous Computing, 6, 2002.

[36] T. Roosendaal and S. Selleri, *The Official Blender 2.3 guide: free 3D creation suite for modeling, animation, and rendering*, No Starch Press, 2004.

[37] R. J. Rost et al., *OpenGL Shading Language*, Third Edition, Addison-Wesley, 2009.

[38] N. Sarris and M.G. Strintzis, *3D Modeling & Animation*, IRM Press, 2005.

[39] D. Shriener et al., *OpenGL Programming Guide*, Eigth Edition, Addison-Wesley, 2013.

[40] A. Silberschatz et al., *Operating System Concepts*, Addison-Wesley, 1998.

[41] M. Singhal and N.G. Shivaratri, *Advanced Concepts in Operating Systems*, McGraw-Hill, 1994.

[42] Sketchup, *http://www.sketchup.com/intl/en/product/gsu.html*, 2013.

[43] A.S. Tanenbaum, *Modern Operating Systems*, Third Edition, Prentice Hall, 2008.

[44] University of Maryland, *Blendshape Face Animation*, *http://userpages.umbc.edu/bailey/Courses/Tutorials/ ModelNurbsHead/BlendShape.html*, 2009.

[45] L. Wood et al., *Document object model (dom) level 1 specification*, W3C Recommendation, 1, 1998.

[46] 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.

[47] 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.

[48] T.L. Yu, D. Turner, D. Stover, and A. Concepcion, "Incorporating Video in Platform-Independent Video Games Using Open-Source Software", *Proceedings of ICCSIT*, Chengdu, China, July 9-11, IEEE Computer Society Press, 2010.

[49] I. Zbib, *3D Face Animation with OpenGL ES: An Android Application*, CSE Master Project Report, CSUSB, 2013.

Self-aligning wireless communication for first responder organizations in interoperable emergency scenarios

P. Dorfinger[1]**, G. Panholzer**[1]**, F. von Tüllenburg**[1]**, M. Cristaldi**[2]**, G. Tusa**[2]**, and F. Böhm**[3]
[1]Advanced Networking Center, Salzburg Research Forschungsgesellschaft, Salzburg, Austria
[2]Intelligence for Environment & Security – IES Solutions s.r.l., Catania, Italy
[3]Roofnode, Unterrabnitz, Austria

Abstract - *This paper presents wireless gateways (WGW) - a wireless communication solution for first responder organizations. The system is based on 802.11 wireless LAN technologies. The core of the system is a self-alignment of directional antennas which span a kind of mesh network. The links between the WGWs can be up to 3.5 kilometers (line of sight), without the need of human interaction for setting up the network. Further the WGW offers wireless connectivity to end devices such as tablets of the first responders. The installation of a network based on WGWs does not need IT know how and can be performed by first responders on their own.*

Keywords: wireless gateway, first responders, communication, interoperability, mesh network, automatic alignment

1 Introduction

Both data and voice communication between first responders of the same emergency organization or across different organizations, is crucial for an effective cooperation during disaster management. Especially data communication offers the possibility to exchange relevant information across organizational and national borders, making the help more targeted and faster, thus saving more people.

One of the major problems right now is that after a large scale disaster the existing broad band network is often partially destroyed or overloaded. As the communication network is essential for a better and more efficient collaboration between first responders there is a critical need for setting up alternative communication means. In such a case, another issue arises: first responder organizations are not experts in setting up communication equipment, thus these systems have to be designed such that they are easy to setup.

The wireless communication system proposed in this paper is part of the work done within the IDIRA (Interoperability of data and procedures in large-scale multinational disaster response actions) [2] European project framework. IDIRA's main aim is to improve cooperation across responding organizations by enabling interoperability between different information systems that can mutually act as data sources or data consumers, and with connected devices used on the field. This leads to more efficient multi-national and multi-organizational disaster response actions.

IDIRA starts from the concept that interoperability has to be addressed at both organizational and technical level. This means that involved actors have to align their workflow and procedures first (organizational level). At technical level, the possibility to share data, leading to more efficient cooperation by working on the same set of information, is addressed through the choice of suitable protocols and software interfaces to interconnect information systems, and through the choice of standard data formats to represent and exchange relevant information. More, at physical level, the setup of a suitable network infrastructure is needed to cope with the requirements posed by such data exchange needs.

In this context, the IDIRA Mobile Container, also called MICS (Mobile Integrated Command and Control Structure) will be brought on scene and works as information hub for all the disaster related information. The MICS provides Inbound and Outbound interoperability software interfaces for data sharing with locally deployed information sources and consumers (like e.g. existing C&C systems). Different standard, non-proprietary data formats are defined to represent, in a structured way, different types of information: incidents, resources, and sensors information just to mention a few. The Emergency Data Exchange Language (EDXL) family of standards [3] is used for data about incidents, resources, and for situation reports. The EDXL-CAP (Common Alerting Protocol) [4] data structure is used for Inbound and Outbound incidents data sharing with alerting sources and C&C systems. The EDXL-RM (Resource Messaging) [5] standard allows sharing of data that are relevant for coordinating needs and availability of resources (resource needs and offers, status of resource deployment, and so on). EDXL-SitRep (Situation Reporting) [5] is mainly used to send observations and situation reports through mobile devices, by commanders in the field. To bring the maximum benefit indeed, the MICS needs the ability that the commanders on the field are able to exchange information with each other.

The basis for the abovementioned information exchange is a working communication network, consequently the wireless gateways (WGW) explained in this paper are a core part of the IDIRA system.

2 Related Work

First responder organizations nowadays use different technologies for communication. In the daily operation often mobile phones are used. After large scale disasters the mobile phone network is often overloaded or down due to power outage or destroyed infrastructure. Specifically equipment for first responders like radio or TETRA are used during their daily business and can also be used in case of larger disasters, as the technology is equipped with backup systems in case of power outage. The interaction of IDIRA with the users is performed via a map-based Web-GUI, the so called Common Operational Picture (COP). The initial load of the COP requests about 10Mbyte of data. During the operation disparate types of data are exchanged such as sensor information, information on incidents, dispatched resources and resources activities, situation reports, simulation results, voice and user positions. For a seamless operation of the Web-GUI a few Mbit of bandwidth are needed. TETRA can only be used for low bandwidth data communication and does not fit the described bandwidth needs of IDIRA. Also a native app is developed, needing lower bandwidths, nevertheless the TETRA bandwidth will not be sufficient to cope with several end-devices at the same time.

In large scale disasters often satellite communication is used, which offers a good possibility to bring communication on scene (used as uplink technology). Different technologies and systems such as BGAN [7], VSAT [8] or Emergency.lu [9] are used by the first responders. For on scene communication between different commanders in the field satellite communication is too expensive and in the case of BGAN has only a very limited bandwidth.

For on scene communication a different system is needed. Bringing independent communication equipment such as WIMAX [10] equipment on scene has the drawback that licenses to operate the system are needed. A system specifically targeted towards broad band communication for PPDR (public protection and disaster relief) organizations is HiMoNN [11]. HiMoNN uses the frequency band 5150-5250 MHz with a transmission power of up to 8W according to the ECC Recommendation (08)04 [12]. It is capable of transmitting 28MBit/s over several kilometers. Unfortunately the system's usage is only allowed in a few countries and thus is not usable for an interoperable communication infrastructure in international disaster relief.

802.11 [13] based systems can be used all over the world, but have the major drawback that the distance between two devices is quite limited.

Meshed networks make use of end-user devices as repeaters (e.g. mobile phones). For example the mobile devices of users in the field are used as relay node for the communication of the forces. These relay nodes can be used to communicate with forces which cannot be reached directly. Nevertheless there has to be a full chain of devices between the communication partners, so that each device is able to reach another device. In case of large disasters it cannot be assumed that the density of devices is high enough that a meshed network across all the devices can be spanned.

The Optimized Link State Routing Protocol (OLSR) [14] is a routing protocol tailored to the requirements of wireless LANs. It is based on multipoint relays which reduce the routing overhead on the network.

In the work presented in this paper we try to extend the distance between 802.11 hosts by self-aligning directional antennas and we are using OLSR as routing protocol across the meshed wireless gateways.

3 System Description

Within IDIRA it has been decided that for on scene communication (in case of the public network is down or overloaded) an independent network based on 802.11 will be installed ad-hoc. The decision to setup an 802.11 based network ad-hoc after the disaster has advantages (e.g. no licenses needed, no pre-installation needed) but also has two major drawbacks:

- First responders are not trained to set up a communication network.
- 802.11 allow only a small distance between peers.

This paper presents a solution which can be used by first responder organizations to set up a communication network.

The system is based on automatically interlinking directional WLAN antennas. Directional antennas are used as this allows increasing the distance between two sites, compared to omnidirectional antennas. Due to the usage of directional antennas the signal strength will be higher and the noise level will be lower. This increase in the signal to noise ratio (SNR) allows higher throughput at similar distances or longer distances with similar throughput.

The full system is equipped into one case which is easy to install, as it only has to be mounted on a pole and switched on. After that the wireless gateway will automatically align its antennas to other wireless gateways and provide a wireless cloud as well as a LAN connection to the spanned network.

This system allows being setup by non trained first responders, and the distance between the peers can be extended compared to a WLAN using omnidirectional antennas, thus overcoming the main two drawbacks.

The main building blocks of the wireless gateway are presented in Figure 1. The system is mounted in a housing which consists of four layers. The top three layers (also referred to as modules) are built up identically. Each one consists of a wireless access point (which can also be configured as wireless client) connected to a directional antenna. Further a motor which allows rotating each antenna individually by 360° is mounted in each of the top three layers. In the bottom layer the core parts of the systems are installed. A small embedded PC as router (router-board) running OpenWRT firmware together with a switch provides connectivity between the WLAN stations and the rest of the system. The self-alignment algorithm is controlled by the router-board. The router-board is connected to a microcontroller which is responsible for the control of the rotation of the modules. It operates the motors and reads the sensor values controlling the rotation from the top three layers. The router-board is equipped with two WLAN modules one is configured to operate at 2.4GHz and used for connecting end user terminals. The second one is configured to operate in the 5GHz range and is used for the self-alignment algorithm of the remote WGWs.

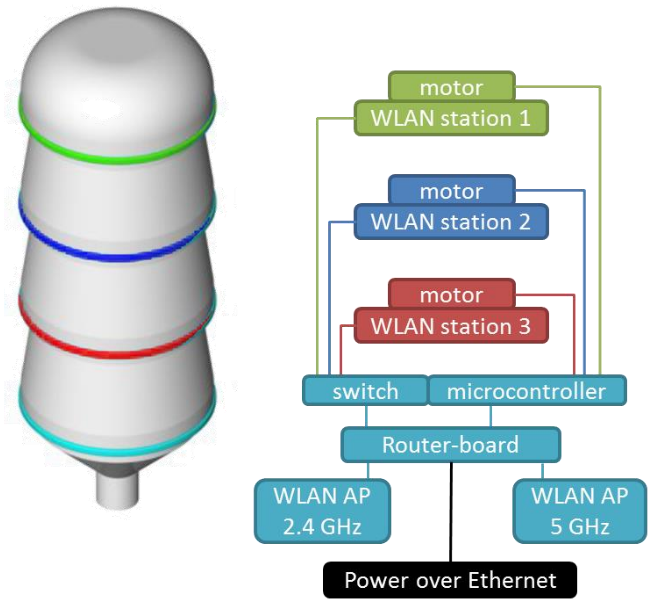


Figure 1. Main building blocks

4 Prototype Description

Based on the system description in chapter 3 a prototype has been designed and built. A modular plastic housing with four stacked layers of similar size has been chosen. A schematic overview of the prototype is shown in Figure 3.

The top 3 layers contain an electrical driven turntable which holds and rotates an embedded system with built-in directional antenna. This embedded system provides a 100Mbps Ethernet interface and a 5GHz 802.11n wireless interface with a 16dBi directional MIMO antenna.

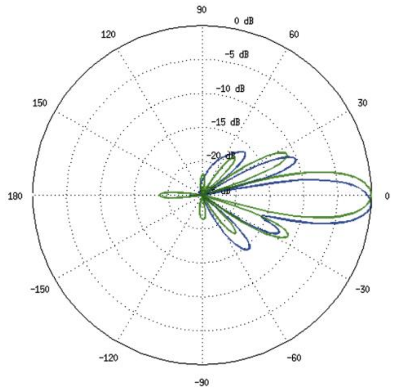


Figure 2. Nanostation M5 elevation [15]

All these features are covered by using a commercial product called Nanostation M5 from manufacturer Ubiquiti Networks. In total three such Nanostation M5s are embedded into the prototype. For easier integration the plastic housing of the Nanostation M5 was removed and only the bare electronics was mounted on the turntables. The antenna inside the Nanostation M5 has a non-symmetrical radiation pattern of about 42° azimuth angle and 15° elevation angle. The Nanostation M5s are mounted 90° rotated so that the relevant radiation angle for the mechanical antenna alignment process is now the narrow 15° angle. The directional antenna is dual-polarized to support the MIMO feature of the 802.11n wireless interface. Figure 2 shows the radiation pattern of both polarization planes for the 15° beam width.

The mechanical design enables to turn all three directional antennas independently 360° on the horizontal plane while having an outer casing that covers and protects all mechanical and electrical parts. Furthermore the electrical design enables to use common and cheap Ethernet interfaces as data link between all three radio devices and a superior network-routing module.

All three turntables are equipped with a DC gear motor and controlled via one central Arduino Leonardo microcontroller installed in the 4th layer. This microcontroller is responsible for turning all three turntables to the desired position. The microcontroller is able to locate the home position of each turntable through a light-barrier attached to the turntable itself together with a reflector attached to the outer housing of each layer.

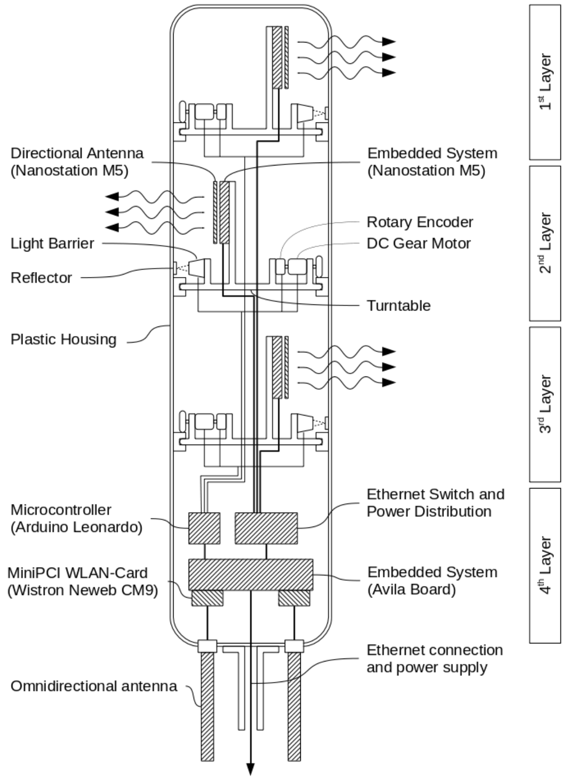


Figure 3. Prototype schematic overview

The exact positioning of the turntables is controlled through a feedback signal from an incremental rotary encoder. This rotary encoder is directly attached to the output shaft of the DC gear motor. The feedback signal also helps to detect mechanical problems like stuck or broken gear elements. It provides a cyclic 2-bit pattern through two output signals. These pulses are used to measure the alignment of the turntable by incrementing or decrementing a software counter. Furthermore this bit-pattern can also be used to detect the actual direction of rotation.

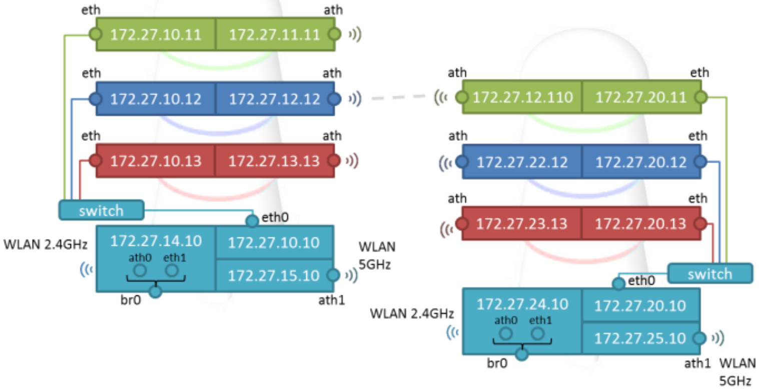


Figure 4. IP addressing scheme for two connected nodes

The software inside the microcontroller is responsible for reading and parsing text commands via serial interface and providing status information like the actual position of each turntable via the same serial interface. An initialization routine is implemented to home all turntables after power-up. During normal operation the microcontroller controls the DC motor and counts the pulses from the rotary encoders until the desired position is reached. As soon as a rotary encoder sends pulses while the motor is not in operation the software indicates a problem.

The microcontroller is not aware of cardinal directions. Instead it's only aware of the angular displacement of each turntable based on its home position. The software accepts additional commands to store the actual turntable positions in a non-volatile memory. This stored position is recovered when the device powers up and the homing procedure is finished. The microcontroller is connected to the prototypes main logic board via serial interface.

The main logic board is an industrial grade embedded board called Avila from manufacturer Gateworks. It is based on an Intel IXP425 CPU and features two 100Mbps Ethernet ports and 4 MiniPCI. Two out of the 4 MiniPCI slots are equipped with CM9 wireless cards from manufacturer Wistron NeWeb. These cards are based on Atheros AR5213A chips and can be configured for 2,4GHz or 5GHz operation. They are well supported via ath5k open-source Linux wireless driver.

The Arduino microcontroller and the Avila embedded board are both installed in the lowest (4th) layer of the prototype setup. As backplane connection between all 4 layers a 5-Port 100Mbps Ethernet Switch is installed into the 4th layer. This switch is also responsible for providing power via Ethernet cable to the 3 Nanostation M5s. As the chosen microcontroller, motors and sensors are powered by 5V also a DC/DC converter is installed in this layer.

The Nanostations are configured to operate as router and run a modified Ubiquiti firmware including the OLSR routing protocol. On the Avila embedded board OpenWRT with the OLSR package is running. The OLSR configuration has been modified such that Ethernet links have a cost of 0.1. Figure 4 shows the IP addressing scheme of two connected WGWs.

The system has been designed such that it can be transported and installed by one person. Therefore a battery-pack which lasts about 12 hours, together with a tripod and a telescopic 6m pole build a full system setup which can be installed in the field.

5 Alignment Algorithm

Figure 5 shows a simplified flow-diagram of the alignment algorithm performed by the WGW.

After providing power to the WGW via PoE the WGW will start to initialize. During the initialization sequence the home position of the modules is identified and a self-check of the system is performed. After the initialization all three modules start to scan for remote WGWs. For each module individual start and stop positions are defined. The default step size is 5°. The modules start with an offset of 120° to each other so that all directions are scanned as fast as possible. The scan will continue until the stop position for the module has been reached or a remote WGW has been identified. Based on experiments we have defined a threshold such that weak signals below -87dBm are not taken into account. The scan is performed for the WLAN spanned by the 5GHz omnidirectional antenna of the remote WGW. The Nanostations M5 are configured to operate in AP mode because in station mode scan results are cached over several scans. As we perform one scan per direction this would falsify our scan results. One scan takes about 7 seconds.

If a remote WGW has been identified one directional antenna will rotate to the position with the best signal and the module will be configured such that it will connect to the remote omnidirectional antenna. After the connection is established the WGW will request a connection with a directional antenna at the remote WGW.

When the requesting node receives the confirmation the connection to the remote WGW omnidirectional antenna is canceled, and the local WGW is configured to connect to the remote directional module. The local module is configured as client with the SSID of the remote module. Further the IP address is set appropriately for the remote module (see Figure 4). If the request is rejected the module will continue with scanning for remote WGWs.

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

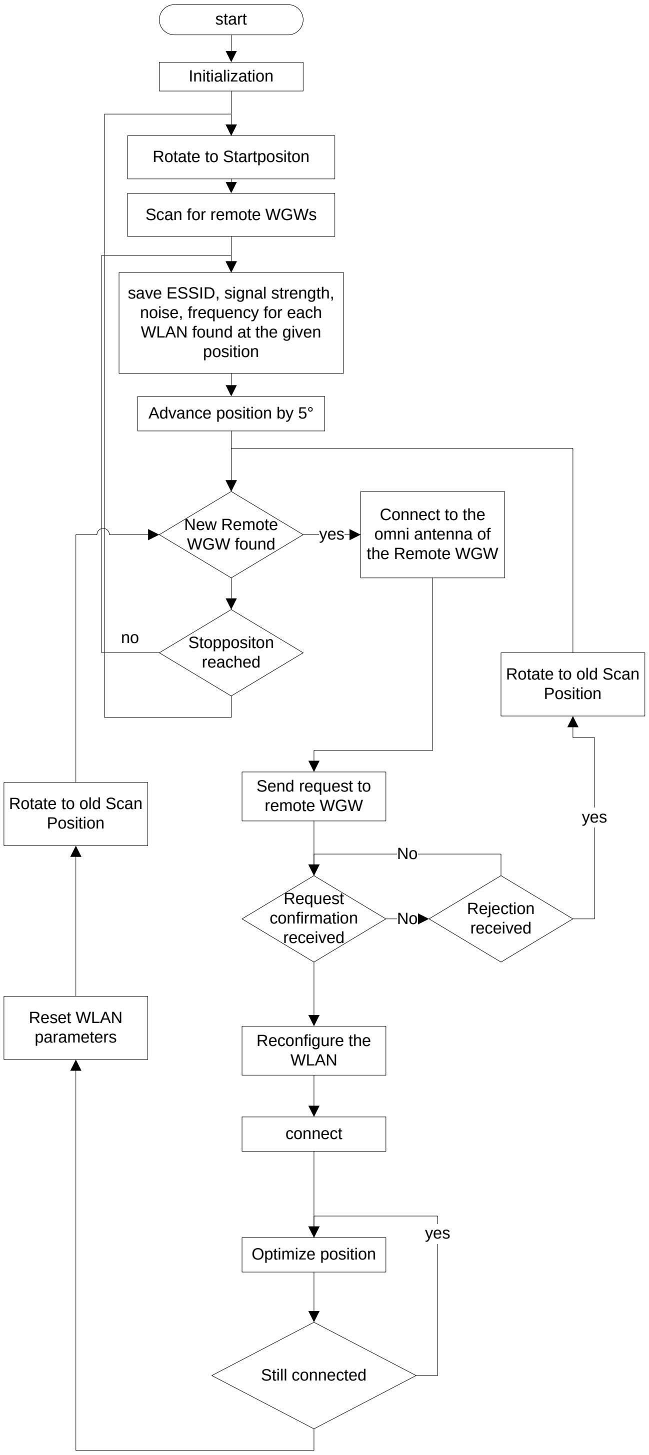

start
Initialization
Rotate to Startpositon
Scan for remote WGWs
save ESSID, signal strength, noise, frequency for each WLAN found at the given position
Advance position by 5°
New Remote WGW found
yes
Connect to the omni antenna of the Remote WGW
no
Stopposition reached
Rotate to old Scan Position
Send request to remote WGW
yes
No
Request confirmation received
No
Rejection received
Rotate to old Scan Position
Reset WLAN parameters
Reconfigure the WLAN
connect
yes
Optimize position
Still connected

Figure 5. Local Alignment Algorithm

Modules which are not connected will continue to scan for remote WGWs. All established connections are monitored, and if one of them is lost the module will be reconfigured and will start to scan for remote WGWs again or connect to another previously located WGW.

If new scan results show that the remote WGW can be reached with a better signal at a different position the position of the module will be adjusted such that the optimal signal strength is ensured.

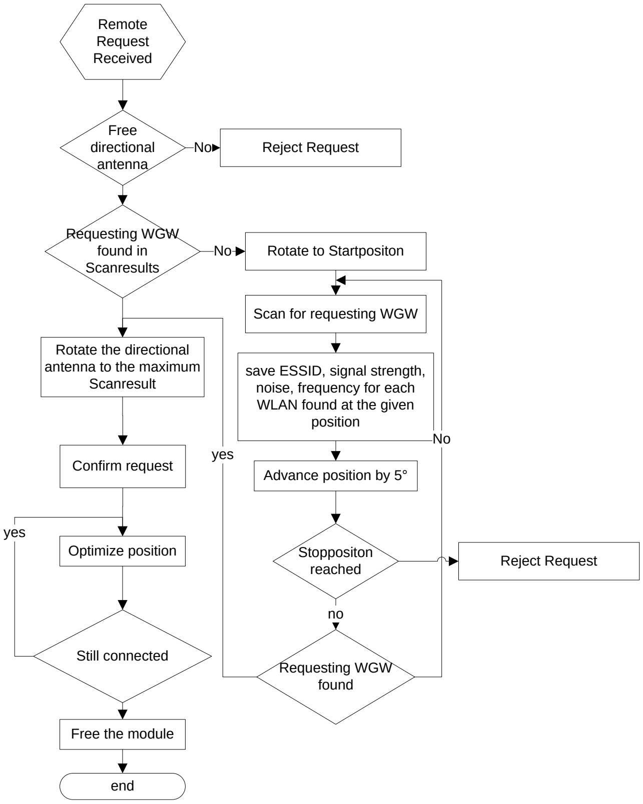


Figure 6. Remote Alignment Algorithm

Figure 6 shows the sequence performed at the remote WGW when a request has been received. Each time when a request is received the remote WGW tries to fulfill the request and assign one of the directional antennas to the requesting node. If at the remote WGW all directional antennas are used by other WGWs the request cannot be fulfilled and will be rejected. Otherwise the scan results will be searched for results of the requesting WGW. If a scan result exists a directional antenna will be rotated towards the direction of the best scan result and a confirmation for the request will be sent to the requesting WGW. If no scan result for the requesting WGW is present a 360° scan will be performed using all not assigned modules. The scan will run until the requesting WGW is found or the stop positions for the involved modules are reached. If the stop positions are reached without identifying the requesting WGW the request is rejected, otherwise the request is confirmed and an antenna is rotated to the position where the requesting node has been identified with the best signal strength. As long as the connection is established the position will be optimized (adjusted) to ensure the maximum signal strength.

The alignment sequence typically lasts between five and 15 minutes.

6 Results in the Field

Due to the rotating directional antennas the system benefits from a better signal to noise ratio, and consequently reaches higher throughput of the connection.

To show the practical gain we have installed two wireless gateways in a distance of about 2.2km with a vertical drop of about 100m. Each of the WGWs has been mounted on a 6m pole. We have rotated one of the directional antennas in a reduced step size of 2° and scanned for the WLAN of the remote 5GHz omnidirectional antenna. The sending power for omnidirectional antenna and for the Nanostations was optimized to 30dBm EIRP. For scanning for remote nodes the Ubiquiti Nanostations are running in AP mode and the command *iwlist ath0 scan* is used for the scanning. Figure 7 shows the noise and signal levels of the remote omnidirectional antenna. The maximum signal to noise ratio (SNR) in this case is 14dB at 186°. This direction is chosen for the connection between the two wireless gateways. After the connection between the two directional antennas is established, the SNR increased to 15dB. Compared to a connection between the two omnidirectional antennas the gain in SNR is 6dB, as the SNR for the omnidirectional connection is only 9dB. The figure also shows the secondary lobe of the antenna where the center is about 25° off the major lobe. The SNR at the secondary lobe is too weak for a successful connection.

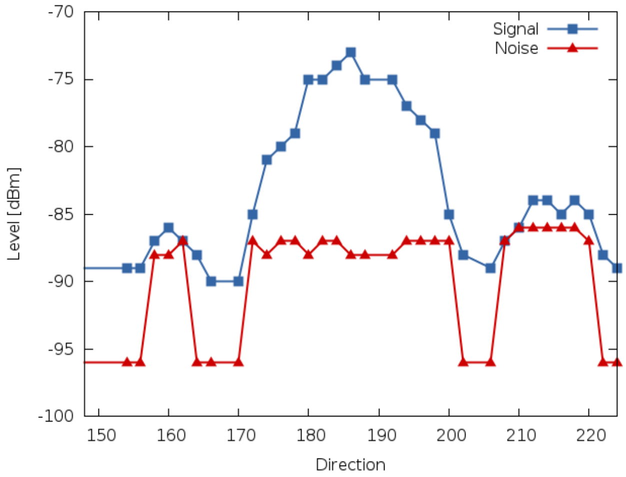


Figure 7. Noise and Signal level of remote omnidirectional antenna

The self-alignment algorithm of the system has been extensively tested in the field. The system has been used to provide access to the COP for first responders. Several installations have been performed therefore. To evaluate the influence of the distance between the WGWs measurements in a rural area north of Salzburg have been performed. Several test points have been identified for the installation of the WGW. The WGWs were mounted on a 6m pole and the automatic alignment has been started. The automatic alignment algorithm successfully establishes links over a distance of about 3.5 kilometers. For larger distances the scan for the remote omnidirectional antenna fails and a connection to the remote omnidirectional antenna is not possible. Nevertheless a software controlled manual alignment of two Nanostations can be performed. The interface for manual alignment is designed such that it can be used by non expert users. After a successful alignment the signal and noise levels and throughput have been measured. The measurements have been performed between two ASUS netbooks directly connected to the WGW. To orchestrate the measurement the MINER software [16] has been used. The throughput measurements were done using iperf [17].

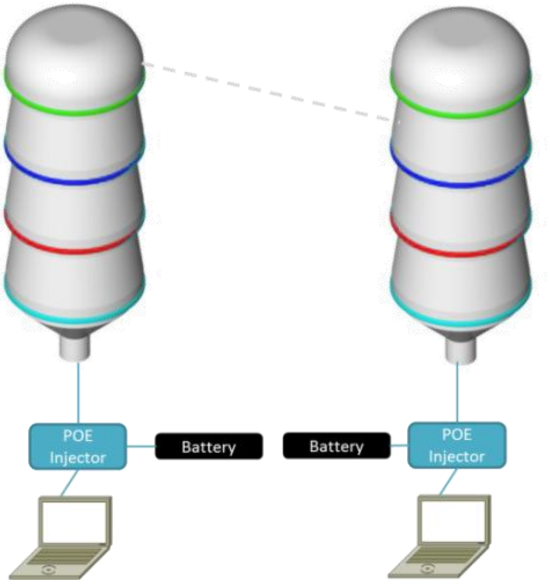


Figure 8. Measurement Setup

Figure 8 shows the setup for these measurements. Figure 9 shows the values for signal and noise levels in dBm and Figure 10 shows the measured throughput across the link. The SNR decreases from 18dB to 5dB. For distances up to 1.5km the throughput is close to the maximum speed of the Fast-Ethernet interface of the Nanostation M5. Even at a distance of about 5.5km the throughput has been in the region of about 20Mbit/s, which is sufficient for first responder communication needs.

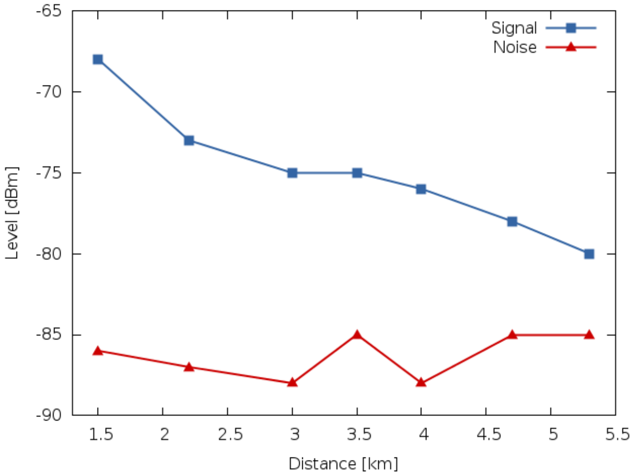


Figure 9. Signal strength and Noise level

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

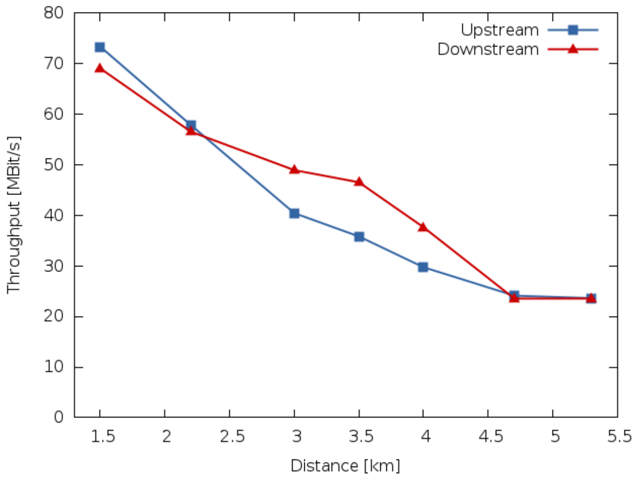


Figure 10. Throughput

7 Conclusion & Future Work

This paper presents a novel approach for a broad band communication network for first responders in case of large scale disasters. The network is based on self aligning directional 802.11 antennas. Using directional antennas increases the distance between two sites, and the self alignment algorithm ensures that the network can be setup by untrained first responders.

The automatic alignment algorithm tries to identify the maximum signal to noise ratio (SNR) for the connection and align the antennas accordingly.

Results of measurement performed with the first prototype show that the self alignment algorithm can establish connections across a distance of up to 3.5km. For larger distances the alignment can be manually controlled via an easy to use software interface.

The prototype fulfills the throughput requirements of the first responders. The major focus here is on the usage of the IDIRA system, as the prototype is part of this project.

In the future it is planned to advance the interface for the manual control of the alignment. Further the robustness of the automatic alignment algorithm will be improved. All the decisions of the algorithm are now based on local available information. This may not lead to the optimal network setup as setting up different links would lead to a more robust meshed network. Consequently it is planned that when several Wireless Gateways (WGWs) are connected to each other, the optimization of the connections will be performed with a global focus. This will allow optimizing the links within the meshed network to be more robust in case of link failures or removing WGWs.

Acknowledgment

This work was partially supported by the IDIRA FP7 261726 project. The authors thank Christoph Kneidinger, Frederic Gierlinger, Günther Breinesberger and Lukas Krämer for supporting the measurements in the field.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)

[2] IDIRA Project, “Interoperability of data and procedures in large-scale multinational disaster response actions, 2011-2015, ”http://idira.eu/

[3] OASIS Emergency Management TC, “Emergency Data Exchange Language (EDXL)”, https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=emergency

[4] OASIS Emergency Management TC, “Common Alerting Protocol” Version 1.2, 2010-07-01 - CAP-v1.2-os, https://www.oasis-open.org/standards#capv1.2

[5] OASIS Emergency Management TC, “Emergency Data Exchange Language Resource Messaging”, EDXL-RM-v1.0-OS-errata-os, 22 Dec. 2009, https://www.oasis-open.org/standards#edxlrm-v1.0

[6] OASIS Emergency Management TC, “Emergency Data Exchange Language Situation Reporting”, edxl-sitrep-v1.0-wd19, Draft 02, 2012-08-07, http://docs.oasis-open.org/emergency/edxl-sitrep/v1.0/cs01/edxl-sitrep-v1.0-cs01.zip

[7] BGAN – Imarsat “Broadband Global Area Network” http://www.inmarsat.com/service/bgan/

[8] VSAT – GVF “Global Very Small Aperture Terminal Forum” http://www.gvf.org

[9] Emergency.lu http://www.emergency.lu

[10] IEEE 802.16 WIMAX “IEEE Standard for Local and metropolitan area networks”

[11] IABG mbH “HiMoNN Higly Mobile Network Node” http://www.himonn.de

[12] Electronic Communications Committee (ECC), “THE IDENTIFICATION OF FREQUENCY BANDS FOR THE IMPLEMENTATION OF BROAD BAND DISASTER RELIEF (BBDR) RADIO APPLICATIONS IN THE 5 GHz FREQUENCY RANGE”, http://www.erodocdb.dk/docs/doc98/official/pdf/REC0804.pdf

[13] IEEE 802.11 “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)” http:// http://standards.ieee.org/about/get/802/802.11.html

[14] Clausen, T. & Jacquet, P. (2003), 'Optimized Link State Routing Protocol (OLSR)' (3626) , Internet Engineering Task Force , IETF , RFC 3626 (Experimental)

[15] Ubiquiti networks, “NanostationM & NanostationlocoM Datasheets” http://dl.ubnt.com/datasheets/nanostationm/nsm_ds_web.pdf

[16] Christof Brandauer, Thomas Fichtel (2009): MINER – A Measurement Infrastructure for Network Research In: Proceedings of the 5th International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (TRIDENTCOM 09), Washington D.C.

[17] Iperf.fr “Iperf – The TCP/UDP bandwidth measurement tool” http://www.iperf.fr

UWB Antenna Equalization of Pulse Modulation for IEEE 802.15.4a UWB PHY

Loay Khalaf

Electrical Engineering Department,
The University of Jordan, Amman, Jordan

Abstract - *Ultra Wideband (UWB) antennas cause pulse distortion due to their frequency response. The antennas frequency response depends on the position and direction of the antenna. When used in consumer small devices, the antenna direction and position are not located for optimal performance; rather, they are located close to the users' devices. The frequency response, in the form of phase and group delay, and impedance variation as function of frequency cause pulse signal distortion. Pulse distortion causes performance degradation, and thus needs to be reduced, especially for coherent reception. Equalization can be used to reshape the pulse distortion caused by the antenna and the radio front end, and reduce pulse dispersion, as shown in this paper.*

Keywords: UWB antenna, radio front end, distortion, equalization

1 Introduction

A communications system performance is affected by several factors, such as random noise (usually Gaussian distributed), and other impairments that can be estimated and corrected, such as channel response, including frequency offset, phase and delay offset, and dispersion. For best operation, a good signal design, with prior knowledge of channel response is chosen. Usually, performance is dictated by a certain bit error rate prescribed at some signal level [1] [2] [3].

An UWB system, as other transceivers, consists of three main parts: a transmitter, a channel, and a receiver. Signal corruption and distortion occurs at each part. The transmitter's power amplifier, if power limited, is operated in a nonlinear region, causing nonlinear distortion. The power amplifier is easily characterized and its effect can be dealt with by performing predistortion of the signal before the power amplifier, thus reducing the distortion. The wireless channel's response is more complicated, however, and might consist of multipath fading, and additive noise [4]. The channel configuration is usually random, and can be characterized by probability density functions. The fading and bandwidth limitation are corrected by equalization, while the noise is best reduced using low noise amplifiers and low phase noise oscillators. The antenna is usually considered a part of the channel, and can cause distortion with dependence on its three coordinates *(r, θ, ϕ)*. While the dependence on the radial distance, *r*, almost disappears in the far field region except for the attenuation due to power spreading over surface area. The other coordinates play a major role, and they influence the impulse response. In addition, the antenna reflection coefficient is a major factor as well. Factors affecting the antenna performance include:

1. Antenna reflection coefficient (at antenna input of the TX antenna, and output at the RX antenna).

2. Antenna response, with input at the transmit antenna and output from the receive antenna.

3. Antenna response is a function of channel frequency, and antenna location and direction.

4. Frequency band, the 802.15.4a standards call for several channels within the 10 GHz band.

Antenna designers optimize the performance by maximizing the radiated power. Narrowband antenna design assumes approximately uniform antenna response within the band, and does not pose high design challenges. UWB antennas, on the other hand, have severe variations over the operating band and require equalization to maximize the performance.

Few authors addressed UWB equalization for antenna response. In this paper, it is shown that equalization is required for coherent demodulation and how it is performed using a simple algorithm.

2 UWB Antenna Pulse Spatial Dispersion and Impulse Response

UWB antenna design focuses mainly on the transmitting antenna reflection coefficient. The power conservation property implies that minimizing reflected energy at the

antenna maximizes the transmitted energy. Most literature addresses the reflection coefficient in terms of S_{11} simulations and measurements in the frequency domain [1], while some authors addressed the time domain properties for a transmitter and receiver antenna pair [1].
UWB antenna-pair response is of most importance for the IEEE 802.15.4a UWB pulse signal dispersion. Proper receiver operations depends on the correlation properties of the received signal and the reference signal as specified in the standards. The antenna impulse response is a function of the azimuth, ϕ. and elevation angle, θ, as discussed in [1]. The variations of the impulse response can be quite considerable to the point of causing considerable pulse distortion. While incoherent reception does not require equalization due to the phase response, none the less, pulse dispersion (pulse broadening) for signals similar to the narrow pulses of the 802.15.4a is significant. As shown below, a 2 nano sec pulse at the TX is spread to around 4 nano sec pulse at the antenna output of the receiver.

A cylindrical patch antenna has been shown to have good properties for the 3 to 10 GHz frequency band. The cylindrical patch antenna has been investigated in the frequency and time domain in [1]. Fig. 1 shows a simulation for the system response, S_{21}, or insertion loss, for a transmit and receive antenna placed directly opposite to each other. The signal processing assumed sampling rate of twice the highest operating frequency o 10 GHz, or a sampling frequency of 20 GHz. Note that the attenuation here is mainly due to the fact that the antenna is not highly directive, thus there is power spread as the reciprocal of the square of the distance.

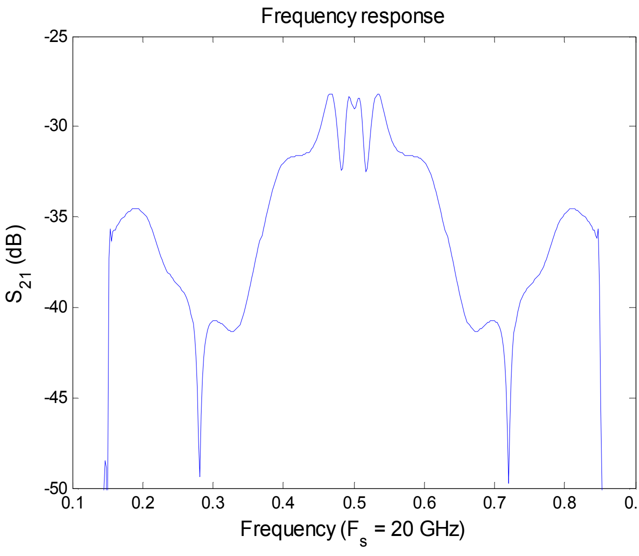


Figure 1: Antenna pair system response

Fig. 2 shows a Gaussian pulse at the input of the cylindrical patch UWB antenna. The received signal at the output of a similar receive antenna (placed directly opposite to each other) is shown in Figure 3. The input signal is a pulse with a width of 2 nsec.

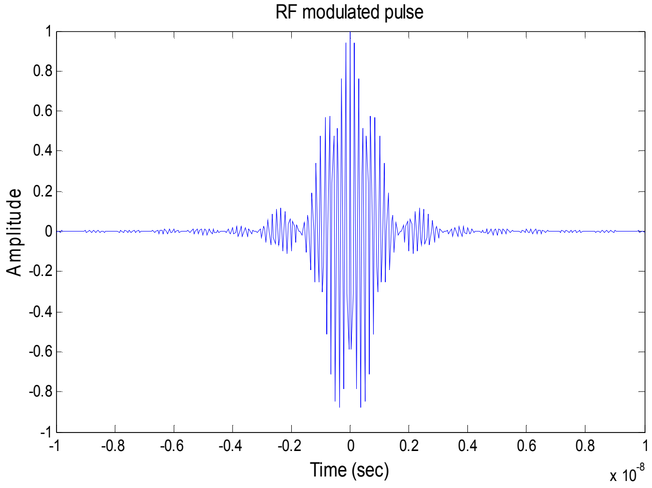


Figure 2: RF modulated pulse

The spectrum of the received pulse is shown in Fig. 3

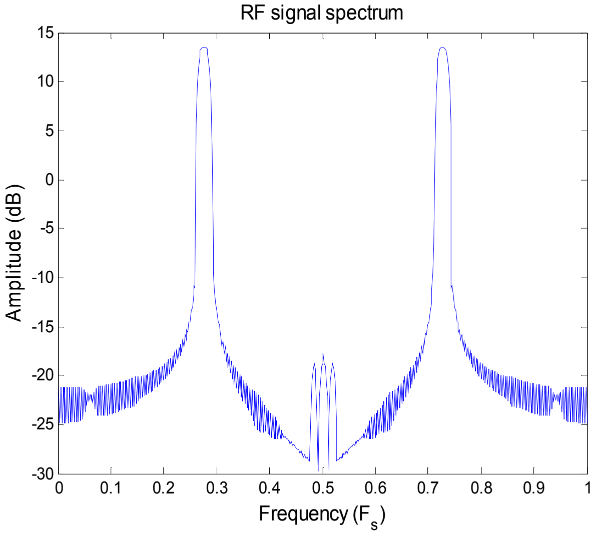


Figure 3: RF received signal spectrum

3 Power Consumption and Equalization of UWB Signals

The requirements for consumer battery operated UWB devices call for low power consumption. The power amplifier of the transmitter consumes most of the source power. The receiver signal processing operations, including the analog to digital conversion, consumes most of the receiver power.

The power consumption of the transmitter is minimized by providing antennas with good match over the whole operating bandwidth, and by operating the amplifier in a class C or E efficient (non-linear amplifiers), while minimizing distortion by digital predistortion methods. Excellent UWB antenna matching is obtained by reducing the reflection coefficient to below -10 dB, and has been demonstrated in the literature [1].

The power consumption of the ADC of the receiver is reduced by decreasing the sampling rate and the resolution (bitwidth) of its output. Given the maximum channel bandwidth of the 802.15.4a of 1.2 GHz, the Nyquist rate is at 2.4 GHz, however; higher sampling rates are required during preamble processing for channel estimation, which raises power consumption during that period. The equalizer will consume large power during the initial phase of training before convergence.

3.1 Equalizer Architecture

Fig. 4 below shows the proposed equalizer architecture. The performance of the equalizer depends on the length of the equalizer (the number of equalizer taps), the resolution, the rate (clock), and the algorithm used for equalization. The equalizer can be thought of as a filter performing deconvolution with the antenna response, with the impulse response being estimated based on the received preamble signal, which is known to the equalizer.

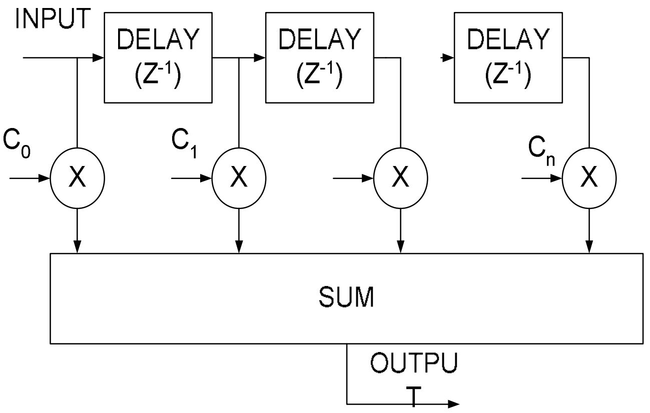


Figure 4 Equalizer architecture

The equalizer algorithm actually estimates the channel impulse response, including the antenna. Several algorithms are known, with different performance in convergence depending on the channel characteristics [4]. Of course, the equalizer can not differentiate between the antenna or the channel response, but all that is needed at the receiver is the total system response which consists of the convolution of the antenna and the channel.

3.2 Equalizer Algorithm

A direct computation for the system response can be done using a deconvolution process implemented with Fourier and Inverse Fourier Transforms. However, such computations require very high signal to noise ratio (SNR) and great computation capability not available in low power consumer devices. Hence, we use a mean square estimation (MSE) algorithm, which is a simple equalization algorithm that has been proven to work in lower SNR situations. The algorithm performance is discussed in section 4.

4 Simulation Results

Major factors in the design of the equalizer are its length, and its algorithm. The length of the equalizer is determined by the system response. Fig. 5 shows the received time domain signal at the RX antenna output. As the pulse is spread over 4 nano-sec, the length of the equalizer is thus around 10 delay units, given a clock of 2.5 GHz (which is the Nyquist rate for the highest bandwidth channel) [5]. The convergence of the equalizer algorithm is also dictated by the length of the preamble as specified in the standards.

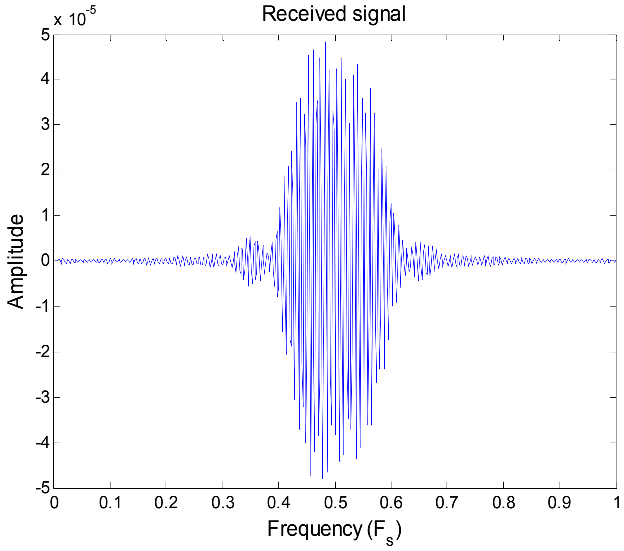


Figure 5 Received RF signal

Fig. 6 shows the baseband response for the RF carrier channel at the output of the equalizer. It is seen how much improvement is done by the equalizer, making the pulse duration closer to the transmitted reference signal. Other carrier frequencies (channels) shows similar results

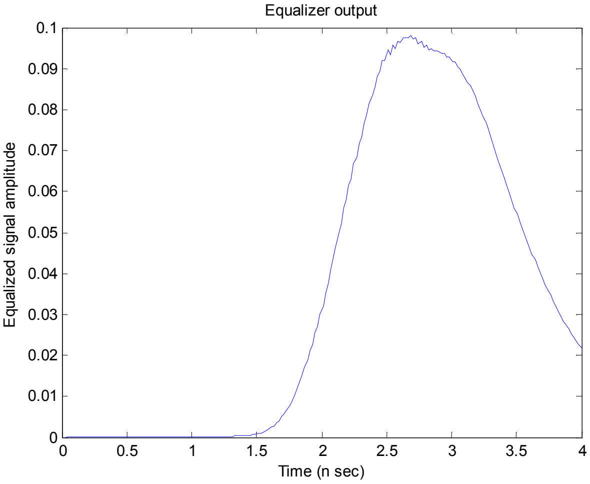


Figure 6 Equalizer output

5 Conclusions

Wireless personal consumer products are placed without regard to antenna directivity. The antenna response is a function of frequency and direction, thus causing spatial signal dispersion that is not easily predicted or controlled. As shown here, equalization can be used to perform signal restoration for optimum results and performance enhancements.

6 References

[1] Mohamed Srifi, Symon Podilchak, Mohamed saaidi, Yahia Antar, "Compact Disc Monopole Antennas for Current and Future Ultrawideband (UWB) Applications", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 59, NO. 12, DECEMBER 2011

[2] Meng-Hsuan Chung, "Receiver Designs for Ultra-Wideband Radios with Eficient Multipath Diversity Utilization ", THESIS, UNIVERSITY OF SOUTHERN CALIFORNIA, 2006

[3] Changhui Hu, Rahul Khanna, Jay Nejedlo, Kangmin Hu, Huaping Liu, Patrick Y. Chiang, "A90 nm-CMOS, 500 Mbps, 3–5 GHz Fully-Integrated IR-UWB Transceiver With Multipath Equalization Using Pulse Injection-Locking for Receiver Phase Synchronization", IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 46, NO. 5, MAY 2011

[4] Bo Zhou, Fei Chen, Woogeun Rhee, Zhihua Wang, "A Reconfigurable FM-UWB Transceiver for Short-Range Wireless Communications", IEEE MICROWAVE AND WIRELESS COMPONENTS LETTERS, VOL. 23, NO. 7, JULY 2013 371

[5] IEEE 802.15.4a 2007 Standards

Wireless-Aware Congestion Control for Transmission over Heterogeneous Networks

Jyun-Siou Fan[1], Sheng-Shuen Wang[2], and Hsu-Feng Hsiao[3]
Department of Computer Science, National Chiao Tung University
1001 University Road, Hsinchu, Taiwan
gree1970@hotmail.com[1], {sswang[2], hillhsiao[3]}@cs.nctu.edu.tw

***Abstract*—TCP is the most widely used transport layer protocol on Internet. The congestion control mechanism implemented in a traditional TCP uses events of packet loss as indicators of network congestion. However, this is not suitable for transmission over heterogeneous networks, since there can be wireless packet losses when data passes through wireless networks. If a congestion control algorithm takes the wireless losses caused by common channel errors due to multipath fading, shadowing, and attenuation as an index of network congestion, it will mistakenly lead to degraded performance because of the incorrect reduction of the congestion window. In this paper, we designed a packet loss classification algorithm which uses two-trend detections on relative one-way trip time to differentiate congestion losses from wireless losses. A wireless-aware end-to-end congestion control algorithm is further proposed. The congestion control algorithm discriminates wireless losses from congestion losses, and then it regulates the congestion window and slow start threshold properly. The extensive simulation results show that the proposed algorithms can significantly improve network performance when compared with the congestion control algorithms implemented in the TCP protocols.**

Keywords—congestion control algorithm; transmission control protocol; heterogeneous networks.

1. INTRODUCTION

With the advances of wireless technologies and ever increasing demands for mobile communications, Transmission Control Protocol (TCP), which is the dominant transport layer protocol on Internet, provides a reliable, connection-oriented service and is being extended to wireless network. However, wireless communications are carried out in the open air and the signal may disperse and travel on different paths due to reflection, diffraction, and scattering caused by obstacles before it arrives at receivers. The resultant signal may have been significantly distorted and attenuated when compared with the original signal. Therefore, wireless losses [29] can appear randomly due to signal fading by unreliable wireless medium. Traditional TCP is not quite appropriate to be used over heterogeneous (mixed with wired and wireless networks) networks because of the mechanisms of its congestion control algorithm. TCP congestion control algorithm uses the event of packet losses as an indicator of network congestion. If the congestion control algorithm takes wireless losses as an index of network congestion, it will lead to dramatic performance degradation due to the reduction of the congestion window by mistake.

In order to solve this problem, several congestion control approaches have been suggested. There are four different types of approaches [1][2][12][25][26]: split connection, link-layer, end-to-end, and network cooperation. The main idea of split-connection approach [10] is to split connection at the base station between wired and wireless network. The base station serves as a relay node to separate congestion control functionality on wireless links from that on wired networks. When bandwidth asymmetry exists between the wired and wireless path, a huge buffer is required at the relay node to store and forward packets toward the mobile host. “Snoop” is a proxy-based link layer enhancement to cache copies of TCP data packets at the base station, and to monitor the ACKs from a receiver to its sender. Besides, some link layer approaches [26][28] attempt to reduce wireless losses using scheme of automatic repeat request (ARQ) in IEEE 802.11 MAC to recover transmission error locally by retransmitting the lost frame at the link layer. This protocol could reduce end-to-end retransmission and prevent the associated reduction in congestion window size by retransmitting the cached copy for local wireless links. ARQ is also used in 802.16 MAC (WiMAX), and also LTE protocols. For the two approaches mentioned above, the control overhead is considerable as the base station needs to maintain a significant amount of state information for each TCP connection.

Instead of the support from base station, the network cooperation approach [16][17] requires the assistance from intermediate routers to send the information about network condition to end-systems for the improvement of congestion control efficiency in presence of wireless losses. The intermediate routers do not require per-flow states. However, it is expensive to deploy all the intermediate routers to enable this function.

The end-to-end methods are promising since significant gains can be achieved without extensive support at the network layer at routers and base stations. The main concept of end-to-end approaches is to distinguish wireless losses from congestion losses at end-system and no explicit modifications are required for the intermediate routers, which are out of the control of the end users. This approach has attracted extensive research attention because this approach can be deployed easily and gradually. Besides, the end-to-end approach can be further classified into reactive and proactive congestion control in [20] according to whether feedback information is used to reallocate network resources.

A packet loss classification based on the work in [6] is proposed in this paper to differentiate packet loss classes. According to the classification result, the congestion control algorithm can adjust the sending rate effectively and proactively based on congestion losses instead of wireless losses. Furthermore, an improved congestion control algorithm based on the TCP congestion control mechanism in NewReno was developed.

In section 2, several packet loss classification (PLC) algorithms proposed in the literature are discussed, as well as several congestion control algorithms that are commonly used in TCP protocols. In section 3, a new PLC algorithm based on the trend detection of relative one-way trip time is proposed. The wireless-aware congestion control algorithm for TCP is further proposed. In section 4, we evaluate the performance of the proposed algorithms and the competing algorithms in the literature, followed by the conclusions in section 5.

2. Related Work

In real network environment, there are two classes of packet losses in heterogeneous networks: one is congestion loss, and the other is wireless loss. A packet loss due to network congestion is called a congestion loss. The other kind of packet loss due to shadowing and/or signal attenuation over wireless networks is called a wireless loss. If a wireless loss is used as an indicator of network congestion, the congestion window will be reduced incorrectly.

Biaz scheme [3] uses packet inter-arrival time to differentiate congestion losses from wireless losses at the receiver side using threshold. Biaz assumes that only the last link along the path is wireless and it is also the bottleneck, and the scheme classifies lost packets according to the temporal range. On the other hand, ZigZag [5] increases the classification threshold according to the number of losses encountered because a more severe loss is associated with higher congestion and higher relative one-way trip time (ROTT). In the delay trend scheme [6], the delay trend scheme could explicitly classify this packet loss as congestion loss or wireless loss, respectively, when the ROTT of the packet received after a loss occurs is relatively large or relatively small. If the ROTT of packets falls in an ambiguous region, the delay trend scheme classifies the packet loss according to the variation of ROTT.

In order to solve the performance degradation of TCP when used in heterogeneous network, there are many versions of TCP protocols, including NewReno [9], Vegas [21], Veno [22], Westwood[2][19], New Jersey [16][17], and SACK [11]. There are four phases [7][8] in a typical TCP congestion control algorithm: slow start, congestion avoidance, fast retransmission, and fast recovery. During the slow start phase, a TCP increases congestion window (*cwnd*) exponentially for each acknowledgment received. When *cwnd* reaches or exceeds slow start threshold (*ssthresh*), the control algorithm will enter the congestion avoidance phase. In the congestion avoidance phase, *cwnd* is at most increased by a segment per round-trip time. When congestion occurs, one-half of the current *cwnd* is saved in *ssthresh*. Additionally, if the congestion is indicated by a timeout, *cwnd* is set to one segment.

If three duplicate ACKs are received by the sender, it is a strong indication that a packet has been lost. So TCP performs a fast retransmission on the lost packet without waiting for the retransmission timer to expire. After the missing packets are transmitted in the fast retransmission phase, the fast recovery phase is performed to control the *cwnd* until a non-duplicate ACK arrives. The fast recovery phase is an improvement that allows high throughput under moderate congestion, because the received ACKs mean that other packets can still be received. When the third duplicate ACK is received, *ssthresh* is set to half of the amount of data that has been sent but not yet acknowledged. The lost segment is then transmitted and *cwnd* is set to *ssthresh* plus three maximum segment sizes (MSS). The *cwnd* is increased by one MSS whenever each additional duplicate ACK is received. The congestion window is inflated in order to reflect the additional segment that has left the network.

However, in TCP Reno [30], fast recovery is terminated while the next new ACK arrives, even if the ACK just acknowledges some but not all of the transmitted packets before the fast retransmission. Therefore, NewReno [9] keeps the highest sequence number before retransmission as *recover* to improve the fast recovery algorithm of Reno that incorporates a response into partial acknowledgments. In NewReno, fast recovery is terminated with either a retransmission timeout or an ACK that acknowledges all of the data that was outstanding at the start of fast recovery procedure. There are two cases for receiving a new ACK. If it acknowledges all of the data up to and including *recover*, fast recovery procedure is terminated in NewReno and *cwnd* is set to *ssthresh*. Then the congestion avoidance phase is performed. Otherwise, if this ACK is a partial acknowledgment, NewReno retransmits the unacknowledged segment. After that, it reduces the *cwnd* by the amount of new data acknowledged. Due to the improved mechanisms mentioned above, our proposed TCP congestion control algorithm is based on the NewReno to incorporate our new design.

TCP Westwood [2][19][30] (TCPW for short) is a modified version of TCP Reno. TCPW sender makes an end-to-end estimate of the available bandwidth along the connection by measuring the rate of returning ACKs. When a packet loss event is observed, TCPW uses the bandwidth estimate to set the *cwnd* and the *ssthresh*. The sender updates the *ssthresh* by the measured bandwidth and the minimum relative trip time instead of the half of the congestion window.

In [2], it shows that TCPW has better throughput than TCP Reno when there are wireless losses. This is because TCPW uses the current estimated rate as reference to reset the congestion window, but TCP Reno simply halves the congestion window. Besides, there are some derivations based on TCPW, such as TCP WestwoodNR [23] with NewReno feature and TCP Westwood+[24] with low-pass filter for the rate of returning ACKs to increase throughput and fairness over wireless links.

TCP Vegas [21] [30] utilizes the minimum RTT sample value observed during the connection to estimate the number of backlog packets in the buffer of bottleneck link by the difference between the expected throughput and the actual throughput. Two thresholds are used to estimate if the network experiences too many or few backlog packets followed by either decreasing or increasing the congestion window linearly to deal with the problem.

TCP Veno [22][30] improves the performance of TCP Reno by utilizing bandwidth estimation scheme of Vegas. In addition, it differentiates wireless losses from congestion losses by checking if the number of backlog packets is less than a threshold when a packet loss is detected. It implicitly means that the available bandwidth is still not fully utilized due to wireless losses.

TCP Jersey [16] consists of two key components, available bandwidth estimation (ABE) and congestion warning (CW) router configuration. CW is like Explicit Congestion Notification (ECN) configuration for network routers to mark the packets when the average queue length exceeds a threshold so as to help a sender to effectively differentiate packet losses caused by congestion, instead of wireless link errors. The ABE algorithm computes the current available bandwidth based on the time interval of ACK packets, the measured RTT, and the size of data that *n*th ACK acknowledges.

Instead of inter-ACK gap at the sender, TCP New Jersey [17] enhances TCP Jersey by computing ABE using t_n as the arrival time of the *n*th packet at the receiver. Thus the time interval $(t_n - t_{n-1})$ of data segments at the receiver can overcome the problem of ACK compression/delay in the reverse path. In addition, ESTCP [27] cooperates between the dynamic AIMD window controller at the source side and traffic controller at the network bottleneck node to achieve better fairness. However, not all routers have the ECN-like support in reality and the performance will degrade dramatically if the router of bottleneck link does not have the ECN-like algorithm implemented. Furthermore, bandwidth estimation algorithms that exploit the gap of inter-ACK or RTT do not work well, especially in asymmetric network where the bottleneck link may be on the reverse path.

3. The Proposed Packet Loss Classification and Wireless-Aware Congestion Control

In this section, a new packet loss classification algorithm extended from delay trend scheme is proposed with both increasing and decreasing trends. Besides, an improved congestion control for TCP transmission is proposed. The congestion control algorithm is based on the congestion control in NewReno with the assistance of the proposed PLC algorithm to regulate the transmitted bandwidth fairly and efficiently.

3.1. Packet Loss Classification Algorithm

The ROTT of received packets is exploited to assist packet loss classification, and two-trend detection method is utilized when it is ambiguous to distinguish the class of packet losses. The ROTT is measured as the time difference between the receiving time and the packet sending timestamp recorded in packet header plus a fixed bias. The end-to-end packet delay can be modeled as the summation of propagation delay, queuing delay, transmission delay, and router processing delay. Propagation delay is the time for the electromagnetic waves to traverse along the path. Router processing delay is required for the router to perform multiplex, reassembling, and packet forwarding. Transmission delay is the time required to send a packet into the link. They are usually constant for a given end-to-end path and the same packet length. The remainder, queuing delay, is the time for a packet to stay in a queue. The model for the end-to-end packet delay T_d is shown in (1):

$$T_d = \sum_i (T_{q,i} + \frac{P_s}{C_i} + T_{p,i}) + \frac{d}{s} \qquad (1)$$

where $T_{q,i}$ is the queuing delay of link *i*, $T_{p,i}$ is the router processing delay, P_s is the packet size, and C_i is the capacity of linke *i*. The final term is propagation delay where *d* is the length of physical link and *s* is the propagation speed of EM waves. We use measured ROTT and the trend of the received ROTTs which carries the information about the status of the queue in the bottleneck to classify packet losses in our proposed method.

The flow chart of the proposed method is shown in Fig. 1. Two thresholds $\mathrm{TG}^{\mathrm{up}}$ and $\mathrm{TG}^{\mathrm{low}}$ as in [6] are adopted to segment three regions to present the upper and lower bound of gray zone. When ROTT is larger than $\mathrm{TG}^{\mathrm{up}}$, it means that ROTT is larger than the time that is required when buffer is filled, and we classify the packet loss as a congestion loss. Otherwise, when ROTT is smaller than $\mathrm{TG}^{\mathrm{low}}$, the packet loss is classified as a wireless loss. When the measured ROTT falls in the gray zone between $\mathrm{TG}^{\mathrm{up}}$ and $\mathrm{TG}^{\mathrm{low}}$, "full search" method is used to calculate the two trends, increasing trend index ($incr_{trend}$) and decreasing trend index ($decr_{trend}$), as shown in (2) and (3).

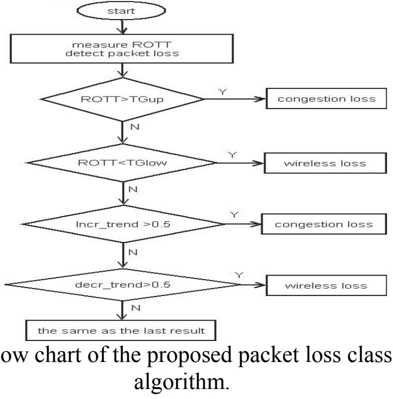


Fig. 1 Flow chart of the proposed packet loss classification algorithm.

$$incr_{trend} = \frac{\sum_{i=0}^{w-2}\sum_{j=i+1}^{w-1} I(D_i > D_j)}{\sum_{i=0}^{w-2}\sum_{j=i+1}^{w-1} I(1)} \quad (2)$$

$$decr_{trend} = \frac{\sum_{i=0}^{w-2}\sum_{j=i+1}^{w-1} I(D_i < D_j)}{\sum_{i=0}^{w-2}\sum_{j=i+1}^{w-1} I(1)} \quad (3)$$

where $I(X)$ is defined as 1 if X is valid, and 0 otherwise; D_i is the ROTT of the i^{th} packet and w is the search range.

When the ROTT of the current received packet falls in the ambiguous area, the increasing trend index and decreasing trend index are compared with the thresholds. We take the ROTTs of recent received w (16 in experiments) packets into consideration to estimate the network condition. If the increasing trend index is larger than an empirical threshold (=0.5), it means that ROTTs have an increasing trend caused by growing up queuing delay in the bottleneck and the packet loss is caused by network congestion with high probability. In addition, if the $decr_{trend}$ is larger than the threshold (=0.5) which means the queue in the bottleneck is not fully utilized; the packet loss is classified as a wireless loss. Otherwise, if ROTT is neither increasing nor decreasing, the packet loss is classified to be the same as the last PLC result. Although clock skew may make the measurement of ROTT imprecise, the result of increasing/decreasing trend detection is not affected because relative D_i is used during the trend detection.

3.2. The Proposed Wireless-Aware Congestion Control Algorithm

With the convenience brought by mobile devices, the heterogeneous networks with wired and wireless channels are more and more common in reality. In heterogeneous networks there are congestion losses caused by network congestion, and wireless losses caused by wireless channel error due to shadowing and attenuation. For the TCP congestion control algorithms which reduce the congestion window regardless of the nature of packet losses, it will mistakenly lead to performance degradation when a wireless loss occurs.

In order to solve this problem, the proposed PLC algorithm is utilized to classify packet losses, and then the TCP congestion control algorithm is modified in response to the packet losses caused by wireless errors to avoid unnecessary performance degradation. The wireless loss flag and packet loss number could be recorded in the reserved field of TCP header. The first bit of the reserved field is used to record the flag. This flag is kept constant for the packets that have the same acknowledgement number until the next packet loss event occurs. The remainder of the reserved field records the difference between the discontinuous sequence numbers. According to the flag and the packet loss number, the sender knows that wireless loss occurs and how many packets are dropped, and then the modified congestion control algorithm can be adopted to adjust the congestion window. In response to wireless losses, the retransmission policy remains the same. That is to say, the sender still retransmits the lost packet when three duplicate ACKs or partial ACKs are received, or when timeout of the retransmission timer occurs. However, the congestion window is increased as if a new ACK was received.

The proposed wireless-aware congestion control algorithm is shown in Fig. 2. The *seqno* means the sequence number of the next packet requested by the receiver. The *last_ack* is defined as the ACK number of the last received one. After the third duplicate ACK is received, *recover* is recorded as the highest sequence number (*highest_seq*) transmitted, and the fast recovery phase starts. The fast recovery phase means that one packet loss has occurred and different policy is used to recover successive packet losses in one window. The main idea of the proposed congestion control is to maximize the *ssthresh* after timeout occurs and also to maximize the size of *cwnd* after fast recovery ends if the loss is caused by wireless channel errors.

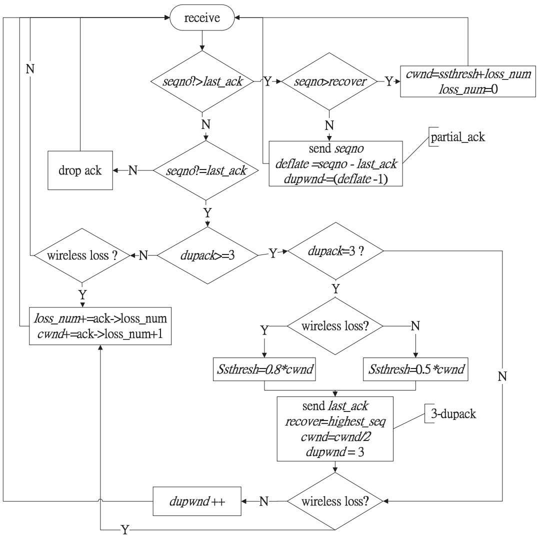


Fig. 2 A flow chart of the proposed wireless-aware congestion control for TCP

When the sender receives a packet, it checks whether this packet is a new ACK by comparing *seqno* with *last_ack*. If *seqno* is equal to *last_ack*, this ACK is a duplicate ACK. According to the number of the duplicate ACKs (*dupack*), different strategies are taken. If *dupack* is less than three, we check whether the loss is caused by network congestion or wireless error. If the loss is classified as a wireless loss, *cwnd* is increased by (*loss_num*+1) where *loss_num* is defined as the number of lost packets at the lost event and it is recorded in the header of corresponding ACK packet. Otherwise, *cwnd* is constant and this action is the same as the control in NewReno.

When the sender receives the third duplicated ACK, the control algorithm keeps *recover*, sets *ssthresh* to be the 0.8**cwnd* if a wireless loss is detected. Otherwise, the value of *ssthresh* is set as 0.5**cwnd*. We set higher *ssthresh* in case of wireless losses so as to acquire larger congestion window in the end of fast recovery. Then *cwnd* is set to be one-half of the current congestion window. The lost packet with sequence number *last_ack* is retransmitted, and the fast recovery phase is initiated.

In addition, the variable *dupwnd* is utilized to keep track of the increment of sending window whenever the duplicated ACK indicates network congestion, because the value of *cwnd* is still used to reset *ssthresh* when timeout occurs during fast recovery phase. Therefore, the sending window size is determined by the sum of *cwnd* and *dupwnd* during fast recovery phase. If the loss is caused by wireless error, we keep the accumulated *loss_num* and increase *cwnd* by (*loss_num*+1). Otherwise, the *dupwnd* is increased by 1 as in NewReno. The same action is also applied when the number of *dupack* is greater than three. The *cwnd* is increased if a wireless loss occurs and *dupwnd* is increased if congestion loss occurs.

If the *seqno* of new ACK is greater than *last_ack* during fast recovery phase, the control algorithm enters partial ACK action or exits fast recovery phase according to the comparison between *seqno* and *recover*. If *seqno* is less than or equal to *recover* (which means partial ACK), we retransmit the lost packet and reduce *dupwnd* by the difference between *seqno* and *last_ack*. Moreover, we set *cwnd* to the sum of *ssthresh* and *loss_num*, and reset *loss_num* to 0 when *seqno* is more than *recover*. Therefore, under the situation of wireless losses, *ssthresh* can be larger than the original one by *loss_num*/2 after timeout occurs, while the size of *cwnd* is increased by *loss_num* after fast recovery phase ends.

4. PERFORMANCE EVALUATION

The performance of the proposed PLC algorithm in terms of classification accuracy is presented first. The proposed wireless-aware congestion control algorithm for TCP is then compared with several congestion control algorithms used in different TCP protocols. Extensive simulation results are conducted at different wireless error rates/traffic patterns, and different network topologies.

4.1. Performance for the Parking-Lot Topology with Multiple Bottlenecks

We use the parking-lot topology with multiple bottlenecks to evaluate the proposed PLC algorithm and the delay trend scheme in literature. The new topology is shown in Fig. 3 and the link delay and link capacity for each link are also indicated. The wireless link is between node W8 and node M0.

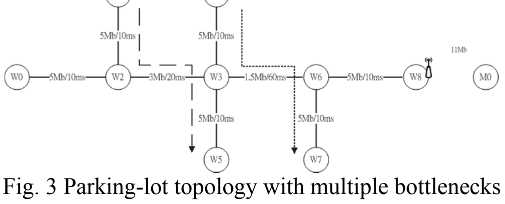


Fig. 3 Parking-lot topology with multiple bottlenecks

There are three TCP flows and all of them are FTP. The traffic is set as the following. TCP1 is from node W0 to node M0 during 0 to100 seconds. TCP2 is from node W1 to node W5 during 20 to 60 seconds. TCP3 is from node W4 to node W7 during 40 to 80 seconds. The total simulation time is 100 seconds. The error model of the wireless links described by a two-state Markov model of Gilbert-Elliott channel [14] is turned on at the second 60 and its average error rate is equal to 0.034.

Ac [3], which is defined as the ratio of the number of congestion losses correctly classified to the total number of congestion losses, is used to evaluate the accuracy of congestion loss discrimination. ***Aw*** is defined as the ratio of the number of wireless losses correctly classified to the total number of wireless losses and it is used to evaluate the accuracy of wireless loss discrimination. ***A*** is defined as the ratio of the number of total packet loss correctly classified to the number of total packet losses for the evaluation of the accuracy of overall discrimination. The accuracy of flow TCP1 is shown in Table 1. The proposed PLC algorithm shows better accuracy when there can be more than one bottleneck in the network.

Table 1 Accuracy comparison

	Ac	Aw	A
Delay trend scheme [6]	0.58	0.8	0.68
The proposed PLC	0.75	0.75	0.75

4.2. The Performance of the Wireless-Aware Congestion Control Algorithm

We use ns-2 as the simulation environment to compare the performance of the proposed algorithm with the congestion controls algorithms implemented in several TCP protocols. The performance metrics used here are throughput, network utility, and fairness index as defined in [13]. The packet size is set to 1000 bytes and queue size is set to 50 packets in these simulations.

4.2.1. *Performance at Various Wireless Error Rates*

The simulation topology is shown in Fig. 4 with link capacity and delay for each link indicated. There are two FTP flows that use the same TCP protocol. One flow TCP1 is from node S1 to node D1; the other flow TCP2 is from node S2 to node D2. The bottleneck with capacity 1.3Mb is the link between node N1 and node N2. The total simulation time is 100 seconds, and both of the two flows exist during the entire simulation time. When the simulation time reaches second 40, the wireless error model is activated to generate wireless losses. We compare total goodput, which is defined as the effective amount of data rate in application layer, of the proposed algorithm to several TCP variants: WestwoodNR, NewReno, Westwood, Westwood+, New Jersey, and Veno at different error rates, and the results are shown in Fig. 5. The proposed algorithm can discriminate wireless losses from congestion losses to avoid unnecessary performance degradation. Therefore, it outperforms other TCP protocols that also do not require congestion control support from the routers. The performance of the proposed algorithm is almost the same as the one of the TCP-New Jersey which differentiates wireless losses by the help of the intermediate routers to give a signal of congestion.

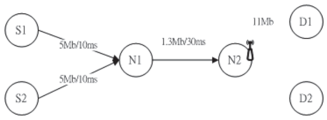


Fig. 4 Network topology used for different wireless error rates

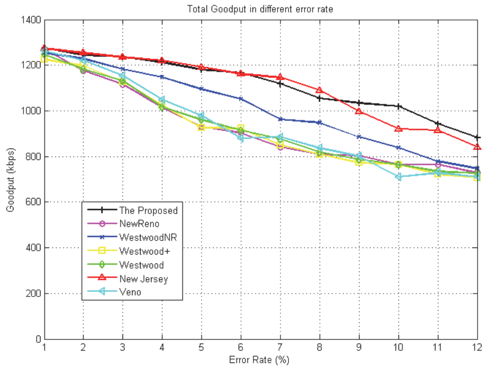


Fig. 5 Total goodput at different error rates. Note that New Jersey needs extra information from routers and it is not an end-to-end protocol.

Table 2 Performance comparisons over bottleneck link (error rate 0.06)

	The proposed	NewReno	WestwoodNR	Veno
Utility	0.94	0.72	0.84	0.72
Fairness	0.998	0.999	0.998	0.997

Table 3 Performance comparisons over bottleneck link (error rate 0.12)

	The proposed	NewReno	WestwoodNR	Veno
Utility	0.84	0.59	0.61	0.62
Fairness	0.998	0.999	0.997	0.999

We measure the data flow at the bottleneck link according to different flows separately, and the average values of the simulations at different error rates are listed in Table 2 and 3. In general, the proposed wirelesss-aware congestion control algorithm outperforms other TCP variants with respect to the throughputs and network utility. We can find that the proposed method shows better utility of the bottleneck link regardless of the wireless error rates. The results of fairness index of these methods are close. The proposed algorithm is better than the others, especially when the error rate is high. The main reason of the better performance is due to the increase of *cwnd* after fast recovery ends as well as the increase of *ssthresh* after timeout occurs if a wireless loss occurs.

4.2.2. Performance with Larger Topology

In this Section, we use Brite [15] topology generator to generate the network topology which is composed of 50 nodes, as shown in Fig. 6. There are 33 nodes (the square ones) in the core network and the others are leaf nodes. There are 3 WiFi base stations connected to the leaf nodes and there are 3 WiFi mobile hosts for each base station. Three leaf nodes are taken as senders to transmit TCP packets to their receivers at mobile hosts. The capacity for each link is generated according to heavy-tailed distribution in the range of 2Mbps and 10Mbps. The wireless error model with average error rate 0.1 is activated to simulate wireless channel error at second 40. Besides, there are 5 UDP flows with 500 Kbps as background traffic. The effective throughputs are shown in Table 4. The total goodput of the proposed algorithm is better than other TCP variants according to the simulation results.

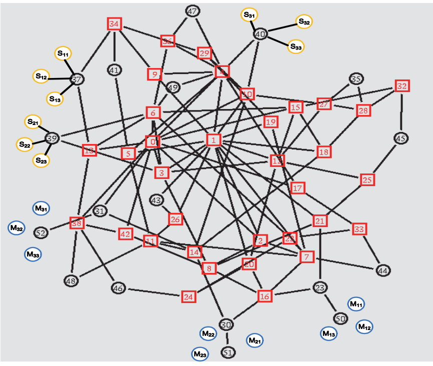


Fig. 6 The large scale network topology generated by Brite

Table 4 Goodput at the receiver

	The proposed	NewReno	WestwoodNR	Veno
M11	493.88	412.18	567.27	393.54
M12	486.23	406.27	362.68	387.72
M13	538.23	414.51	386.97	457.61
M21	563.19	481.07	515.52	428.9
M22	557.03	440.64	470.09	494.05
M23	556.28	422.75	543.64	355.19
M31	421.42	335.55	344.46	307.93
M32	408.44	306.43	417.09	379.98
M33	454.78	336.05	341.63	335.8
Total	4479.44	3555.42	3949.29	3540.69

5. CONCLUSION

A new packet loss classification algorithm is proposed with two-trend delay detection to differentiate congestion losses from wireless losses in the ambiguous region of ROTT distribution. A wireless-aware congestion control algorithm for TCP is further proposed. The proposed algorithm takes advantage of the packet loss classification to react correctly for real congestion situations. The algorithm adjusts the congestion window after fast recovery ends and increases the slow start threshold after timeout occurs. From the simulation results in Section 4, the proposed PLC algorithm shows better classification accuracy and it is more insensitive to the variation of the thresholds. The end-to-end wireless-aware congestion control algorithm also demonstrates better throughputs, better utility in the bottleneck link, and higher fairness than other congestion control algorithms in many TCP protocols, especially over heterogeneous networks. The only algorithm with performance similar to ours is the TCP

New Jersey, which, unlike our approach, is not an end-to-end algorithm and TCP New Jersey needs router support. Although the results are encouraging, the further evaluation is desired to take the mobility of the mobile host and handoff between base stations into consideration, and also prove the efficiency in real wired-wireless network environments.

ACKNOWLEDGMENT

This work was supported in part by the Ministry of Economic Affairs of Taiwan under Grant 102-EC-17-A-03-S1-214, and by the National Science Council of Taiwan under Grant 101-2221-E-009-086-MY3.

REFERENCES

[1] D. Mitzel, “Overview of 2000 IAB wireless internetworking workshop”, RFC 3002 (Dec 2000)

[2] C. Casetti, M. Gerla, S. Mascolo, M.Y. Sanadidi, and R. Wang "TCP Westwood: bandwidth estimation for enhanced transport over wireless links", in *Proceedings of ACM Mobicom*, 287-297(July 2001)

[3] S. Biaz, and N. Vaidya, "Discriminating congestion losses from wireless losses using inter-arrival times at the receiver," in *Proceedings of the 1999 IEEE Symposium on Application - Specific Systems and Software Engineering and Technology*, 10 –17(1999)

[4] Y. Tobe, Y. Tamura, A. Molano, S. Ghost, and H. Tokuda, "Achieving moderate fairness for UDP flows by path-status classification, " in *Proceedings of the 25th Annual IEEE Conference on Local Computer Networks*, 252-261(2000)

[5] S. Cen, P. C. Cosman and G. M. Voelker, "End-to-end differentiation of congestion and wireless losses", IEEE/ACM Trans Netw. 11(3), 703-717(Oct. 2003).

[6] H.-F. Hsiao, A. Chindapol, J. Ritcey, Y.-C. Chen, J.-N. Hwang, “A New Multimedia Packet Loss Classification Algorithm for Congestion Control over Wired/ Wireless Channels”, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005. ICASSP 2005, 1105-1108(2005).

[7] W. Stevens, “TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms”, RFC 2001 (Jan. 1997)

[8] M. Allman, V. Paxson, and W. Stevens, “TCP congestion control”, RFC2581 (Apr 1999)

[9] S. Floyd, T. Henderson, and A. Gurtov, “The NewReno modification to TCP’s fast recovery algorithm”, RFC3782 (Apr 2004)

[10] A. Bakre, and B. R. Badrinath, “I-TCP: Indirect TCP for mobile hosts”, in *Proceedings of the 15th IEEE International Conference on Distributed Computing System*, 1995, ICDCS 1995, 136-143(1995).

[11] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow, “TCP Selective Acknowledgment Options”, RFC2018, (Oct 1996)

[12] H. Balakrishnan, V. N. Padmanabhan, S. Seshan , R. H. Katz, “A comparison of mechanisms for improving TCP performance over wireless links”, IEEE/ACM Trans Netw. 5(6), 756-769 (Dec 1997).

[13] D. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks," Computer Networks and ISDN Systems, 17, 1-14 (1989)

[14] M. Zorzi, R. R. Rao, and L. B. Milstein, “On the accuracy of a first-order Markov model for data transmission on fading channels”, in *Proceedings of IEEE International Conference on Universal Personal Communications*, 211-215 (Nov 1995).

[15] A. Medina, A. Lakhina, I. Matta, J. Byers, Brite, http://www.cs.bu.edu/brite/

[16] K. Xu, Y. Tian and N. Ansari, “ TCP-Jersey for wireless IP communications,” IEEE J Select Areas Commun. 22, 747-756 (2004).

[17] K. Xu, Y. Tian, and N. Ansari, "Improving TCP performance in integrated wireless communications networks," Comput Netw., 47, 219-237 (Feb 2005).

[18] S. Henna, “A Throughput Analysis of TCP Variants in Mobile Wireless Networks,” in *Proceedings of IEEE International Conference on Next Generation Mobile Applications, Services and Technologies,* 279-284 (2009).

[19] S. Mascolo, M.Y. Sanadidi, C. Casetti, M. Gerla, and R. Wang, ”TCP Westwood: End-to-End Congestion Control for Wired/Wireless Networks,” *Wireless Networks Journal* , 8, 467-479 (2002).

[20] Y. Tian, K. Xu, and N. Ansari, "TCP in Wireless Environment: Problems and Solution," IEEE Commun Mag. 43, 2005.

[21] L. Brakmo and L. Peterson, “TCP Vegas: End to End Congestion Avoidance on a Global Internet,” IEEE J Select Areas Commun. 13, 1465-80 (1995).

[22] C. P. Fu and S. C. Liew, “TCP Veno: TCP Enhancement for Transmission over Wireless Access Networks,” IEEE J Select Areas Commun. 21, 216–28 (2004).

[23] UCLA Computer Science Department, TCP WESTWOOD - Modules for NS-2, http://www.cs.ucla.edu/NRL/hpi/tcpw/tcpw_ns2/tcp-westwood-ns2.html

[24] L. A. Grieco, and S. Mascolo, ”Performance evaluation and comparison of Westwood+, New Reno and Vegas TCP Congestion Control,” ACM SIGCOMM Comput Commun Rev. 34 , 25 –38(2004).

[25] S. Chen, X. Hei, J. Zhu and B. Bensaou, "Congestion Control in the Wired-cum-wireless Internet", in Heterogeneous Wireless Access Networks: Architectures and Protocols, ed. by Ekram Hossain (Springer,2009), p. 333

[26] S. A. Mondala and F. B. Luqman, “Improving TCP Performance over Wired–Wireless Networks,” Comput Netw. 51(13), 3799-3811(Sep 2007).

[27] H. Nishiyama, N. Ansari, and N. Kato, “Wireless Loss-Tolerant Congestion Control Protocol Based on Dynamic AIMD Theory,” IEEE Wireless Commun. 17(2), 7-14(Apr. 2010).

[28] J.-H. Yun, “Cross-Layer Explicit Link Status Notification to Improve TCP Performance in Wireless Networks,” *EURASIP Journal on Wireless Communications and Networking* (2009). doi: 10.1155/2009/617818

[29] K.-C. Leung and V. O. K. Li, “Transmission Control Protocol (TCP) in Wireless Networks: Issues, Approaches, and Challenges”, *IEEE Communications Surveys & Tutorials*, 8 (4), 64-79(2006).

[30] A. Afanasyev, N. Tilley, P. Reiher, L. Kleinrock, "Host-to-Host Congestion Control for TCP," IEEE Commun Surveys & Tutorials, 12(3), 304-342(Aug. 2010).

Performance Improvement in Subcarrier QAM Free Space Optical Links in the Presence of Atmospheric Turbulence and Pointing Error with Spatial Diversity and Beam Optimization

D. Chadha and Bhaskar Chakravarty
Department of Electrical Engineering, Indian Institute of Technology, Delhi, India

Abstract — Free-space optical communication suffers from irradiance variations caused by random atmospheric temperature fluctuations and pointing error due to jitter and misalignment between transmitter and receiver. In this paper, the simulation results for the average symbol error rate performance of Free Space Optical links over gamma-gamma turbulence fading channels in the presence of pointing error with subcarrier QAM are studied and numerical results are obtained using spatial diversity and beam optimization. The received signal at the receivers is combined using maximal ratio combining. Furthermore, optimum beamwidth is calculated taking into account various metrics such as the electrical signal-to-noise ratio, the normalized jitter, and the SER. The results obtained can be a useful outcome for FSO system designers in order to limit pointing error effects and achieve an optimum performance.

Index Terms — Diversity Reception, Free-space optics, Fading, MRC, Pointing error, QAM

1. INTRODUCTION

Free-space optical (FSO) communication has emerged as a promising and commercially viable technology in today’s communication infrastructures. It can provide high-speed links for a variety of applications; can be considered as a supplement or an alternative to RF for the next generation broadband in order to support large bandwidth, unlicensed spectrum, excellent security, and quick and inexpensive setup. However, atmospheric turbulence produces scintillation of the transmitted optical beam at the receiver end, severely degrades the link performance. Due to line-of-sight (LoS) connection and the directional reception of narrow FSO beams another cause of concern is the pointing error (PE) which arises due to misalignment between the transmitter and receiver due to weak earthquakes, dynamic wind loads resulting in sway of high rise buildings that causes vibrations of transmitted beam and thus misalignment.

An appropriate selection of the modulation technique is vital to circumvent turbulence induced fading. Though the on-off keying (OOK) scheme is the simplest and extensively used modulation technique, but it requires an adaptive threshold scheme to perform optimally in atmospheric turbulence induced fading. This adaptive threshold is complex to implement and practically not suitable. Therefore, it is a reasonable approach to use modulation techniques that carry the information in the phase or the frequency of the RF carrier signal. The phase shift keying (PSK) or the quadrature amplitude shift keying (QAM) based subcarrier intensity modulation (SIM) requires no adaptive threshold scheme, thereby offering superior performance compared to OOK in the presence of atmospheric turbulence induced fading channels. They require lower bandwidth and have higher spectral efficiency as compared to binary modulation schemes. Further, M-ary QAM delivers better BER performance due to more distance between constellation points for values of M >4. However, the penalty paid is higher SNR to attain a desired BER performance. Further, multiple RF SIM is the preferred choice when increased capacity is more important than the power requirement. Also, QAM is becoming the standard for 4G wireless communications, which is another driver, in order to have seamless integration with the system.

Diversity schemes with different combining systems have received much attention in RF due to their ability to mitigate the performance degradation of multipath fading through diversity. The random pointing error in FSO systems has a profound effect on the BER performance of the atmospheric fading channel. This can also be improved by using diversity and optimum combining techniques at the receiver. In the present paper, therefore, study is carried out using diversity with single-input multiple-output (SIMO) and maximal ratio combining (MRC) for the FSO link affected by turbulence fading and pointing error.

The remainder of this paper is organized as follows. The channel, system model and beam optimisation are described in Section II followed by simulation results and conclusions in section III and IV, respectively.

2. CHANNEL AND SYSTEM MODEL

A. Channel Model

The channel state *h* models' the random attenuation of the propagation channel. In our model, *h* arises due to two factors: atmospheric turbulence $h_{a,}$ and pointing errors h_p. The combined optical channel model is defined as:

$$h = h_a.h_p \qquad (1)$$

In (1), '*h*' is the normalized channel fading coefficient considered to be random but constant during the symbol duration. The coefficients; h_a and h_p are random with distributions discussed in the following paragraphs.

The atmospheric turbulence follows the Gamma–Gamma distribution which recently has emerged as a useful turbulence model as it has excellent fit with measured data over a wide range of turbulence conditions [1]. According to the Gamma–Gamma model, the irradiance '*I*' can be modeled as weak eddies induced irradiance fluctuation modulated by strong eddies induced irradiance fluctuation. The probability density function (PDF) of a normalized Gamma–Gamma random variable *I* is given as [1].

$$f_{h_a}(I) = \frac{2(\alpha\beta)^{\frac{\alpha+\beta}{2}} I^{\left(\frac{\alpha+\beta}{2}-1\right)} K_{\alpha-\beta}(2\sqrt{\alpha\beta I})}{\Gamma(\alpha)\Gamma(\beta)} \qquad (2)$$

where Γ(.) is the Gamma function and $K_{\alpha\text{-}\beta}$(.) is the modified Bessel function of the second kind of order α-β. The shaping parameters α and β are related to the Rytov variance (σ_l^2) and are given by [1]

$$\alpha = [\exp\left(\frac{0.49\sigma_l^2}{\left(1+1.11\sigma_l^{12/5}\right)^{7/6}}\right) - 1] \qquad (3)$$

$$\beta = [\exp\left(\frac{0.51\sigma_l^2}{\left(1+0.69\sigma_l^{12/5}\right)^{5/6}}\right) - 1] \qquad (4)$$

Pointing errors which can arise due to mechanical misalignment, errors in the tracking system, or due to mechanical vibrations present in any system, is composed of two components [2]: a fixed error, called *boresight*, and a random error, called *jitter*. The spatial intensity profile of the beam is assumed to be Gaussian with a beam waist 'w_z' at the receiver plane at a distance '*z*' from the transmitter with a circular aperture of radius 'r'. The probability density function of the pointing error is given by [2]

$$f_{h_p}(h_p) = \frac{\gamma^2}{A_0^{\gamma^2}} h_p^{\gamma^2-1}, \quad 0 \leq h_p \leq A_0 \qquad (5)$$

where,

$$\gamma = w_{z_{eq}}/2\sigma_s \qquad (6)$$

h_p represents fraction of the power collected by the detector, σ_s is the pointing error displacement standard deviation due to jitter at the receiver. $w_{z_{eq}}$ is the equivalent beam width at the receiver. The parameters $w_{z_{eq}}$ and A_o are given in [2].

B. System Model

The block diagram of the transmitter of subcarrier FSO system for a single RF carrier and employing QAM-SIM is shown in Fig.1. At the transmitter, the serial data signal is converted to two parallel streams and the radio frequency signal is first pre-modulated with the data signal *d*(*t*). After proper biasing, this RF signal *s*(*t*) is used to modulate the irradiance of a continuous wave optical laser beam. The transmitted irradiance will have the following waveform:

$$i_t(\mathrm{t})=\mathrm{P}[1+\xi s(t)], \qquad (7)$$

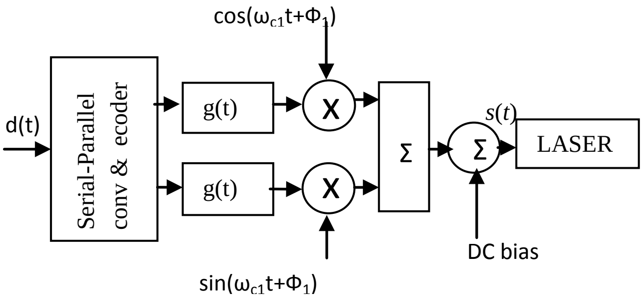


Fig.1.Transmitter of a single subcarrier SIM FSO system.

where 'P' is the transmitted optical power which is normalized to unity and 'ξ' is the modulation index satisfying the condition -1≤ ξ$s(t)$ ≤ 1 in order to avoid over-modulation. Here QAM is chosen as the electrical modulation format with the in-phase and quadrature-phase of the RF signal. The output signal of the QAM modulator can be written as [3]

$$s(t) = s_i(t)\cos(2\pi f_c t) - s_q(t)\sin(2\pi f_c t) \tag{8}$$

In Eq. (8), $s_i(t) = \sum_{j=-\infty}^{\infty} a_j(t) g(t - jT_s)$ and $s_q(t) = \sum_{j=-\infty}^{\infty} b_j(t) g(t - jT_s)$, where $a_j(t)$ and $b_j(t)$ are the in-phase and quadrature components of the jth data symbol, respectively, g(t) is a shaping pulse, and T_s is the symbol interval.

At the receiver shown in Fig.2, in a spatial diversity system, aperture area of each detector in the N-receiver system is assumed to be A/N, where A is the aperture area of detector under single-transmitter single-receiver link and the background radiation noise on each link with detector diversity is 1/N that of a SISO link resulting in $\{n_i(t)\}_{i=1}^{N} \sim N(0, \sigma_i^2/N)$.

This approach is particularly valid for the case where noise from the background radiation is the dominant source as is the case for FSO. By assuming identical PIN photodetector on each link, the individual detector output is given by [4]

$$i_{ri}(t) = \frac{\mathcal{R}}{N} h\{1 + A_j g(t)\cos(w_{cj} t + \theta_j)\} + n_i(t) \tag{9}$$

where, i= 1, 2, 3, . . . , N. '$\mathcal{R}$' is the responsivity of the photodetector, $n(t)$ is the additive white Gaussian noise (AWGN).

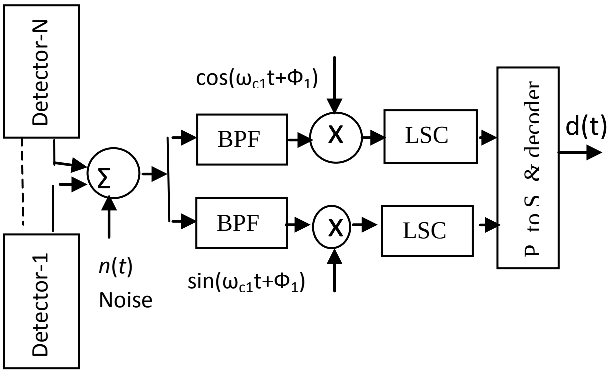


Fig.2. Diversity receiver of single subcarrier SIM FSO system.(LSC-Low pass filter, Sampler, Comparator)

After removing the DC bias, the electrical signal is first demodulated by an electrical QAM demodulator. It is then sampled to recover the transmitted data. The MRC combiner weights each output signal $\{i_{ri}(t)\}_{i=1}^{N}$ from each link by gain $\{a_i\}_{i=1}^{N}$proportional to the received intensity. The weighted signals are then co-phased and coherently added to obtain the combined output current. In a SIMO MRC system, channel state information (CSI) is perfectly known at the receiver. The received signals are combined in such a way that the signal-to-noise ratio (SNR) at the receiver combiner output is maximized.

C. Beam Width Optimisation

To overcome the problem of misalignment and keep line-of-sight between the transmitter and receiver, the beamwidth and transmitted power need to be increased. However, a wide beamwidth increases the required signal-to-noise ratio (SNR), leading to increased cost and complexity, whereas, a narrow beam may result in the signal outage. A beam width which is larger than receiver aperture, reduces the SNR due to spreading of optical power. On the other hand a beam width which is smaller than receiver aperture radius would result in the increase of ambient noise, reducing the effective SNR. Hence, a proper optimization of the beamwidth is required.

3. SIMULATION RESULTS

The RF SIM-QAM SIMO FSO system described above is simulated using Matlab with single subcarrier. The system parameters and that of the atmospheric turbulence and pointing error (considering a fast tracking system with negligible bore sight error) are given in Table 1.

The simulation has been carried out for low turbulence (σ_l^2 =0.2) and moderate turbulence (σ_l^2=1) by varying jitter standard deviation σ_s for each case.

The pulse shaping function is assumed to be rectangular in the simulation.

Table 1: Simulation Parameters

Data rate	1Mbps
Laser wavelength	1550nm
Distance between Tx and Rx	1Km
Receiver diameter	8cm
Jitter standard deviation, σ_s	4cm, 6cm, 8cm
Variance of Turbulence	$0.2 \leq \sigma_l^2 \leq 1$
Number of receivers	2

In Fig.3, we have shown the performance improvement of the FSO SIMO system over the single-input-single-output (SISO) system. Fig. 3 has plots of the SISO system with and without PE. It is observed that the performance of the FSO system degrades considerably by a value of 9dB because of pointing error at a SER of 10^{-4} for the SISO syatem. With receiver diversity of order two and MRC combining, a diversity gain of the order of 7-13dB is achievable for jitter values of 8cm $\geq \sigma_s \geq$4cm at a SER of 10^{-4}.

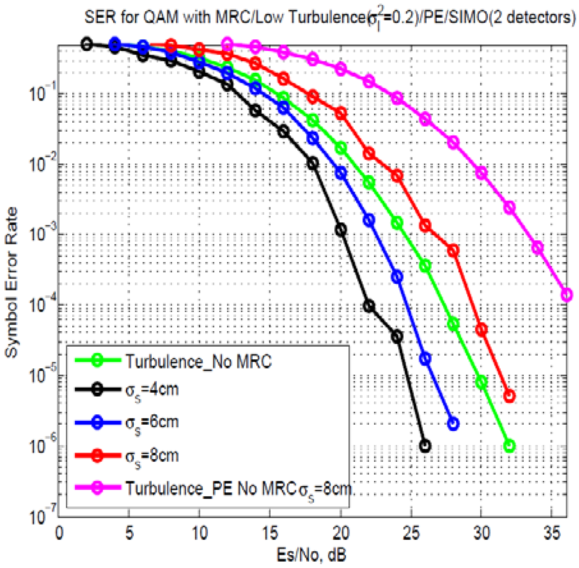


Fig.3. SER performance of QAM-SIM with SIMO-MRC in the presence of low turbulence and pointing error.

In Fig. 4, once again, similar order of diversity gains has been obtained for the FSO system with PE for moderate turbulence conditions.

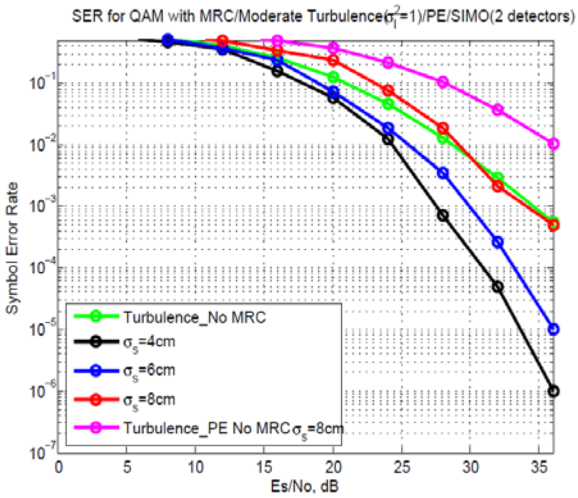


Fig.4.SER performance of QAM-SIM with SIMO-MRC in the presence of moderate turbulence and pointing error.

Next, we have obtained the results for the optimisation of the beam. From Fig.5 we observe that the value of optimum beam-width remains constant even on varying the turbulence from low to a high value keeping the pointing error jitter constant. The minimum beam-width remains unchanged for different atmospheric turbulence at 1.7 for a fixed PE. Therefore, turbulence has no effect on the optimum beam-width required.

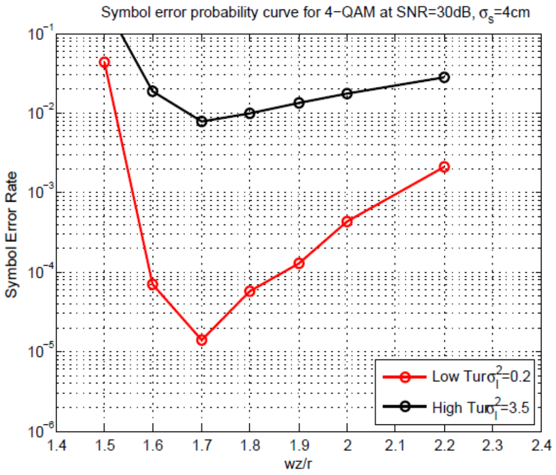


Fig.5. SER vs w_z/r curves for QAM-SIM at fixed SNR and jitter standard deviation with varying turbulence.

In Fig.6, it can be seen that by keeping the turbulence constant at a moderate level and varying the pointing error jitter standard deviation by only a small amount, the optimum beam width required has grown very large in size (from 1.6 times to 2.85 times the receive aperture radius**).** This is due to the fact that a higher jitter results in larger variations in radial displacement of the incoming beam center demanding a larger beam width for optimum SER performance.

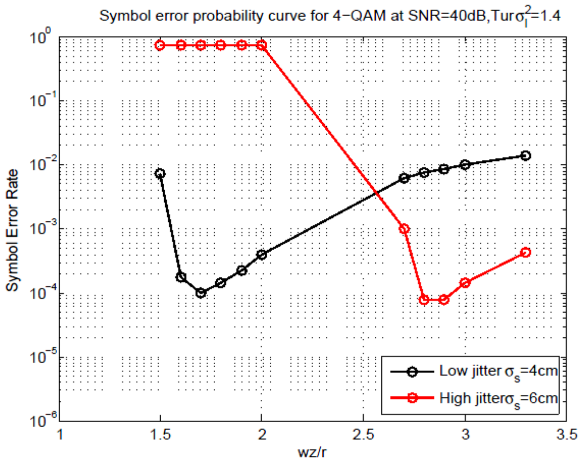


Fig.6. SER vs w_z/r curves for QAM-SIM at fixed SNR and turbulence with varying jitter standard deviation.

4. Conclusion

In this paper, SER performance of QAM-SIM FSO link over Gamma-gamma atmospheric turbulence channels with receiver diversity and MRC has been investigated. The effect of varying the pointing error standard deviation on the SER performance for low and moderate turbulence has been obtained. It is concluded that the degradation in SER due to turbulence induced fading with PE can be reduced considerably by employing receive diversity. Also, the performance of the FSO link with pointing error has been investigated with the beam-width radius. The effect of varying the pointing error standard deviation and turbulence on the beam-width has been obtained. The effect of misalignment may be reduced by using broad transmitted beam but this would be at the expense of higher transmitting power. In this regard, beam width optimization was found to be a viable option for improving SER performance along with automatic tracking systems. It is concluded that, increase in turbulence doesn't have any effect on optimum beam width required whereas increase in pointing error jitter increases the optimum beam width required.

5. References

[1] Nistazakis H.E, Tombras G.S, Tsiftsis T.A.: 'Performance analysis of free-space optical communication systems over atmospheric turbulence channels', *IET Commun.*, 2009, vol. 3, 8, pp. 1402–1409.

[2] Ahmed A. Farid , Steve Hranilovic ,"Outage capacity optimization for free-space optical links with pointing
errors", *J. of Lightwave Technol*, vol. 25, no. 7, July 2007,pp. 1702-1710.

[3] Md. Zoheb Hassan, Xuegui Song, and Julian Cheng, "Subcarrier intensity modulated wireless optical communications with rectangular QAM", *J. Optical Commun Netw*, vol. 4, no. 6, June 2012, pp. 522-532.

[4]W.O. Popoola, Z. Ghassemlooy, J.I.H. Allen, E. Leitgeb and S. Gao. "Free-space optical communication employing subcarrier modulation and spatial diversity in atmospheric turbulence channel", *IET Optoelectron.*, vol. 2, no. 1, Feb 2008, pp.16-23.

An Implementation of the TDMA Baseband Modem for Voice Relay

Yeonbo Kim [1], Byoungchul Ahn[2] and Yong-Wook Bae[3]
[1] School of Electronic and Electrical Engineering, Daegu University, Gyungsan, Korea
[2] Department of Computer Engineering, Yeungnam University, Gyungsan, Korea
[3]Department of Avionic Electronics Engineering, Kyungwoon University, Gumi, Korea

Abstract - *This paper presents an implementation of the TDMA baseband modem to support M:N voice communications by using ad-hoc method. The designed baseband modem is applied to communicate a 14-node relay network. Each node can send and receive signals by its assigned time slot. When odd-numbered nodes of 14 nodes can transmit signals at the same time, even-numbered nodes can receive signals, and vice versa. To implement the relay function, the baseband modem has a cycle network controller, a priority dual-port I/O controller with an asynchronous serial device and data buffer memories, and has been programmed on Xilinx Spartan-6 FPGA with Verilog HDL. Since the baseband modem configures a network for maximum 14-hop relay in serial, the communication distance is extended up to 1.4 Km. The baseband modem is tested and verified its functions for voice communications, and its measured maximum delay time is less than 230.4msec for the end-to-end voice transmission.*

Keywords: Wireless Ad-hoc Network, Relay Protocol, Baseband Modem, WPAN

1 Introduction

Wireless personal area networks(WPANs) are used in many places such as small group meetings, short distance multi-way communications, remote speakers and so on. Typically WPANs do not have a function to join the networks or relay voices using ad-hoc function. Also WPANs are very limited distance and outputs 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.

To implement ad-hoc function for voice communications, there are several technical challenges in ad-hoc networks. First, ad hoc networks have 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 the wireless network 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 an ad-hoc network is short in the wired network. There is also a disadvantage in a multi-hop wireless networks 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. There are no products to support the relay protocol and ad-hoc function among groups.

As the processors speeds are increased, it is possible to overcome the above technical challenges. Ahn has proposed a relay protocol based on TDMA technology with ad-hoc function[1]. This paper presents an implementation of the baseband modem for voice communications using the protocol

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 the possibility, and the chip of BSR has not been implemented. Kwong *et al.* have used multi-path routing protocol called MSDR in order to 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. Some researches have used only one network interface due to cost. 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 are no related papers which have been implemented the baseband modem with the relay and ad-hoc functions. This paper presents an implementation of the TDMA baseband modem using Xilinx Spartan-6 FPGA and Verilog HDL. The basic protocol is described in Section 3 and the design of baseband modem is described in Section 4 and 5.

3 Network Protocol

Figure 1 shows the TDMA network cycle for the relay protocol for voice communication. One network cycle has 15 slots. The time length of each slot is 15.36*msec* and one network cycle time is 230.4*msec*. Each slot has 15 frames. Each frame is 1.0*msec*. Each node uses one of its own control section assigned in its frame. Slot 0 is used for the start of the network cycle(SoC) and used 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 the other slots is used for the control frame, which is used by the master node and all slave nodes to exchange network information. The other 14 slots are used for transmitting voice data.

The basic structure of each frame as shown in Fig 2 consists of five sections, which are preamble, SyncWord, Header, Payload and EoF(end of frame). The length of the frame is 1024-bit(128B). The preamble is used for stabilizing time of the frequency generator for RF modules and users can specify the length of preamble for specific RF chips. SyncWord is 4-byte synchronous and has a special pattern in order to synchronize the payload data. The Header has the addresses of the sender and the parent node.

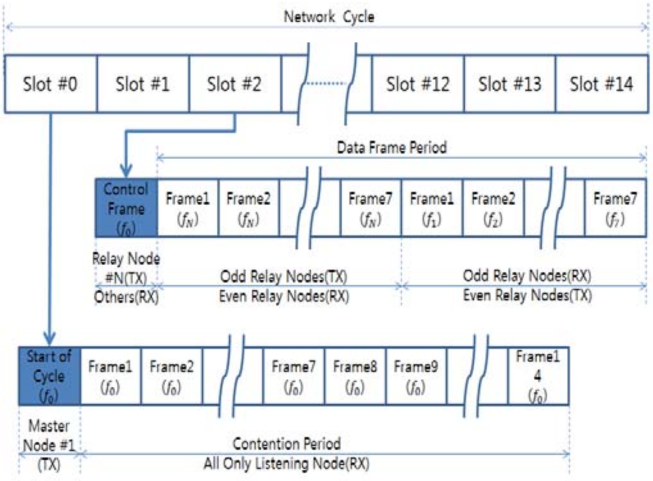


Figure 1. Network Cycle of Relay Protocol

Preamble (24Bytes)	SyncWord (4Bytes)	Header (4Bytes)	Payload (92Bytes)	EoF (4Bytes)

Figure 2. Basic Structure of a Frame

3.1 Start of Cycle

The SoC is generated and transmitted by the master. It indicates the start of the network cycle and all slave nodes should synchronize this signal to work with the network. This includes special codes for security, which are group code and message security code. To synchronize this cycle, the frequency is a designated frequency or default frequency to all slave nodes.

The contention frames followed by the SoC are used to request to join the network and total contention frames are 14 frames. The slave which joins the network sends the message “connection request (JoinREQ)” to the master and the master replies the message “connection permission (JoinACK)” with its node number and network address for routing. For this purpose, the designated frequency is used. After new connection is set up, the frequency hopping is used.

3.2 Control Frames

The control frame is generated and sent by the master and all slave nodes participated in the network. The control frame is used to manage the network using the designated frequency. This frame synchronizes the network, transmits and receives data. The master node and relay nodes use the payload of the control frame to exchange routing information to relay data and ad-hoc function.

3.3 Data Frames

Data frames are divided into 14 small frames to transmit and receive voice data. As shown in Figure 1, 14 data frames are divided into two sections to relay voice data. Two sections have different roles upon hop number from the master. If hop numbers are odd number, the first section (Frame1 ~ 7) is used to transmit voice data generated by odd numbered nodes with designated frequency. The second section (Frame8 ~ 14) is used by even numbered nodes and used to receive voice data. After this stage, odd numbered nodes are changed to the reception mode and even numbered nodes are changed to the transmission mode. The frequency hopping is used for this operation.

4 Baseband Modem Design

The baseband modem for the relay protocol is implemented on Xilinx Spartan-6 using Verilog HDL. The baseband modem block diagram is shown in Figure 3. The modem has a network cycle controller, a priority dual-port I/O controller, an asynchronous serial transceiver, and data buffer memories.

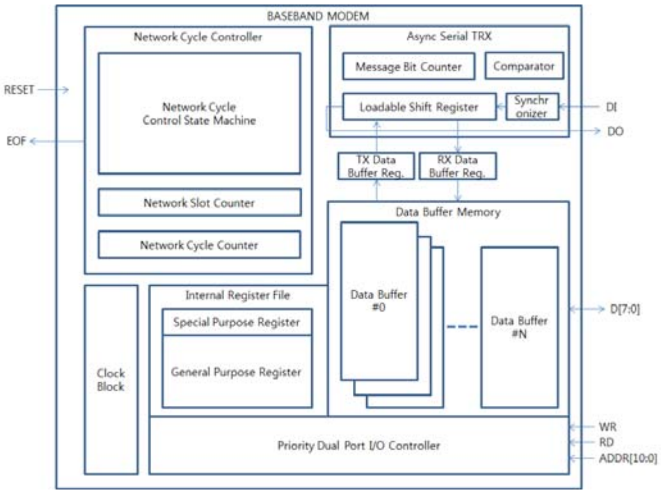


Figure 3. Block Diagram of the Baseband Relay Modem

4.1 Asynchronous Serial Tranceivers

The Baseband modem is designed to communicate 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 SyncWord and a shift register is used to convert serial data to parallel data or vice versa.

4.2 Buffer Memory

To interface the baseband modem with a processor, buffer memories are implemented. The buffer memories store voice data from the processor to the RF module or store data from the RF module to relay or play voice data. Therefore, seven 512-bit buffers are implemented to handle data simultaneously. Each buffer is assigned to one of seven data frames and each buffer is used to a dedicated relay node participating in the voice relay. In order to reduce the size of logic gates, the data buffer memory is implemented by a single port memory. When a dual-port memory is implemented for fast data access, the power consumption and the number of logic gates increased dramatically.

4.3 Priority Dual-Port I/O Controller

To meet the processing time of the baseband modem, a priority I/O controller is implemented. The priority I/O controller has designed to give higher priority to a network cycle controller when both the network cycle controller and the processor access the buffer memories at the same time.

4.4 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 network service for slave nodes to join TDMA network cycles. If the slave mode is set, it searches the SoC with the designated frequency. After it receives the SoC, it sends out the message “connection request (JoinREQ)” and searches the control frame of the master and retrieves its unique address assigned by the master. If a slave node receives all information from the master, it synchronizes the network cycle times and operates for transmitting and receiving voice data. The slave node decodes the control frame of the master and compares the SoC continuously, and compares the state of the internal current modem continuously. If the address of slave node is not match the address received from the master, the slave node changes its state to searching mode again since it considers that it is not synchronized to the master. This means that it must be initialized and rejoin the network and performs synchronization process again.

5 Experiment

In order to evaluate relay protocol functions of the baseband modem, it has been assembled on a system board shown in Figure 4. The system board has a Cortex-M4 processor from ST Micronics, a XC6SLX9 FPGA chip from Xilinx, a CC2400 chip from TI and a WM8976 audio codec chip from Wolfson Microelectronics. The baseband modem is programmed on Xilinx FPGA using Verilog HDL. The CC2400 chip is 2.4GHz RF module and the WM8976 chip is used for encoding and decoding voice data. Cortex-M4 processes routing function to relay voice data and controls the system.

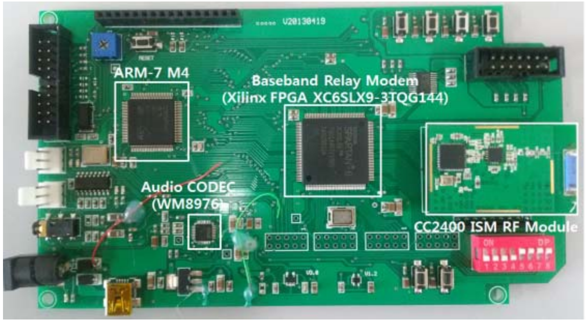


Figure 4. Evaluation Board

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. Slave nodes are joined the network one by one. In Figure 5, a relay node

#2 is connected to the master and a relay node # 4 is connected to the relay node # 2.

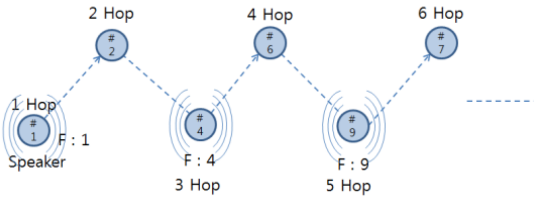


Figure 5 Network topology to test voice relays

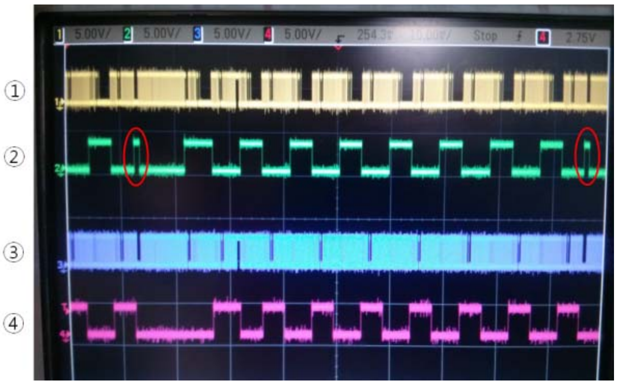


Figure 6. Output Signals of Network Cycle Controller

Figure 6 shows the output signals measured on FPGA pins when the master node transmits data to relay node #2 at Figure 5. At Figure 6, the number ① shows the audio data signal to be transmitted and received by the master. The number ② shows the transfer mode of the master node. The high state of pulses means that master node is a transmission mode and the low state of the pulses means that the master node is a receive mode. The narrow pulses with a circle illustrate the SoC of the first frame of the network cycle. Audio data signals received by the relay node #2 are shown at the number ③**.** The number ④ shows that the relay node #2 is operating in the communication mode.

When the RF module is set to the maximum power, the radius of RF reaches about 100*m*. Experiment results show that the maximum distance using the implemented baseband modem reaches about 1.4*Km* when 14 nodes are connected in serial. When the data bit rate is 1Mbps, the end-to-end delay time from the first node to the 14th node is about 230.4*msec*. The average node-to-node delay time is about 17.7*msec*. When this baseband modem is applied to the personal wireless communication, the delay time guarantees the QoS.

6 Conclusion

In this paper, a TDMA baseband modem has been implemented on FPGA and tested its performance to relay voice data for WPAN environment. The maximum relay nodes are configured up to 14 nodes with ad-hoc function. Experiment results show that the maximum end-to-end delay time of 14 nodes is 230.4 *msec* with the data rate of 1Mbps. The PWAN network can be configured as various topology such as line, start or tree and so on. Also the modem can be used with various RF modules for specific applications with different data rates. It requires to re-design in a chip to apply to WPAN applications.

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] Person. “Title of Research Paper”; name of journal (name of publisher of the journal), Vol. No., Issue No., Page numbers (eg.728—736), Month, and Year of publication (eg. Oct 2006).

Region Authority (RA) Collaborated Certificate Organization and Management in VANET

Shahnawaj Khan
CSE Department
National Institute of Technology, Hamirpur
Hamirpur, India
shahnawaj.khan1990@gmail.com

Naveen Chauhan
CSE Department
National Institute of Technology, Hamirpur
Hamirpur, India
naveenchauhan.nith@gmail.com

***Abstract*- Vehicular ad hoc networks (VANETs) are receiving increasing attention from academics due to the various applications and potential tremendous benefits they offer for future VANET users. Safety information exchange enables life-critical applications, such as the alerting functionality during medical emergencies, and thus, plays a key role in VANET applications. In a VANET, vehicles rely on the integrity of receiving data for deciding when to present alerts to drivers. The communication between car to car, car to the roadside unit done through wireless communication. That is why security is an important concern area for vehicular network application. For authentication purposes large amount of bandwidth is consumed and the performance becomes low. In VANET some serious network attacks such as man in the middle attack, masquerading is possible. In this paper various previous researches done in this area are analyzed and compared on the basis of drawbacks of those approaches. After that the different issues on VANET are discussed and finally conclude with proposed idea.**

Keywords-Security, Region authority (RA), on board unit (OBUs), certificate revocation lists (CRLs).

I. INTRODUCTION

Vehicular Ad-Hoc Network is a special class of Mobile Ad-Hoc Networks (MANETs) in which communication link is established between road side units (RSUs) and on board units (OBUs), OBUs to OBUs in a short range of 100 to 300 m and between RSUs to RSUs. To enable application for safety, traffic, driver assistance, infotainment vehicular communication is evolving very rapidly. But the race of providing various services raises the security concern and makes VANET vulnerable to various attacks like jamming, forgery, privacy violation, on board tampering. Existing protocols to secure VANETs resolves these issues up to some extent but raises some concerns on the basis of which, this paper reviews the different schemes developed for VANET.

More specifically, the purpose of the paper is to survey the literature, and provide an overview of the extent of the research done in the area of VANETs and also provide some protocols to resolve the security issues. This paper is organized as follows: In the second section the challenges for the security in VANET are discussed. In the third section, previous work that is close to this approach has been discussed. In the fourth section, system model which we assumed for the proposed scheme is discussed. In the fifth section, evaluation criteria are discussed based on which proposed scheme is evaluated. In section sixth proposed scheme is discussed. In section 7 paper is concluded.

II. CHALLENGES

The most significant challenges of VANET are

A.Network Volatility

The connectivity among nodes can often be highly transient and a one-time event (same vehicles may not get the chance to communicate again). For example, two vehicles (nodes) traveling on a highway may remain within their transceiver range, or within a few wireless hops, for a limited period of time. Hence password-based establishment of secure channels, gradual development of trust by enlarging a circle of trusted acquaintances, or secure communication only with a handful of endpoints and may be impractical for Securing vehicular communication (VC).

B. Authentication vs. Privacy

In the process of providing authentication for communication between OBUs, The privacy of the sender can be revealed and can pose threat to the sender by many ways like tracking someone's location, journey details.

C. Delay Sensitive Applications

Many of the safety and driver-assistance applications Pose strict deadlines for message delivery or are Time-sensitive. Security mechanisms must take these constraints into consideration and impose a low processing and messaging overhead.

D. Network Scale

The scale of the network, with roughly a billion vehicles around the globe, is another challenge.

E. Heterogeneity

The heterogeneity in VC technologies and the supported applications are additional challenges, especially taking into account the gradual deployment. With nodes possibly equipped with cellular transceivers, digital audio and Global Positioning System (GPS) but with the current standard architecture using the vehicular public key infrastructure (PKI)

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

and tamper proof device (TPD) and various protocols these problems are resolved up to some point.

III. PREVIOUS WORK

In VANETs, the primary security requirements are identified as entity authentication, message integrity, non-repudiation, and privacy preservation. The PKI is the most viable technique to achieve these security requirements.

In [6] author proposed TACKs for certificate organization and vehicle revocation in a VANET, which we consider to be the most relevant and closely related scheme to the work we propose in this paper. TACK adopts a hierarchy system architecture consisting of a central trusted authority and regional authorities (RAs) distributed all over the network. The authors adopted group signature where the trusted authority acts as the group manager and the vehicles act as the group members. Upon entering a new region, each vehicle must update its certificate from the RA dedicated for that region. The vehicle sends a request signed by its group key to the RA to update its certificate, the RA verifies the group signature of the vehicle and ensures that the vehicle is not in the current Revocation List (RL). After the RA authenticates the vehicle, it issues short lifetime region-based certificate. This certificate is valid only within the coverage range of the RA. It should be noted that TACK requires the RAs to wait for some time, e.g., 2 seconds, before sending the new certificate to the requesting vehicle. It restricts the vehicle to send messages to neighboring vehicles within this period, which makes TACK not suitable for the safety applications. Also, TACK requires the RAs to completely cover the network, otherwise, the TACK technique may not function properly. This requirement may not be feasible especially in the early deployment stages of VANETs.

In [1] author considers the deployment stage of VANET and proposes 3 protocols revocation using tamper proof device (RTPD) which uses the tamper proof device (TPD) to revoke all the certificates of the malicious vehicle with the help of radio or FM to broadcast in case of vehicle is not in the range of road side units (RSUs), distributed revocation protocol (DRP) which uses group based revocation technique to revoke the certificate. In case if any vehicle is suspected of doing malicious activity by its neighbor and if then the numbers of neighbors are greater than a certain threshold then they inform it to the CA to remove its certificate. The last is revocation using compressed certificate revocation list (RCCRL) which uses the distribution of the only updated and compressed list. For privacy it proposes using a set of anonymous keys that change frequently. These keys are preloaded in the vehicle's TPD for a long duration. For authentication vehicles will sign each message with their private key and attach the corresponding certificate. To reduce the security overhead, it uses the approach of elliptic curve cryptography (ECC).

In [2] author does not consider the deployment stage problem. It uses the RSU aided certificate revocation scheme in which RSU checks all passing vehicles for revoked certificates which are already stored at RSUs distributed by the CAs to the RSUs. If revoked certificate is found from any malicious vehicle it inform to all local vehicles about the revoked certificate by broadcasting it locally. This helps in reduction in the size of the CRL and high cost of the distribution of the CRL. It also considered revocation using tamper proof device (RTPD), DRP, RCCRL which reduce the size of the CRL. It follows group signature and identity based signature (GSIS) to preserve the privacy of the vehicle.

In [3] author only focuses on the CRL size and its distribution. It divides the CRL into various parts uses network coding and erasure coding to reassemble complete CRL with few pieces. Using erasure coding, a node will simply send out the same pieces it received without making any changes. Using network coding, a node will generate linear combinations of all of the pieces currently possessed, requiring greater processing capability at every OBU. Erasure coding has less overhead, both in packet overhead carry the coding information, and in processing overhead to reconstruct the file. It uses vehicle to vehicle (V2V) communicate to forward pieces between vehicles. The Most Pieces Broadcast method creates a situation where only the node with the most number of CRL file pieces is selected to broadcast within a given radio broadcast range.

In [4] author focuses on the issue of authentication and privacy. Here CA uses a pseudo random key generator (PRNG) to generate all the certificates of a single vehicle so that only CA can backtrack the detail of the source. CA generate all the certificates by selecting a random number "n" then generate all the certificates from it and send all the certificates to the corresponding vehicle and hold the random no to itself so that in case of need of detail of the sender only CA can back track the identity of the vehicle. It also uses the group certificate policy in which vehicles have a very large database of certificate up to 25000 each certificate is valid up to a very short time period to ensure a high level of privacy.

In [5] author uses a new approach to accelerate the certificate validation procedure by adding 2 new attributes credibility and issued date. Credibility is the measure of the authenticity of a particular vehicle. If a vehicle is having high credibility then it is a more trust full vehicle and the other one shows the date at which the particular certificate has been issued. With k-mean clustering it divide CRL into the k-cluster. Each cluster is divided based on these two new attributes. Whenever a request arrives, to check its validity it is compared with the certificate resides in its cluster only which in term reduces the overhead to search the entire CRL.

IV. SYSTEM MODEL

As shown in Fig. 1, the system model under consideration consists of the following:

A. Certificate Authority(CA)

It is responsible for providing anonymous certificates and distributing secret keys to all RAs and OBUs in the network. It is assumed that it cannot be compromised.

B. Region Authorities (RAs)

These are fixed units dedicated only one for a region. RAs are the middleware between CAs and RSUs. It is assumed that it cannot be compromised.

C. Roadside Units (RSUs)

RSUs are considered to be fixed and installed throughout the network. The RSUs can communicate securely with their RAs. It can be compromised because these lies near the road so attacker can easily reach to them.

D. On Board Unit (OBUs)

These are embedded in vehicles. OBUs can communicate either with other OBUs through V2V communications or with RSUs through vehicle to infrastructure (V2I) communications. These have the highest chances of being attacked.

According to the WAVE standard, each OBU is equipped with a TPD, which is a tamper-resistant module used to store the security materials, e.g., secret keys and certificates of the OBU. Also, the TPD in each OBU is responsible for performing all the cryptographic operations such as signing messages, verifying certificates, keys updating. We consider that legitimate OBUs cannot collude with the revoked OBUs as it is difficult for legitimate OBUs to extract their security materials from their TPDs. Finally, we consider that a compromised OBU is instantly detected by the RA.

V. PROPOSED SCHEME

In this approach the functionality of RAs and RSUs are very distinct as compare to the other scheme, we have seen so far. OBUs are loaded with certificates in its tamper proof devices (TPDs) by the certificate authorities (CA), which are valid for a long time. These are loaded with large numbers so that OBUs need to update only once in a year. Certificates are generated by the CA for each OBUs and only CA can recover the original identity of the OBUs.

Algorithm 1 *Certificate Generation Algorithm*

```
M = no. of certificates per time interval for vehicle.
I = no. of time intervals during a reload period.
Begin
n = get random number()
for i = 1 to I do
    $S_i = H^i (n)$
    for j = 1 to M do
        $(PK_{j,i}, SK_{j,I} )$ = generate public private key pair()
        $SIG_{CA,j,i} = SIGN(H\{E_{Si}(j), PK_{j,i}\}), SK_{CA})$
        $CERT_{j,i} = \{E_{Si}(j), PK_{j,I,} SIG_{CA,j,i}\}$
        UPLOAD($(CERT_{1,i,PK1,i,SK1,i})$..........,$(CERT_{M,i, PKM,i , SKM,i})$) at OBU
    end for
end for
end
```

Whenever the vehicle enters a new region, it needs to obtain the certificate for that region. With the help of road side units (RSUs) which are used only for communication purpose only in this approach because they are located near the road and highly vulnerable to attack. RSUs here are used only to cover the entire network. Once RSU get any request for the certificate by a vehicle. It forwards it to the RA of its region. RA has two types of certificate revocation lists (CRLs). Revocation list of RSUs as well as the revocation list of the OBUs. Revocation list of the OBUs at RA is provided by the CA while the RL of RSUs is generated by RA itself by detecting any malicious activity by any RSU. On getting any request by any OBU it checks it against all the entries of the CRL for the OBUs. If no entry is found then it generate a temporary certificate, valid for a short time period and in that region only. These certificates are sent to the corresponding vehicle

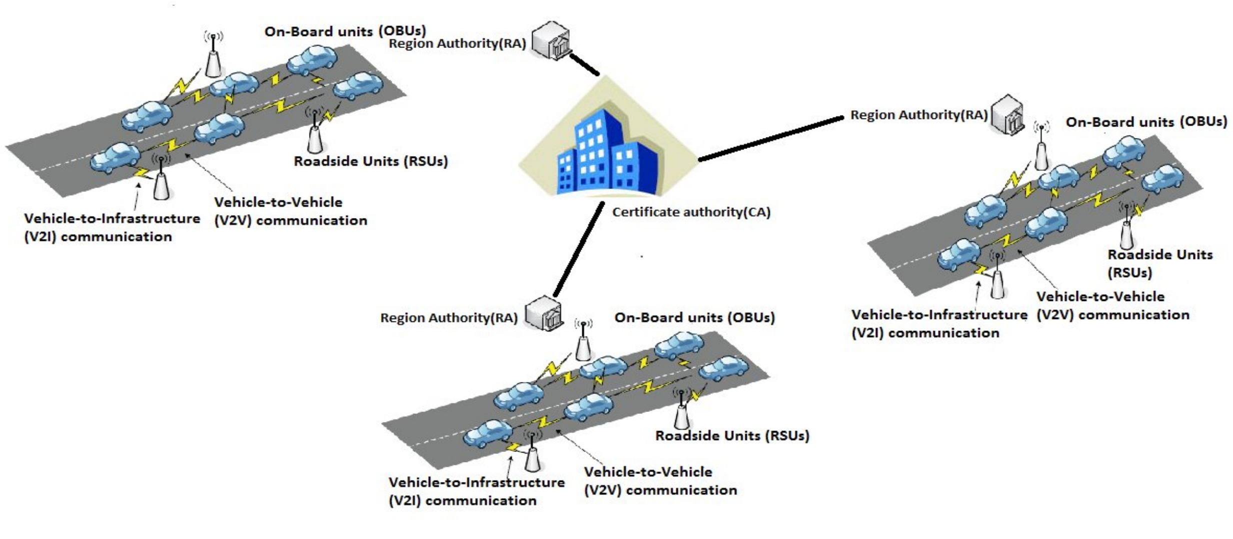

On-Board units (OBUs)
Region Authority(RA)
Roadside Units (RSUs)
Vehicle-to-Infrastructure (V2I) communication
Vehicle-to-Vehicle (V2V) communication
Certificate authority(CA)
Region Authority(RA)
On-Board units (OBUs)
Roadside Units (RSUs)
Vehicle-to-Infrastructure (V2I) communication
Vehicle-to-Vehicle (V2V) communication
Region Authority(RA)
On-Board units (OBUs)
Roadside Units (RSUs)
Vehicle-to-Infrastructure (V2I) communication
Vehicle-to-Vehicle (V2V) communication

Fig.1. System model

along with the CRL of OBUs as well as RSUs of that region only. CRL is very small in size to distribute and search. CRL is also updated time to time on any revocation of certificate by distributing only required pieces of information.

Algorithm 2 *Certificate updation algorithm*

```
AT OBU

N = no. of possible regions in path.
Begin
    for i = 1 to N do
        X = SIGN_SK (PK,CERT_CA)
        SEND(X,PK,CERT_CA) to RA
    end for
end

AT RSU
Begin
    verify(X,CERT_CA)
    (PK_V, SK_V,) = generate public private key pair()
    Y = SIGN_SKRA (PK_V, CERT_CA)
    CERT_RA  = (Y, Expiration, RA_id)
    add (X, CERT_RA) in history table of RA
    SEND (CERT_RA, PK_V, SK_V, RA_id, CRL_OBU, CRL_RSU)to OBU
End
```

A. Certificate Generation

To upload certificates initially in an OBU, it requires a number of certificates for a time interval as well as the number of time intervals during a reload period. Here n is a random number, the S_i is a key to some block of certificates. It is generated by hashing n by i times using some hash function 'H'. PK and SK are public and secret key pair. Signature is used to ensure integrity of the certificates by hashing the public key and encrypted value of 'j' using block identifier 'S' and then applying some signature algorithm with the secret key of CA. Certificates are generated by using sign, public key, and encrypted value. Now these are uploaded to the OBU.

B. Certificate Updation

N is the number of possible region a vehicle can enter. OBU send its request to each RA for their certificates by sending a sign of PK and certificate CERT issued by the CA using its SK. Now on receiving the request from an OBU each RA verifies its signature and then its certificate. If it is not found in CRL then generate new public-private key pair and sign OBU's public key and its CERT issued by a CA by its SK. now it generates a new regional certificate with sign, expiration period and its region authority id. It further adds the mapping detail in its history table and then sends the certificate of its region and ID of its region to the OBU.

C. Certificate authentication and verification

Each OBU need to authenticate to each other before the start of communication. Sender OBU initially broadcast its PK to all the other OBUs as well as RA. For authentication, sender OBU generate the signature by signing the message M by its SK

$Y = SIGN_{SK} (M)$

And then send ($Y,M,CERT_{RA}$) to the receiving OBU.

On receiving the above parameters receiving OBU needs to verify before to start the communication. First receiving OBU checks the validity of the sign Y, then check the $CERT_{RA}$ against all entries of the CRL. IF match found M is dropped else further communication will be established.

In TACK [6] RA delay the around 2 min to process the request of OBUs which is not considerable by VANET applications. In this paper RA collaborated approach is proposed to resolve such issue. Each RA is connected with their neighbor RAs to reduce the delay in the process. Before entering a new region OBUs can be easily facilitated by the functionality of early request of the certificate. Hence OBUs can have the certificate of the region even before entering it. Sometime it may also be possible that at any point vehicle can have a choice to enter into more than one region but generally it is restricted to either two or three regions possibility. In such case vehicle gets the early certificate of each region and with few choices it cannot be an accountable waste of resources.

It considers all the challenges and facts about the VANET. The Deployment stage problem is resolved by providing only one RA for each region. It also restricts the work of RSUs with covering the network only. It also provides the low cost deployment with most of the coverage. It also provides the authentication with RA certificate as well as group based key while preserving the privacy. With pseudonymous certificates and the hierarchy approached used here. Search and distribution of CRLs also cost very low because of very few entries in each CRL to deal with.

VI. EVALUATION CRITERIA

In this section we define a set of evaluation criteria which will help us in the comparison of the different schemes. Following is a list of the evaluation criteria used in the revocation schemes in VANETs.

A. Deployment Stage of VANET

The very first thing to consider is deployment stage because currently either VANET is an idea for most of the world or it is in the deployment stage. To fulfill the complete objective of secure VANET. Deployment stage must be considered because each protocol has their different performance in partial and full deployment of VANET.

B. Size of Certificate Revocation List and its Distribution

As the revocation of certificates takes place frequently. And the size of the VANET is very large with having millions of vehicles around the globe. The problem of distribution of the certificate revocation list (CRL) costs very huge amount of time and bandwidth. And with such huge size of CRL it also needs huge storage AT on board units (OBUs) as well as high processing speed to search that huge CRL.

C. Authentication vs. Privacy

Although each proposed scheme provides different methods for authentication between vehicles as well as providing privacy for the sender but there is a need to evaluate the performance of each protocol to ensure a high level of authentication while providing complete privacy in which only higher authorities can have the right to access the detail of the sender under any case of malicious activity.

VII. CONCLUSION

This paper proposes RAs based certificate organization and management scheme which resolves the basic challenges of VANET and provide a feasible solution of the deployment of VANET, The huge size of CRLs and its distribution and search, authentication and privacy. In comparisons with the available protocols, it can be deduced that the RA collaboration scheme will give better results than the existing ones. As the next step towards our research, we would like to augment this research with mathematical analysis and simulation result.

REFERENCES

[1] M Raya, P Papadimitratos and JP Hubaux , “Securing Vehicular Communications,” IEEE Wireless Communications, vol. 13, no. 5, pp. 8-15, October 2006.

[2] X. Lin, R. Lu, C. Zhang, H. Zhu, P. Ho and X. Shen, “Security in Vehicular Ad Hoc Networks,” IEEE Communications Magazine, vol. 46, no. 4, pp. 88-95, April 2008.

[3] Michael E. Nowatkowski and Henry L. Owen, “Certificate Revocation List Distribution in VANETs Using Most Pieces Broadcast,” proceedings of the IEEE SoutheastCon, pp. 238-241, 2010.

[4] Jason J. Haas, Yih-Chun Hu, and Kenneth P. Laberteaux, “Efficient Certificate Revocation List Organization and Distribution,” IEEE Journal on Selected Areas In Communications, vol. 29, no. 3, pp. 595-604, March 2011.

[5] Qingwei Zhang, Mohammed Almulla, Yonglin Ren and Azzedine Boukerche, “An Efficient Certificate Revocation Validation Scheme with k-Means Clustering for Vehicular Ad hoc Networks,” IEEE Symposium on Computers and Communications (ISCC), pp. 862-867, 2012.

[6] Ahren Studer, Elaine Shi, Fan Bai and Adrian Perrig, “TACKing Together Efficient Authentication, Revocation, and Privacy in VANETs,” 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad-Hoc Communications and Networks, pp. 1-9, 2009

[7] Albert Wasef and Xuemin Shen, “EMAP: Expedite Message Authentication Protocol for Vehicular Ad Hoc Networks,” IEEE Transactions on Mobile Computing, vol. 12, no. 1, pp. 78-89, January 2013.

[8] Vighnesh N V, N Kavita, Shalini R. Urs and Srinivas Sampalli, “A Novel Sender Authentication Scheme Based on Hash Chain for Vehicular Ad-Hoc Networks,” IEEE Symposium on Wireless Technology and Applications (ISWTA), pp. 96-101, September 2011.

[9] Hind AI Falasi and Ezedin Barka, “Revocation in VANETs: A Survey,” IEEE International Conference on Innovations in Information Technology (IIT), pp. 214-219, 2011.

[10] Albert Wasef, R. Lu, X. Lin and X. Shen, “Complementing Public Key Infrastructure to Secure Vehicular Ad Hoc Networks,” IEEE Wireless Communications, vol. 17, no. 5, pp. 22-28, October 2010.

[11] Ghassan Samara, Wafaa A.H. Al-Salihy and R. Sures, “Security Analysis of Vehicular Ad Hoc Networks (VANET),” Second International Conference on Network Applications Protocols and Services (NETAPPS), pp. 55-60, 2010.

[12] Nurain Izzati Shuhaimi and Tutun Juhana, “Security in Vehicular Ad-Hoc Network with Identity-Based Cryptography Approach: A Survey,” IEEE 7th International Conference on Telecommunication Systems, Services, and Applications (TSSA), pp. 276-279, 2012.

Performance Evaluation of 6LoWPAN Based Networks for Ubiquitous Health Monitoring System

Waiser Mehmood[1], Ammad Hassan[1], Rohan Tabish[1], Farid Touati[1], Adel Ben Mnaouer[2], and Brahim Gaabab[3]
[1]Dept. of Electrical Engineering, College of Engineering, Qatar University, Doha, Qatar
[2]School of Engineering,Canadian University of Dubai, Dubai, UAE
[3]MACS Labs., School of Engineering, University of Gabes, Gabes, Tunisia

Abstract—*Machine-to-Machine (M2M) networks have become very useful in a wide range of applications such as ubiquitous healthcare (u-healthcare) monitoring, air quality monitoring, insect monitoring, etc. Such applications require providing Internet connectivity to wireless sensor networks for the purpose of data collection and gathering. IPv6 over low power wireless personal area network (6LoWPAN) protocol specifications support the Internet-of-Things (IoTs), through an adaptation layer that provides efficient header compression. The 6LoWPAN middleware constitutes a crucial enabling technology for M2M networking. An optimum deployment of a 6LoWPAN-based system requires an accurate evaluation of the 6LoWPAN network's performance. Among various metrics, throughput and Packet Error Rate (PER) are of paramount importance, especially for time-critical and delay sensitive applications. In U-Health System where real-time ECG monitoring requires high throughput with reduced PER, an intolerable level of latency could result into lethal effects on patients. Thus, the performance of Wireless Sensor Network (WSN) used in U-Health should conform strictly to the timeliness requirement of health monitoring applications. In this study, we consider a system architecture where a 6LoWPAN node acts as a proximity body sensor to form a Wireless Body Area Network (WBAN) for a given patient. The objective of the study aims at assessing the scalability range of 6LoWPAN based WBANs and at estimating whether the performance bounds of such networks can serve efficiently u-healthcare applications. We evaluate the network performance by simulating different scenarios using Cooja (a Contiki-Network Simulator). Scenarios of a Constant bit rate (CBR) traffic and a Variable-Bit-Rate traffic are simulated, emulating real life biomedical data (e.g., ECG, temperature, etc.). The CBR-based scenario is conceived to determine the verge of the 6LoWPAN network in acute cases of U-Health system. Through simulation we demonstrate the viability of the performance bounds of the 6LoWPAN-based WBAN for u-healthcare applications.*

Keywords: 6LoWPAN, WBAN, Contiki OS, Performance evaluation, Scalability

1. Introduction

Wireless Sensor Networks (WSNs) have been increasingly deployed in the past few years and have earned significant importance in various application areas such as medical, environmental, agricultural, telecommunications, etc. Several applications require connectivity between the wireless sensor network and the Internet [1].

Sensor networks have been a fertile research area, during the recent years [2], for health monitoring systems. Examples of application include remote patient monitoring, wearable/portable health monitoring systems, patient data logging for analysis and diagnosis and so forth [3]. Sensor networks in general and wireless sensor networks (WSNs) in particular are constrained with respect to networking capabilities. Recently, the advent of the 6LoWPAN as an enabling technology allowed IPv6 packets to be carried on WSNs up to the level of the final sensing node, thus materializing the Internet of Things paradigm [4]. The 6LoWPAN feature is specified as an adaptation layer that does header compression allowing large IPv6 headers to shrink into smaller headers with sizes between 40 and 211 bytes [5].

Numerous U-Healthcare systems have been presented in the literature [6], [7], [8]. In these systems, researchers have developed either their own protocols on top of IEEE 802.15.4 MAC or have used Zigbee or other proprietary protocols. However, with all these protocols, the interoperability of their smart devices remains a challenge. In this context the 6LoWPAN specifications (an open standard proposed by IETF), have been branded as the solution for the efficient use of IPv6 packets over low-power, low-rate wireless networks, thus allowing network visibility for all involved embedded devices [9]. In addition to efficient header compression, key features of 6LoWPANinclude automatic network configuration using neighborhood discovery, unicast, multicast and broadcast support, fragmentation support, IP routing (using the Routing Protocol for Lossy channels (RPL)) and support for link layer mesh topology formation (e.g., IEEE 802.15.4) [10], [11].

In this paper, we carry out a scalability analysis based performance evaluation of a 6LoWPAN network intended for use with u-health monitoring applications. We are using common performance metrics such as data throughput, Packet Error Rate (PER), delay and deliverability ratio in order to characterize the 6LOWPAN network's behavior for a U-Healthcare monitoring system.

The remainder of this paper is as follows. Section 2 describes the simulation setup. Simulation results are presented in section 3, and their analysis are given in Section 4. Finally, section 5 presents concluding remarks.

2. Simulation setup

We present a simulation-based evaluation of a 6LoWPAN network using Cooja, a Contiki Network simulator. We used a built-in Zolertia mote model provided by the Contiki OS [12], an operating system dedicated to execute IoTs applications. Contiki OS provides a complete set of tools that are required to create 6LoWPAN-based networks further, in Contiki OS program can directly simulated or emulated on a device. We simulated 6LOWPAN using the Cooja simulator and the emulated built-in Zolertia motes. Cooja allows large and small networks of Contiki motes to be simulated. Motes can be emulated up to the hardware level.

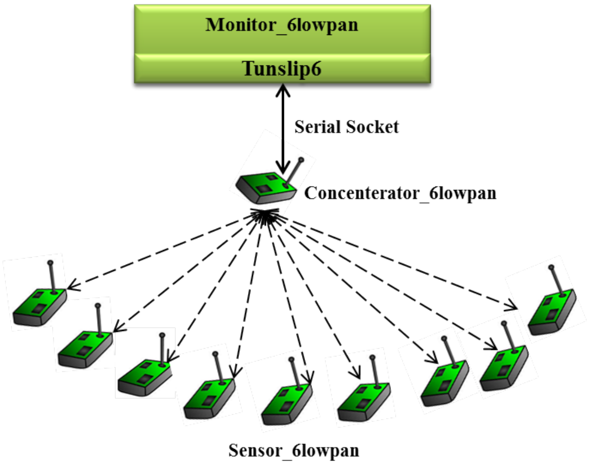


Fig. 1: Complete test bench with Cooja scenario for performance evaluation of 6LOWPAN network

To simulate the 6LoWPAN Network two scenarios are investigated; Constant Bit Rate Scenario (CBR) and Variable Bit Rate Scenario (VBR). In fact, in U-Healthcare system each sensor node has different rate of transmission rate. To emulate U-Healthcare system VBR scenario is simulated, however in some acute cases or in emergency cases transmission rate of the sensor can be changed depending upon the condition of the patient. In order to emulate that case we simulated CBR traffic. In CBR every sensor node is generating equal traffic at the same time and making equal contribution in throughput. If a certain level of the throughput can be achieved with reasonable PER in CBR then acute cases of U-Healthcare system can be managed.

The simulated network is sketched in figure 1. It is essentially made up of several sensor nodes (referred to as *sensor_6lowpan*), one edge router (referred to as *concenterator_6lowpan*), a Serial Line Internet application "tunslip6", and multithreaded application (referred to as *monitor_6lowpan*).

As sensors send packets to the concenterator_6lowpan, the monitor_6LoWPAN computes the average transmission delay between two contiguous received packets. Tunslip6 emulates a network interface on a PC. Tunslip6 acts as a bridge between the concentrator_6LoWPAN and monitor_6LoWPAN. Monitor_6lowpan performs computations according to the test discussed later in this paper. To obtain the results for analysis, number of simulations was performed. The simulation parameters are shown in table 1.

Table 1: General simulation parameters

Parameter name	Value
MAC layer	CSMA/CA
Radio duty cycling algorithm	Contiki MAC
Radio model	Undirected graph model
MAC layer queue size	8 packets
Bit rate	250 kbps
Node transmission range	50 Meter
Node carrier sensing range	100 meter
Simulated node type	Zolertia

3. Simulation results

During simulation, we measured the following quality of service (QoS) parameters:

- throughput against Packet Generation Interval (PGI) and offered traffic, and
- PER against the Packet Generation Interval

Throughput and PER were measured in both CBR and VBR. In each case, several network densities were considering, varying from 1 to 6.

CBR is used to determine the upper bound of throughput for the U-Healthcare monitoring system.

3.1 Constant bit rate scenario

Throughput vs Packet Generation Interval

The purpose of this simulation is to determine the effect of Packet Generation Interval on 6LoWPAN throughput, which affects the frequency of packet transmission of biomedical sensor. This simulation was conducted to measure the throughput as a function of the PGI for different network densities (Sensor_6Lowpan varying from 1 to 6) and with packet retransmission time 100ms.

We consider the topology shown in figure 1, where sensor_6lowpan transmitting directly to a concentera-tor_6lowpan. Various measurements were carried out in correspondence to different values of PGI. The repetition in simulation was made by incrementing the network density from 1 to 6. We analyzed the 6LoWPAN throughput as function of PGI up to six sensor nodes to accommodate six biomedical sensors (ECG, Accelerometer, Temperature, Heart Rate, Glucose and Blood Pressure) used in our U-Healthcare monitoring system.

figure 2 shows the performance of the throughput as a function of PGI. In figure 2, as we proceed from left to right, throughput increases linearly till particular value, for every incrementing sensor node, thereafter it starts decreasing exponentially due to decrease in rate of offered traffic, hence increasing PGI. We observed that offered traffic cannot reach the maximum data rate of the physical layer due to hardware and channel limitation.

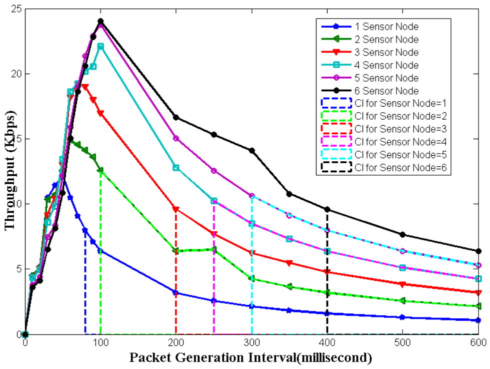


Fig. 2: Throughput measurement results for 6LOWPAN network as a function of Packet Generation Time

Packet Error Rate (PER) /Packet Delivery Ratio (PDR) vs Packet Generation Interval (PGI)

The communication system can be characterized in terms of connectivity or homogenously by the PER/PDR, as the connectivity of the link depends upon the routing protocol and on the packet retransmission mechanism under CSMA/CA [9]. Contiki uses RPL for routing and CSMA-CA for retransmission mechanism.

Figure 3 shows PER with respect to PGI, at different network densities of the 6LoWPAN network. Initially when the PGI is low PER is very high even in the case of one sensor node. As PGI increases PER starts decreasing up to an instant at which PER approaches to zero, from that instant onwards interval is called Confidence Interval (CI).

PDR/PER was measured using the simulation setup shown in figure 1. In figure 3, simulation results for PER/PDR shows that for lower value of PGI, the packet loss is high, even for one Sensor_6lowpan, packet lost is present up to 80ms PGI. Table II shows the confidence intervals of PGI for each network density, with 100% PDR. These confidence intervals are estimated from the throughput and from PER shown in figure 2 and figure 3 respectively. Confidence interval indicates the CBR's maximum rate of the packet transmission, at which PER becomes zero. Performance parameter of 6LoWPAN i.e., for one Sensor_6lowpan PER ≈ 0 for more than and equal to 80ms PGI.

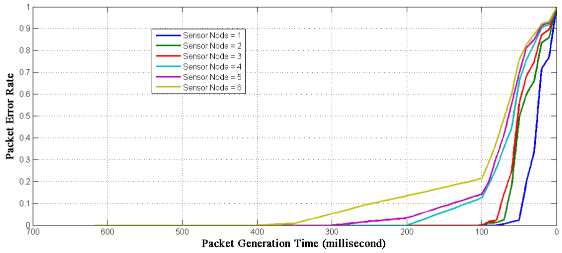


Fig. 3: Packet Error Rate

Table 2: CBR upper limit PER vs Confidence Interval (CI)

Network density	PER/PDR	CI
1	≈ 0 / ≈1	>=80ms , @80ms ~8Kbps
2	≈ 0 / ≈1	>=100ms, @100ms ~12.56Kbps
3	≈ 0 / ≈1	>=200ms, @200ms~9.6Kbps
4	≈ 0 / ≈1	>=300ms, @300ms~8.5Kbps
5	≈ 0 / ≈1	>=400ms, @400ms~8Kbps
6	≈ 0 / ≈1	>=600ms, @600ms~6.5Kbps

Throughput VS Offered Traffic

Among the commonly used sensors in U-Healthcare systems, the highest bandwidth requirement comes from ECG sensor. Figure 2 and table 2 shows that the maximum throughput with high PDR is around 100ms PGI. In acute cases of the U-Healthcare, if every sensor node transmits around 100ms PGI then there will be less chances of collision and will have high throughput, To attain maximum throughput of the system, we simulate the effect of increasing nodes on throughput at Packet generation Interval (PGI) = 100ms, Packet Retransmission Time = 100ms and Packet Length = 80bytes. The intention of this simulation is to measure the maximum throughput that can be offered by the 6LOWPAN. We measured the throughput by increasing the number of nodes, but our simulation shows that a practical network performance is still far from theoretical performance level as shown in figure 4.

In fact, only a throughput of 31 Kbps can be achieved in the presence of the maximum offered traffic load. We also noticed the throughput reached at its peak value for around 23 nodes. Thereafter, the throughput starts deteriorating as

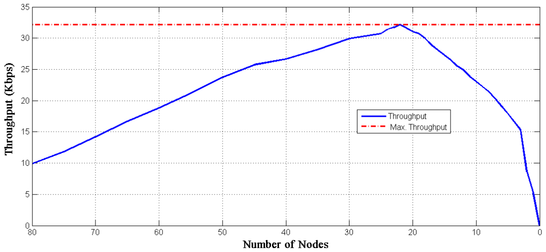


Fig. 4: 6LoWPAN Throughput as a function of Nodes

the number of nodes increases. This is due to high queuing and packet drops that starts occurring at high loads where the system has reached instability.

3.2 Variable bit rate scenario

CBR with low PGI is the worst case, where all nodes transmit with same transmission rate; which increase the collision probability. Whereas as described in table 3, VBR has less collision with respect to CBR because most of the sensors are in sleep mode except ECG node. To analyze the 6LOWPAN feasibility for U-Healthcare monitoring system, we emulate biomedical sensor in the Cooja with following parameters (i) rate of transmission frequency (ii) size of the data, the actual VBR model is described in table 3.

Table 3: Biomedical Sensor Data Statistics

Biomedical sensor type	Data to transmit	Transmission frequency (PGI)
ECG	Max ECG Frequency = 200Hz 1 second data with sampling frequency of 600Hz	15 packets PGI = 80ms
Temperature	2 bytes	1 packet PGI = 120s
Accelerometer	6 bytes	1 packet PGI = 120s
Respiration rate	1 bytes	1 packet PGI = 60s.
Glucose	4 bytes	1 packet PGI = 2s

VBR is considered a case where different proximity 6LoWPAN body sensors are connected to the patient and sending data to the system. As in VBR every sensor has particular PGI and sending at its own particular rate, the probability of collision of the packets is very low and it requires low throughput. The maximum throughput required by the system with these biomedical sensors (ECG, temperature, Accelerometer, Respiration rate) is: ~8Kbps -10Kbps (600Sample/second, each sample of 2 bytes with overhead of processing time of 200ms). In U-Healthcare system only sensor which is frequently using the channel is ECG sensor and using the maximum payload and other biomedical sensor can be integrated to one sensor node to efficiently use the payload and packet collision can be avoided. The WBAN can be emulated by only two sensor nodes (One sensor node is required for ECG sensor and the other sensor can be used to emulate the aggregated traffic of all the other biomedical sensors).

Table 4: Statistical Results: CBR and VBR

Network setup	Traffic description	Results
2 sensors	VBR $\text{PGI}_{\text{node1}}$ = 80ms ; $\text{PGI}_{\text{node2}}$ = 60second	Calculated Throughput = 8.64Kbps Measured Throughput ≈ 8.54Kbps PER = 0, PDR = 100%

	CBR
	Constant PGI = 80ms
	80 bytes payload for both sensors
		Throughput = 13.916Kbps
		PER = 0.1243, PDR = 87.57%

6 sensors	VBR $\text{PGI}_{\text{node1}}$ = 80ms, $\text{PGI}_{\text{node 2,3,4}}$ = 120 s, $\text{PGI}_{\text{node5,6}}$ = 60 s	Calculated Throughput = 10.24Kbps Measured Throughput ~ 9.4Kbps PER = 0.023, PDR = 98.4%

	CBR
	PGI = 80ms, 80 bytes payload for all sensors
		Throughput = 20.573Kbps
		PER = 0.3848, PDR = 61.52

4. Simulation results analysis

Table 4 shows the simulation results of the CBR and VBR of the 6LoWPAN body sensor network. For the simulation of two sensor nodes, if we compare throughput and PER of CBR and VBR, the throughput of the CBR is high as compare to VBR, which has zero PER. The difference in throughput is due to difference in offered traffic by these two systems. As we increase the number of sensor nodes up to six throughput of both system is increased but with some values of PER. The reason for such difference in PER is, in CBR each sensor node has same PGI and probability of packet collision is very high whereas in VBR each sensor has its own particular PGI and chances of collision is very low but it comes with a low value of throughput. The throughput achieved by VBR is reasonable for the U-Healthcare system with achievable packet delivery ratio is 98.4%.

5. Conclusions

In this paper we conducted a scalability based performance evaluation of the 6LoWPAN. In U-Healthcare system 6LoWPAN node acts as a proximity body sensor and these sensors form WBANs. In order to estimate throughput and PER of the WBAN we consider two traffic scenarios CBR and VBR. CBR scenario is used to simulate the acute cases of the U-Healthcare system and it exploits to estimate the max throughput limits of the 6LoWPAN network for VBR as shown in Table II; which is actually emulated as WBAN for U-Healthcare monitoring system. We found that 6LoWPAN network limits the node's transmission capability; primarily it limits the throughput and requires more packet generation interval for less PER. We record the Confidence Interval of the network for different network densities and it is different for every density. CI defines the node's transmission capability with zero PER, which helps us to design Wireless Body Area Network for U-Healthcare monitoring system. Simulated result indicates that the 6LoWPAN based system has reasonable throughput and PER for U-Healthcare system requirement for WBAN and has a big potential for U-Healthcare monitoring system.

Acknowledgment

This publication was made possible by the NPRP grant #[4-1207-2-474/1/2012] from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1] T. Yashiro, S. Kobayashi, N. Koshizuka, and K. Sakamura, "An internet of things (iot) architecture for embedded appliances," in *Humanitarian Technology Conference (R10-HTC), 2013 IEEE Region 10*, Aug 2013, pp. 314–319.

[2] F. Touati and R. Tabish, "U-healthcare system: State-of-the-art review and challenges," *Journal of Medical Systems*, vol. 37, no. 3, pp. 1–20, 2013. [Online]. Available: http://dx.doi.org/10.1007/s10916-013-9949-0

[3] B. Xu, L. Xu, H. Cai, C. Xie, J. Hu, and F. Bu, "Ubiquitous data accessing method in iot-based information system for emergency medical services," *Industrial Informatics, IEEE Transactions on*, vol. 10, no. 2, pp. 1578–1586, May 2014.

[4] R. Silva, J. S. Silva, and F. Boavida, "Evaluating 6lowpan implementations in wsns," *Proceedings of 9th Conferncia sobre Redes de Computadores Oeiras, Portugal*, pp. 1–5, 2009.

[5] I. K. Samaras, G. D. Hassapis, and J. V. Gialelis, "A modified dpws protocol stack for 6lowpan-based wireless sensor networks," *Industrial Informatics, IEEE Transactions on*, vol. 9, no. 1, pp. 209–217, 2013.

[6] C. Otto, A. Milenkovic, C. Sanders, and E. Jovanov, "System architecture of a wireless body area sensor network for ubiquitous health monitoring," *Journal of Mobile Multimedia*, vol. 1, no. 4, pp. 307–326, 2006.

[7] M. Rosu and S. Pasca, "A wban-ecg approach for real-time long-term monitoring," in *Advanced Topics in Electrical Engineering (ATEE), 2013 8th International Symposium on*. IEEE, 2013, pp. 1–6.

[8] K. Ernest, C. Lamei, S. Mohamed, M. Shakshuk, I. Badreldin, and I. ElBabli, "A zigbee-based telecardiology system for remote healthcare service delivery," in *Biomedical Engineering (MECBME), 2011 1st Middle East Conference on*, Feb 2011, pp. 442–445.

[9] F. Touati, R. Tabish, and A. Ben Mnaouer, "Towards u-health: An indoor 6lowpan based platform for real-time healthcare monitoring," in *Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP*. IEEE, 2013, pp. 1–4.

[10] D. Chen, J. Brown, and J. Khan, "6lowpan based neighborhood area network for a smart grid communication infrastructure," in *Ubiquitous and Future Networks (ICUFN), 2013 Fifth International Conference on*, July 2013, pp. 576–581.

[11] X. Wang, S. Zhong, and R. Zhou, "A mobility support scheme for 6lowpan," *Computer Communications*, vol. 35, no. 3, pp. 392 – 404, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366411003537

[12] "Contiki: The open source operating system for the internet of things." Online. [Online]. Available: http://www.contiki-os.org/index.html

An approach for analysis the impact of LTE-based femtocell network - Case study based on discrete simulation

K.C. Silva[1] , C.P.A, Silva[1], A.C.S, Donza[1], C.R.L, Frances[1], N.L Vijaykumar[2]

[1]Department of Electrical Engineering and Computer Science, Federal University of Pará, Belém, PA, Brazil

[2]Laboratory of Computing and Applied Mathematics, National Institute for Space Research, São José dos Campos, SP, Brazil

Abstract - *As in any network, LTE (Long Term Evolution) also has a concern with respect to coverage and data rate. A key element in LTE is the deployment of multiple femtocells in order to improve both coverage and data rate. Naturally, one has to remember that handover mechanism may become complex due to arbitrary overlapping of coverage and this complexity becomes a challenging issue to deal with. Therefore, the paper described here simulated deployment of femtocells in a scenario to obtain QoS (Quality of Service) and handover. The paper also discusses limitations of integrating LTE with femtocells. This alternative of employing femtocells is quite interesting but several uncertainties arise in this topic. The results obtained from simulations showed handover getting worse and thus impacting QoS.*

Keywords: LTE, handover, femtocell, simulation, OPNET

1 Introduction

Currently, Internet and mobile communications are converging to a new paradigm, the Mobile Internet. The ability to access information and services anytime and from anywhere has been shaping not only new user profiles but also demands for new applications.

With the popularization of the third and fourth generation technologies, mobile communication systems suffer from the addition of new services and functionalities, which also involve critical problems, such as interference, limited coverage, restrictions on the use of triple play applications, among others.

The multimedia services are becoming increasingly popular. These services generate intense traffic on the network, that demand for higher data rates and are sensitive to delay and delay variation, experienced in the network.

In indoor environments, channel quality between the cellular base station and the mobile node may be affected by walls and obstacles. The wireless communication for indoor requires more resources, including time, bandwidth, transmission power so that they can ensure the quality of service required by customers. However, the lack of resources in wireless cellular networks will be accelerated, since over 60% of the voice traffic and 90% of data traffic is generated indoors [1].

Thus, it is necessary to investigate access technologies to ensure satisfactory levels of quality of service, taking into consideration the growing demand for data services. In this respect, for a wireless operator, femtocells are an attractive alternative since they are cost-effective to significantly increase the user data rates of their wireless networks at the customer premises.[2]

This growing demand for indoor wireless multimedia and ongoing trends of mobile convergence are paving the way for the installing femtocells industries. Femtocells may be open access or closed access. [3] Open access allows an arbitrary user to use the femtocell, whereas in closed access, the use is restricted to users that are explicitly approved by the owner. While the ultimate goal of femtocell is to improve the efficiency, coverage and services at a reduced cost of operation, the possibility of arbitrary handovers between the existing eNodeB (enhanced NodeB) and HeNB (Home eNodeB), poses new challenges [4].

LTE (Long Term Evolution) is a technology for wireless broadband 4G (4th Generation) mobile networks for voice and data that ensures greater data speeds, better performance and more efficient use of spectrum. In this context, LTE networks have gained much attention, mainly because this technology can be used to improve voice services where there is a limited local coverage [5].

However, even supporting high data rate, LTE frequency can result in poor indoor coverage in some areas. In this case, an LTE femtocell network could be an alternative to improve the indoor signal and avoid overloading the macrocell. It is still too early to predict the success of LTE femtocell. Factors such as safety, interference and management still need to be studied.

The integration of femtocell LTE network appears to be a promising approach due to its homogeneous nature and characteristics. However, for the mobile unit to provide transparent communication in a change of radio resources, it is necessary to handover the network capacity. The handover is the process that characterizes a cellular network and assures its mobility feature.

From a technology point of view, a femtocell is not only characterized by short communication range and high throughput, but also by its ability to seamlessly interact with the traditional cellular network at all layers of the network stack, performing tasks like handovers, interference management, billing and authentication. This necessitates substantial support by the appropriate standards bodies [6].

Therefore, this paper aims to conduct a study on handover from the use of femtocells and evaluate the correlation between the indicators of handover and QoS. The study also includes the effectiveness of using LTE femtocell and especially on what will be the impacts, earnings and disadvantages that this combination of technologies offer.

The following sections are organized as follows: section 2 will discuss some concepts of technologies on which this work is based. In section 3 related work will be presented. The description of the methodology used and the results are discussed in section 4. Finally, in Section 5 are final remarks.

2 Theoretical Fundaments

2.1 LTE (Long Term Evolution)

Searching for solutions to make data transmission more efficient, while dealing with more and more volumes of such traffic, LTE has been proposed as the next step in the 4G mobile system, preceded by 2G and 3G . Its development is intended to provide performance improvements, while reducing the cost per bit, allowing for a greater dissemination of mobile services. Its standardization is the responsibility of 3GPP [7].

LTE networks have a new architecture, totally different from what had been used in previous technologies. An example of this is the base station, called eNodeB in which LTE starts processing tasks previously performed in RNC (Radio Network Controller).

Note that the eNodeB will also be responsible for handover decisions through communication between the elements using X2 interface However it is possible that due to the lack of communication over X2 (optional interface), communication between the eNodeB will be accomplished through other alternative, the Access Gateway. [8] Figure 1 illustrates the components of network architecture integrating LTE and femtocell.

Over the next years, it is expected that billions of devices will be connected to the Internet and cloud-based applications using mobile wireless 3G and LTE networks. So a huge demand for wireless mobility and ubiquitous coverage will definitely be necessary. Global mobile data traffic will increase 26 times between 2010 and 2015, also known as "mobile data tsunami" [9].

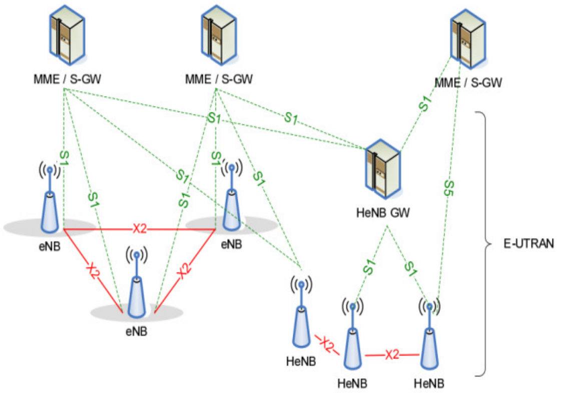


Figure 1. Overall E-UTRAN Architecture with deployed HeNB GW[10].

2.2 OPNET

The OPNET Modeler accelerates the process of research and development enabling the analysis and design of communication networks, devices, protocols and applications. It is widely used as a simulator Instrument for modeling telecommunications networks [11].

It allows one to create a network from a library of templates and define parameters not only for the environment, but also of each object that makes up, and the impacts of its variations. For educational purposes, its use has also a leverage as one of its major advantages is the graphical interface provided to the user to configure settings and to view results.

2.3 Handover

The handover is a difficult procedure because it involves several tasks that may cause interruptions in service delivery and performance degradation of applications. This fact becomes worse if there is an increase in the frequency of migration and transition. As a result, there is a greater number of handovers.

Recently, the concept of handover has not only been linked to continuity of a phone call, but also to the continuity of streaming sessions, maintaining QoS and access to the Internet. One of the research challenges for cellular systems is to improve the call admission system that controls and reduces blocking probability and improves the quality of service.

This extension of the concept of handover occurs due to the popularity of tablets and smartphones, which have allowed the collective experience of users sharing the same coverage area. Recently, the scenario of mobility at different speeds with applications in use has been increasingly common.

As the EU moves in the network, it may experience different propagation conditions and interference. Can happen to a neighboring cell presents the best conditions (RSRP higher) than the current cell. Therefore, the UE monitors the current cell (Senb - serving eNodeB) and neighboring cells (NeNBs - neighboring eNodeBs) performing periodic measurements of downlink radio channels (RSRP). The HO is triggered by the eNodeB based on measurement reports (reports measure) received from the EU. [12]

Control messages are exchanged across the interface between the two X2 eNodeBs and downlink data packets are also forwarded from the source to the target eNodeB through the same X2 interface. [13],[14].

2.4 Femtocell

The femtocell concept is part of the effort of telecommunication industry to provide communication of high performance, high-quality services for home users. In contrast to conventional types of cells, which are well planned by the operators, the femtocell base stations must be installed by customers themselves, similar to a wireless access point. [15]

The Femtocells are small base stations with the same functionality as the macrocells, but they have power to meet only a restricted environment (10-20 meters). They are of low cost, supporting a small number of users and installed by the user to best receive voice and data in closed environments. [16]

It is estimated that 2/3 of the calls and over 90% of data traffic in a cellular network, occur in indoor environment. Some research shows that 45% of households and 45% of companies have a bad experience regarding indoor coverage [17]. To provide good indoor coverage for customers has become a major challenge for operators, because it is not anymore just offering a good voice service, but also high data transfers including video streams.

A factor to be taken into consideration is that the process of installing these femtocells would be up to the user. It must be very simple such as plug and play. These cells must have ability for self-configuration and must be built to minimize impact on the macrocell through self-provisioning parameters.

In this paper, simulations of deploying femtocells in an indoor environment have been carried out to study its effects on the handover and evaluation of quality of service experienced by users.

3 Related Works

Some of the literature analyzed was intended to establish the best way of balancing the factors involved in mobile communication and manage users make better use of network resources and thus get a higher efficiency. Mechanisms for handover in LTE networks have been intensively studied in both academia and industry.

In [18] present a detailed literature review with the main features of femtocell technology and raising technical and regulatory issues. Such networks face a lot of uncertainties as the infrastructure is not preplanned. Moreover, there are technological issues to be considered: Can Femtocells handle unloading data and video streams from conventional networks? Will they create more problems and thus jeopardizing the careful work on installing base stations considering unpredicted interferences?

[19] proposed a strategy for handover between macrocell and femtocell for LTE networks. The paper presents a strategy that tries to avoid failures of handover and the occurrence of unnecessary handovers. [20], [21] analyze challenges with respect to their potential for use in LTE femtocell networks as an alternative for coverage.

Into [22] the handover procedure in LTE femtocells is discussed focusing on the significant increase in the number of femtocells in certain environments. Simulations for the handover between macro and femto and between femto and femto were performed and an optimization algorithm was proposed and compared to conventional algorithm.

In other studies, the process of handover between HeNB and LTE eNodeB in a modified [23] version has been proposed. A new handover algorithm based on the speed of the EU and on QoS. Three different speed settings were considered in the algorithm: low mobility (0-15 km/h), medium (15-30 km/h) and high mobility (>30 km/h). The analysis showed that the proposed algorithm has the best performance, and then the algorithm is compared with the traditional algorithm.

The following articles show some concerns on the principal aspects for limitation on employing femtocells: [24], [25], [26] concerns are in order in spite of all the advantages. Unfortunately, deadlocks are not easily sorted out due to several technical and non-technical issues that are still pending to be solved.

4 Methodology

First of all, it was necessary to elaborate the methodology to consider same modeling for both the scenarios to be simulated. Figure 2 shows a flowchart that describes the sequence of activities to be conducted. This methodology may be generalized for several other real world problems that can be modeled and simulated.

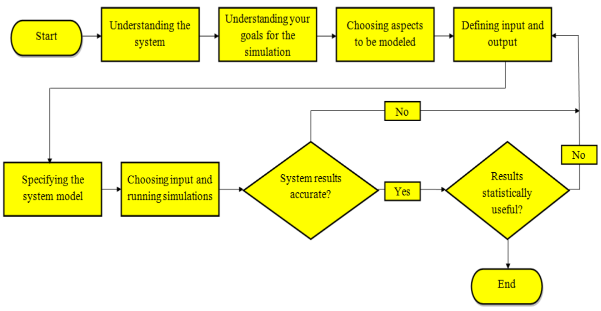

Start
Understanding the system
Understanding your goals for the simulation
Choosing aspects to be modeled
Defining input and output
No
No
Specifying the system model
Choosing input and running simulations
System results accurate?
Yes
Results statistically useful?
End

Figure 2. Flowchart

In this paper, the modeling of scenarios of interest was performed using the Opnet Modeler 17.5 (release 8). Figure 3 illustrates the modeled scenario that will be simulated

In the analysis conducted for this study, two identical scenarios were created: one without the use of femtocell and the other with 9 femtocells per cell. The configuration and parameterization of antennas femtocell are shown in Table III.

In both the scenarios, the network consists of 5 eNodeB and with respect to mobility, random waypoint mobility model [27] was assumed for all 100 network users. The structure also includes EPC (Evolved Packet Core) and gateway elements that will communicate with the application server.

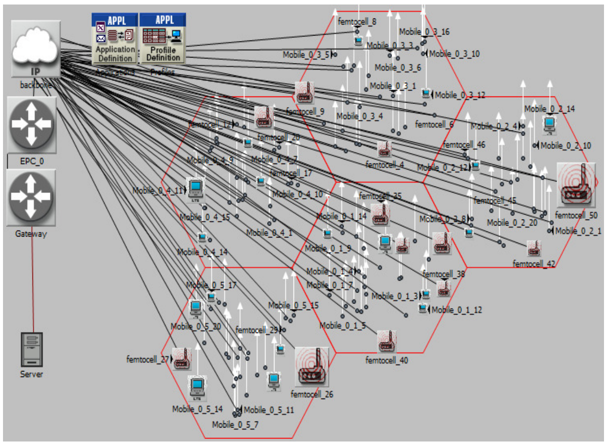


Figure 3. Scenario modeled in OPNET tool.

4.1 The Simulation Parameters

The parameters and settings of the antennas are described in Table II. To generate traffic on the network, VoIP (Voice over Internet Protocol) application was used.

VoIP application represents the class of inelastic applications, real-time, interactive, which is sensitive to delay end-to-end, but can tolerate packet loss. Today, the emergence of real-time application requires more resources, so it is necessary ensure rapid and reliable voice communication for a large number of users on the network.

All users have been configured to establish a VoIP call to an external server. In Table I, we list the most relevant parameters used for configuring the application.

TABLE I. CONFIGURATION OF VOIP APPLICATION

Silence Length (s)	Exponentially distributed, mean 0.65
Talk Spurt Length (s)	Exponentially distributed, mean 0.352
Encoder Scheme	GSM FR
Voice Frames per Packet	1
Type of Service	Best effort (0)
De-Compression Delay (s)	0.02

TABLE II. CONFIGURATION OF LTE ANTENNAS

Parameter	Value
Transmission Power	26 dBm
SC-FDMA (UL) Frequency	1920 MHz
OFDMA (DL) Frequency	2110 MHz
Bandwidth	10 MHz
Gain Antenna	17 dBi
Antenna Height	40m
Radius Coverage	7 Km
Propagation Model	Urban Macrocell
Duration of simulation	900s

TABLE III. CONFIGURATION OF FEMTOCELLS

Parameter	Value
Transmission Power	23 dBm
Gain Antenna	2 dBi
Propagation Model	Indoor Environment
Antenna Height	1m

4.2 Handover Performance Indicator (HPI)

In this section the main evaluation metrics used as performance indicators handover will be described. The metrics are described below.

Handover Delay: This enables to identify the number of handovers performed, as well as the position in time where the delay for the handover has successfully occurred.

Handover Failure: The ratio of handover failure (HF) is the ratio between the number of failed handovers (NHfail) on the number of attempts made. The number of handover attempts is given by the sum of the number of failed handovers (NHfail) plus the number of successful handovers (NHsuc).

$$HPI_{HF} = \frac{NHfail}{(NHfail + NHsuc)} \qquad (1)$$

Blocking Probability: The blocking probability is the ratio of the number of failures (Nfail) over the total number of failures (failure of handover + failures radio) added to the total number of handovers (TH).

$$HPI_{BP} = \frac{Nfail}{((RadioFail - NHfail) + TH)} \qquad (2)$$

4.3 Simulation Results

After the initial configuration in OPNET, simulation was conducted. Several instances were run and simulations of scenarios with the same configuration were repeated. Time to elapse was the same and with same parameters but with different seeds. The simulation time was 900 seconds, long enough for the environment to be stabilized and users that were testing could navigate the established trajectory.

Figure 4 shows the chart extracted directly from OPNET. This chart expresses a joint display of two key metrics for analyzing the handover performance indicators. Association of eNodeB allows viewing the instant in which the user has joined and in which eNodeBs he or she has joined.

In the same figure and at the same instant, the handover delay is reported, identifying delays of handovers that were performed successfully.

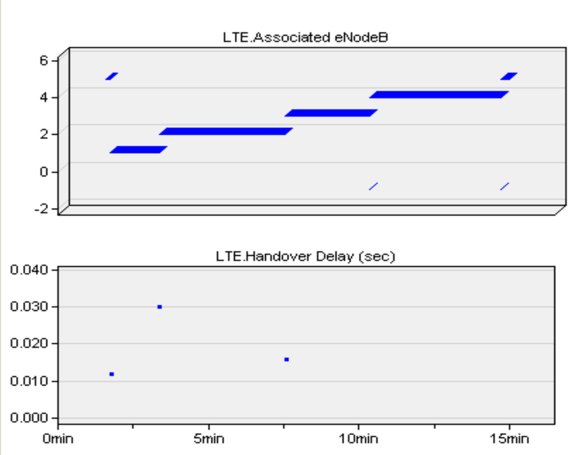


Figure 4. Associated eNodeb e Handover delay.

The results are an indicative of the mechanisms that actually impact on system performance. Some of the variable parameters include the speed of the user, the type of traffic, application, etc.

In the analysis, 100 mobile users were initially observed individually, since each user traveled a random trajectory, which guaranteed number of failures and handovers specific to each one.

The results presented here compare the two simulated environments, i.e. with and without femtocell. The analysis allows us to infer that with the deployment of femtocells in the network, users are conducting a much larger number of unnecessary handovers, which impacts heavily on indicators of handover performance. The comparison can be seen in Table IV.

TABLE IV. Indicators of Handover Performance

Statistic	**LTE scenario**	**LTE scenario + *Femtocell***
Handover Failure	0.1	0.71
Blocking Probability	0.259	0.764

Some indicators were collected to evaluate QoS experienced by the users behavior. We notice that the deployment of femtocells did not represent a significant improvement over the parameters of QoS. An assessment of the general behavior of 100 users was carried out. The figure below shows the delay in both scenarios. We notice that there was no great variation; both had behavior around 200 milliseconds.

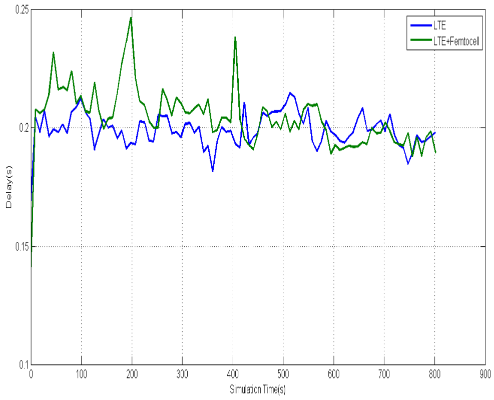


Figure 5. Delay of VoIP application.

Figures 6 and 7 show the values of MOS (Mean Opinion Score) and jitter obtained. The MOS is the mean of results from users that tested the scenarios. Using a scale from 1 to 5, where a score of average equal to or greater than 4 is considered toll-quality. The MOS achieved was considered poor, as average behavior was around 1.5 even after the addition of femtocells.

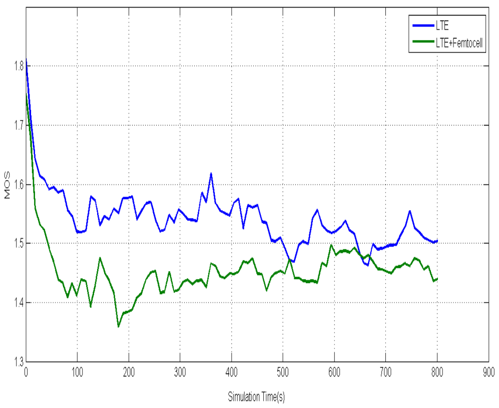


Figure 6. MOS of the VoIP application.

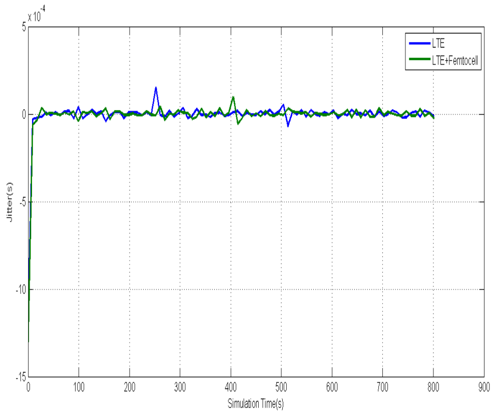


Figure 7. Jitter of VoIP application.

5 Conclusion

Based on the preliminary analyses, one can note that integrating LTE and femtocells was not a good option as expected. For the considered parameters and scenarios, inclusion of femtocells would improve QoS. The results stress the necessity of self-configuration for proper functioning of femtocells. It is also important to mitigate the degradation of the performance due to the interference between macrocell and femtocell as well as among femtocells, especially when installations are conducted without proper planning. Without these issues, it is impracticable and there is a significant impact on QoS, on handover and on the overhead of signaling associated to mobility procedures.

In spite of some of the aspects mentioned, use of these small cells, at least for the considered scenarios, did not turn into a panacea. Management of handover mechanism, interference and self-configuration still poses a major challenge and it is relevant for the success of integration of LTE and femtocells. Finally, it is important to point out that the study conducted in the paper should not be considered as conclusive and other parameters must be taken into consideration. Besides, there must be a forum to discuss the employing of femtocells.

6 References

[1] G Mansfield, in Proceedings of the FemtoCells Europe Conference. Femtocells in the US market-business drivers and consumer propositions, (London, UK, p. 2008).

[2] D Calin, H Claussen, H Uzunalioglu, On femto deployment architectures and macrocell offloading benefits in joint macro-femto deployments. IEEE Commun. Mag. 48(1), 26–32 (2010).

[3] P. Xia, V. Chandrasekhar, and J. G. Andrews, “Open vs. closed access femtocells in the uplink,” IEEE Trans. Wireless Commun., vol. 9, no. 12, pp.3798–3809, Dec. 2010.

[4] A. Roy, J. Shin, and N. Saxena, Multi-objective handover in LTE macro/femto-cell networks. In Proceedings of Journal of Communications and Networks. 2012, 578-587.

[5] Airvana, "Femtocell Network Architecture", whitepaper, May 2010.

[6] J.G. Andrews, H. Claussen, M. Dohler, S. Rangan, M. C. Reed, "Femtocells: Past, Present, and Future" IEEE JSAC on Femtocellular Networks, 30(3):497--508 April 2012.

[7] Danish Aziz, Rolf Sigle, “Improvement of LTE Handover Performance through Interference Coordination”, IEEE 69th Vehicular Technology Conference, 2009.

[8] Neissi Shooshtari, Ali; Optimizing handover performance in LTE networks containing relays ; School of Electrical Engineering, Department of Communications and Networking, Master’s theses, 2011.

[9] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update, 2010-2015,” Whitepaper, Feb. 2013.

[10] 3GPP TS 36.300, “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRAN); Overall description; Stage 2, September 2010.

[11] Opnet Modeler. [Online]. Available: http://http://www.opnet.com/

[12] 3GPP TS 36.214 V8.2.0 (2008-03), Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer-Measurements (Release 8).

[13] 3GPP TS 25.467 V9.2.0, “UTRAN architecture for 3G Home Node B (HNB); Stage 2; March 2010.

[14] 3GPP TR 23.839 V0.1.1, “Study on Support of BBF Access Interworking”, May 2010.

[15] Ying Li, Andreas Maeder, Linghang Fan, Anshuman Nigam, and Joey Chou, “Overview of femtocell support in advanced WiMAX systems,” IEEE Communications Magazine, July 2011.

[16] Vikram Chandrasekhar and Jeffrey G. Andrews. Femtocell Networks: A Survey. IEEE Communication Magazine, Vol. 46, Issue 11, 09/08.

[17] J. Zhang, G. D. L. Roche, and G. L. D. Roche, Femtocells: Technologies and Deployment. John Wiley and Sons, Ltd., 2010.

[18] Andrews J.G. et al. Femtocells: Past, Present, and Future. IEEE Journal, Selected Areas in Communications. 2012.

[19] S. J. Wu, "A new handover strategy between femtocell and macrocell for LTE-based network,” presented at the Fourth International Conference on Ubi-Media Computing, IEEE 2011.

[20] Dionysis Xenakis, Nikos I. Passas, Lazaros F. Merakos, Christos V. Verikoukis: Mobility Management for Femtocells in LTE-Advanced: Key Aspects and Survey of Handover Decision Algorithms. IEEE Communications Surveys and Tutorials 16(1): 64-91 (2014).

[21] H. Zhou, D. Hu, S. Mao, P. Agrawal, and S. A. Reddy, "Cell association and handover management in femtocell networks," in Proc. IEEE WCNC 2013, Shanghai, China, Apr. 2013, pp. 1-6.

[22] Khalid. W, "Handover optimization in femtocell networks," in ICT Convergence (ICTC), 2013, Jeju Island, South Korea 14 – 16 October 2013.

[23] H. Zhang, X. Wen, B. Wang, W. Zheng& Y.Sun, A Novel Handover Mechanism between Femtocell and Macrocell for LTE based Networks, (ICCSN) International Conference on Communication Software and Networks, 2010.

[24] Vivier, G. et al. Femtocells for next-G Wireless Systems: the FREEDOM approach, Future Network & Mobile Summit, Florence, 2010.

[25] Zahir, T. et al. Interference Management in Femtocells. IEEE, Communications Surveys and Tutorials. 2013.

[26] Tyrrell, A. et al. Use Cases, Enablers and Requirements for Evolved Femtocells. IEEE VTC2011, Budapeste, 2011.

[27] T. Camp et al., A Survey of Mobility Models for Ad Hoc Network Research, Wireless Communications and Mobile Computing, vol. 2, no. 5, pp. 483–502, 2002.

A Simulation Study of Effect of MPLS on Latency over a Wide Area Network (WAN)

Adeyinka A. Adewale, Samuel N. John, and Charles Ndujiuba
[1]Department of Electrical and Information Engineering, Covenant University, Ota, Ogun State, Nigeria

Abstract - *A scenario of a service provider network when heavily congested due to heavy traffic flow is brought into play to confirm the attribute of MPLS to significantly improve Latency especially in a highly congested network. The service provider network was congested by sending heavy packets through the ingress and egress port of the network. The network was simulated and ping tests were carried out from a company's head office to one of its branch offices via the service provider network without MPLS implementation and with MPLS enabled. A graphical result showed that latency was reduced drastically with MPLS engaged as the packet forwarding technique on the same Wide Area Network (WAN).*

Keywords: Service provider network, Latency, MPLS, Congestion

1 Introduction

Ethernet Latency can be defined as the time it takes for a network packet to get to a destination or the time it takes to come back from its destination. [1]. It is the delay from the time of the start of packet transmission at the sender to the time of the end of packet reception at the receiver. It also refers to the time taken to deliver an entire data unit (packet) [2]. Latency in a packet-switched network is stated as either one-way latency or Round-Trip Time (RTT). One-way latency is the time required to send a packet from the source to the destination or the RTT divided by two (RTT/2) which means the one-way latency from the source to the destination plus the one-way latency from the destination back to the source divided by two (RTT/2) [1][11]. Latency also refers to any of several kinds of delays typically incurred in processing of network data like the time an application must wait for data to arrive at its destination. Ethernet Latency is also known as End-to-end latency which is a cumulative effect of the individual latencies along the end-to-end network path. Since latency is cumulative it means that adding more links and router between the sender and receiver will increase the latency. Network routers are the devices that create the most latency of any device on the end-to-end path. These network devices (routers) are usually found in network segments. Packet queuing due to link congestion is most often the reason for large amounts of latency through a router [5]. A low latency network is one that generally experiences small delay times which also enables the possibility of having a fast internet connection, while a high latency network generally suffers from long delay times which results in a slow internet connection [3].

Latency affects all network applications to some degree. The degree to which latency affects an application's performance depends on the application's programming model. Latency impacts an application's performance by forcing the application to stall while waiting for the arrival of a packet before it can continue to the next step in its processing. Excessive latency creates bottlenecks which prevent data from filling the network pipe (bandwidth) and delays packet arrival therefore limits the performance of network application [5]. This hinders high-quality network performance needed time-sensitive applications e.g. VoIP, Online games, algorithmic trading [1]. Latency is one of the two key elements that affect a network's performance alongside bandwidth which is the capacity of the network [4]. Speed and capacity are networking concepts that are often commonly misunderstood. For example latency describes 'how fast an internet connection is' while 'how much data can be transmitted per second' is determined by the bandwidth [6]. When it comes to web browsing experience, it turns out that latency, not bandwidth, is the more likely constraining factor at present [7]. A large bandwidth connection only gives the ability to send or receive more data in parallel but not faster as the data still needs to travel the distance and experience the normal delay [8].

MPLS (Multi-Protocol Label Switching) is an Internet based technology that uses short, fixed-length labels to forward packets through the network. MPLS has the attributes of both the layer 2 switching and the layer 3 routing which makes it a very efficient protocol [9] [10]. 'Label switching' indicates that the packets switched are no longer IPv4 packets, IPv6 packets or even Layer 2 frames when switched, but they are labeled [12]. The MPLS labels are advertised between routers so that they can build a label-to-label mapping. These labels are attached to the IP packets enabling the routers to forward the traffic by looking at the label and not the destination IP addresses. The packets are forwarded by label switching instead of by IP switching [12].The fact that MPLS uses labels to forward packets and no longer the destination IP address have led to the popularity of MPLS [12].

Multi-protocol labeled packets are switched after a label lookup/switch instead of a lookup into the IP table. Label lookup and label switching in MPLS is seen to be faster than a routing table or RIB (Routing Information Base) lookup because they could take place directly within the switched fabric and not the CPU [14], [15]. Devices used in an MPLS

Network include customer-edge (CE) routers which is the network device at the customer location that interfaces with the service provider, Provider-Edge (PE) routers which are the device at the edge of the service provider network that interfaces with the customer devices and the provider or core (P) routers which are the devices building the core of the MPLS-enabled network. The PE devices are often also called label switching router edge (LSR-Edge) because they sit at the edge of the MPLS-enabled network. While the provider router have their main functionality which is to label switch traffic based on the most external MPLS tag imposed to each packet and for this reason are often referred to as label switching routers (LSRs)[13].

2 Network design

The scenario of the service provider network to show the effect of MPLS was simulated using GNS3 Software.GNS3 (Graphical Network Simulator version 3) is a network simulator that allows the emulation of complex networks. It provides the user with a realistic feel when configuring the various devices. It is a good tool testing and implementing new infrastructure and devices into an existing architecture. It is also an open source free program that may be used on multiple operating systems [14]. GNS3 provides an estimate of 1,000 packets per second throughput in a virtual environment. A normal router will provide a hundred to a thousand times greater throughput. The devices used in this scenario include routers (customer, provider edge routers and the core routers), Ethernet switches (access and core switches), IP phones, laptop computers and printers.

The simulated service provider network is made of five companies (A- E) and their respective branches connected to the service provider via optical fiber cables as shown in Figure 1.

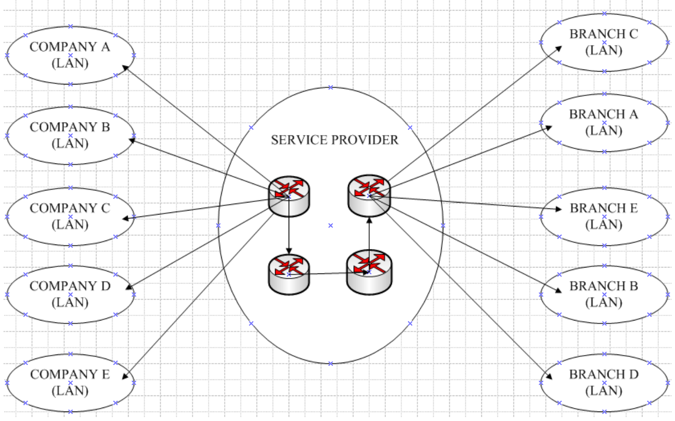


Figure 1: Companies (LAN) and Branches (LAN) connected to the Service Provider Network

The topology area in Figure 2 below shows the design of a local area network (LAN) of one company's Headquarters connected to the LAN of one of its branches via a service provider network. The LAN located at the company's headquarters consists of the customer router, five Cisco Ethernet switches representing five Virtual Local Area Networks (VLANs) for five departments each consisting of IP phones, laptop computers and printers. The main office (headquarters) is linked with the service provider office and the branch office via optical fiber while the LANs at both the companies and their branches make use of Fast-Ethernet twisted-pair cables .The LAN at a branch office is the same with the LAN at the headquarters.

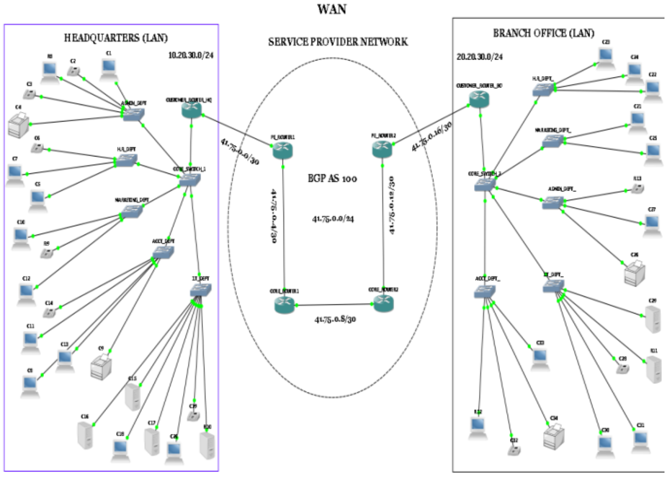
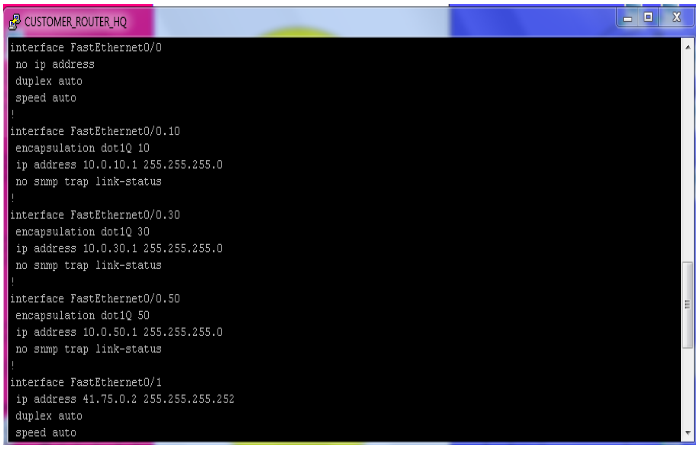


Figure 2: Network Topology Area

The Local Area Network (LAN) of each Branch and Headquarter is linked with Service Provider network via the Provider-Edge (PE) routers while Core routers swap the labeled packets in the Service Provider network (MPLS network).

3 Implementation

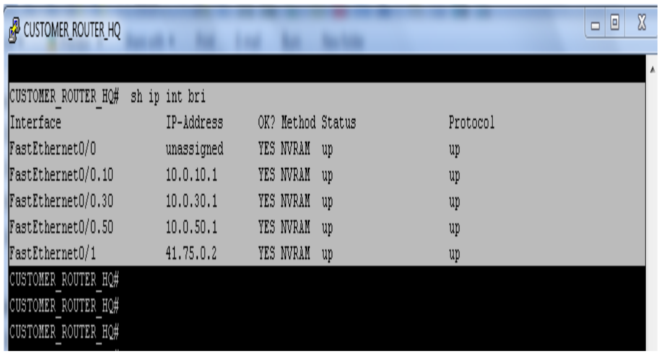
The Figure 3 below shows the running configuration of one of the customer-edge router



```
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
!
interface FastEthernet0/0.10
 encapsulation dot1Q 10
 ip address 10.0.10.1 255.255.255.0
 no snmp trap link-status
!
interface FastEthernet0/0.30
 encapsulation dot1Q 30
 ip address 10.0.30.1 255.255.255.0
 no snmp trap link-status
!
interface FastEthernet0/0.50
 encapsulation dot1Q 50
 ip address 10.0.50.1 255.255.255.0
 no snmp trap link-status
!
interface FastEthernet0/1
 ip address 41.75.0.2 255.255.255.252
 duplex auto
 speed auto
```

Figure 3: Figure showing running configurations on the Customer-edge router

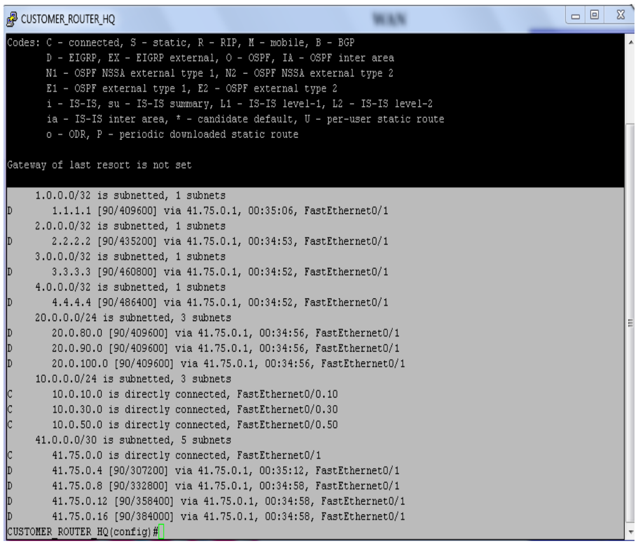
Figure 4 below shows all the working interfaces, and their IP addresses on the customer-edge router using the ‘show interface brief’ command.



```
CUSTOMER_ROUTER_HQ#  sh ip int bri
Interface              IP-Address      OK? Method Status                Protocol
FastEthernet0/0        unassigned      YES NVRAM  up                    up
FastEthernet0/0.10     10.0.10.1       YES NVRAM  up                    up
FastEthernet0/0.30     10.0.30.1       YES NVRAM  up                    up
FastEthernet0/0.50     10.0.50.1       YES NVRAM  up                    up
FastEthernet0/1        41.75.0.2       YES NVRAM  up                    up
CUSTOMER_ROUTER_HQ#
CUSTOMER_ROUTER_HQ#
CUSTOMER_ROUTER_HQ#
```

Figure 1: Configured interfaces on the Customer-edge router

Figure 5 below shows the directly connected routers or routes of the customer-edge router and the EIGRP (Enhanced Interior Gateway Routing Protocol) learned routes that form its routing table which assists the router in routing packets across the network. The configured interfaces can be viewed using the ‘show ip route’ command.

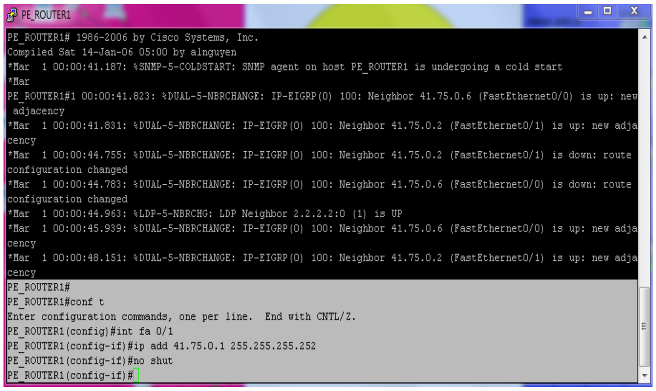


```
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

     1.0.0.0/32 is subnetted, 1 subnets
D       1.1.1.1 [90/409600] via 41.75.0.1, 00:35:06, FastEthernet0/1
     2.0.0.0/32 is subnetted, 1 subnets
D       2.2.2.2 [90/435200] via 41.75.0.1, 00:34:53, FastEthernet0/1
     3.0.0.0/32 is subnetted, 1 subnets
D       3.3.3.3 [90/460800] via 41.75.0.1, 00:34:52, FastEthernet0/1
     4.0.0.0/32 is subnetted, 1 subnets
D       4.4.4.4 [90/486400] via 41.75.0.1, 00:34:52, FastEthernet0/1
     20.0.0.0/24 is subnetted, 3 subnets
D       20.0.80.0 [90/409600] via 41.75.0.1, 00:34:56, FastEthernet0/1
D       20.0.90.0 [90/409600] via 41.75.0.1, 00:34:56, FastEthernet0/1
D       20.0.100.0 [90/409600] via 41.75.0.1, 00:34:56, FastEthernet0/1
     10.0.0.0/24 is subnetted, 3 subnets
C       10.0.10.0 is directly connected, FastEthernet0/0.10
C       10.0.30.0 is directly connected, FastEthernet0/0.30
C       10.0.50.0 is directly connected, FastEthernet0/0.50
     41.0.0.0/30 is subnetted, 5 subnets
C       41.75.0.0 is directly connected, FastEthernet0/1
D       41.75.0.4 [90/307200] via 41.75.0.1, 00:35:12, FastEthernet0/1
D       41.75.0.8 [90/332800] via 41.75.0.1, 00:34:58, FastEthernet0/1
D       41.75.0.12 [90/358400] via 41.75.0.1, 00:34:58, FastEthernet0/1
D       41.75.0.16 [90/384000] via 41.75.0.1, 00:34:58, FastEthernet0/1
CUSTOMER_ROUTER_HQ(config)#
```

Figure 2: Figure showing IP route on the Customer-edge router

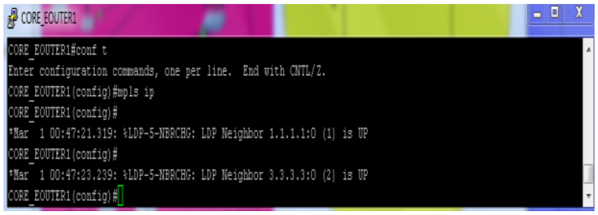
The provider-edge/customer-edge routers and the provider/core routers in the service provider network were configured with IP addresses and they were also configured to disable MPLS and also to enable the same at different points of the testing process.

Figure 6 below shows the IP address configuration of the provider-edge router



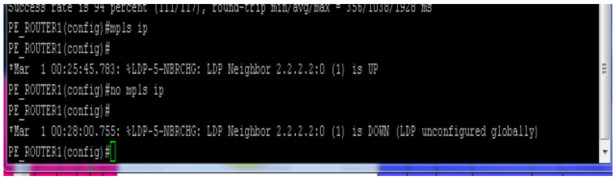
```
PE_ROUTER1# 1986-2006 by Cisco Systems, Inc.
Compiled Sat 14-Jan-06 05:00 by alnguyen
*Mar  1 00:00:41.187: %SNMP-5-COLDSTART: SNMP agent on host PE_ROUTER1 is undergoing a cold start
*Mar
PE_ROUTER1#1 00:00:41.823: %DUAL-5-NBRCHANGE: IP-EIGRP(0) 100: Neighbor 41.75.0.6 (FastEthernet0/0) is up: new
 adjacency
*Mar  1 00:00:41.831: %DUAL-5-NBRCHANGE: IP-EIGRP(0) 100: Neighbor 41.75.0.2 (FastEthernet0/1) is up: new adja
cency
*Mar  1 00:00:44.755: %DUAL-5-NBRCHANGE: IP-EIGRP(0) 100: Neighbor 41.75.0.2 (FastEthernet0/1) is down: route
configuration changed
*Mar  1 00:00:44.783: %DUAL-5-NBRCHANGE: IP-EIGRP(0) 100: Neighbor 41.75.0.6 (FastEthernet0/0) is down: route
configuration changed
*Mar  1 00:00:44.963: %LDP-5-NBRCHG: LDP Neighbor 2.2.2.2:0 (1) is UP
*Mar  1 00:00:45.939: %DUAL-5-NBRCHANGE: IP-EIGRP(0) 100: Neighbor 41.75.0.6 (FastEthernet0/0) is up: new adja
cency
*Mar  1 00:00:48.151: %DUAL-5-NBRCHANGE: IP-EIGRP(0) 100: Neighbor 41.75.0.2 (FastEthernet0/1) is up: new adja
cency
PE_ROUTER1#
PE_ROUTER1#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
PE_ROUTER1(config)#int fa 0/1
PE_ROUTER1(config-if)#ip add 41.75.0.1 255.255.255.252
PE_ROUTER1(config-if)#no shut
PE_ROUTER1(config-if)#
```

Figure 3: IP address configurations on the Provider-edge router



```
CORE_EOUTER1#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
CORE_EOUTER1(config)#mpls ip
CORE_EOUTER1(config)#
*Mar  1 00:47:21.319: %LDP-5-NBRCHG: LDP Neighbor 1.1.1.1:0 (1) is UP
CORE_EOUTER1(config)#
*Mar  1 00:47:23.239: %LDP-5-NBRCHG: LDP Neighbor 3.3.3.3:0 (2) is UP
CORE_EOUTER1(config)#
```

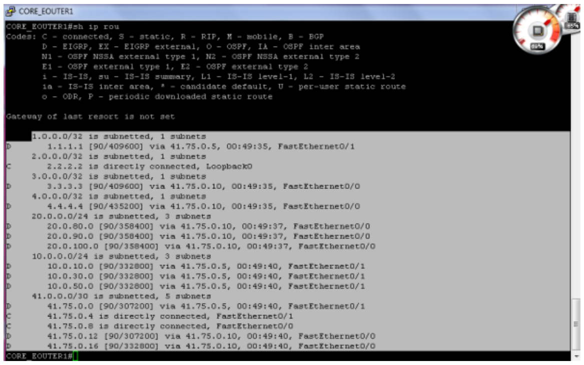
Figure 4: MPLS configurations on the Core router



```
Success rate is 94 percent (111/117), round-trip min/avg/max = 356/1036/1928 ms
PE_ROUTER1(config)#mpls ip
PE_ROUTER1(config)#
*Mar  1 00:25:45.783: %LDP-5-NBRCHG: LDP Neighbor 2.2.2.2:0 (1) is UP
PE_ROUTER1(config)#no mpls ip
PE_ROUTER1(config)#
*Mar  1 00:28:00.755: %LDP-5-NBRCHG: LDP Neighbor 2.2.2.2:0 (1) is DOWN (LDP unconfigured globally)
PE_ROUTER1(config)#
```

Figure 5: Disabling of MPLS (using the ‘no MPLS IP’ command

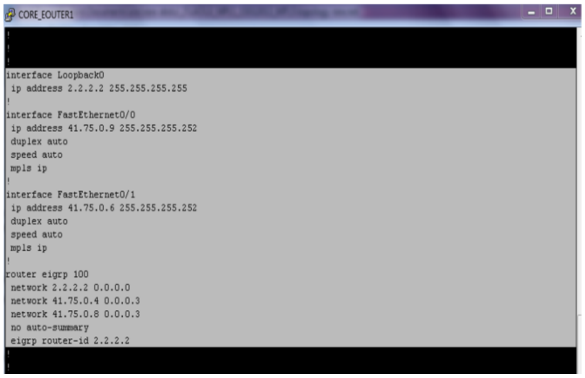
Figure 9 below shows the directly connected routers or routes of the core router and the EIGRP learned routes that form its routing table in which the router uses to route packets across the network. This configuration can be viewed using the ‘show ip route’ command



```
CORE_EOUTER1#sh ip rou
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

     1.0.0.0/32 is subnetted, 1 subnets
D       1.1.1.1 [90/409600] via 41.75.0.5, 00:49:35, FastEthernet0/1
     2.0.0.0/32 is subnetted, 1 subnets
C       2.2.2.2 is directly connected, Loopback0
     3.0.0.0/32 is subnetted, 1 subnets
D       3.3.3.3 [90/409600] via 41.75.0.10, 00:49:35, FastEthernet0/0
     4.0.0.0/32 is subnetted, 1 subnets
D       4.4.4.4 [90/435200] via 41.75.0.10, 00:49:35, FastEthernet0/0
     20.0.0.0/24 is subnetted, 3 subnets
D       20.0.80.0 [90/358400] via 41.75.0.10, 00:49:37, FastEthernet0/0
D       20.0.90.0 [90/358400] via 41.75.0.10, 00:49:37, FastEthernet0/0
D       20.0.100.0 [90/358400] via 41.75.0.10, 00:49:37, FastEthernet0/0
     10.0.0.0/24 is subnetted, 3 subnets
D       10.0.10.0 [90/332800] via 41.75.0.5, 00:49:40, FastEthernet0/1
D       10.0.30.0 [90/332800] via 41.75.0.5, 00:49:40, FastEthernet0/1
D       10.0.50.0 [90/332800] via 41.75.0.5, 00:49:40, FastEthernet0/1
     41.0.0.0/30 is subnetted, 5 subnets
D       41.75.0.0 [90/307200] via 41.75.0.5, 00:49:40, FastEthernet0/1
C       41.75.0.4 is directly connected, FastEthernet0/1
C       41.75.0.8 is directly connected, FastEthernet0/0
D       41.75.0.12 [90/307200] via 41.75.0.10, 00:49:40, FastEthernet0/0
D       41.75.0.16 [90/332800] via 41.75.0.10, 00:49:40, FastEthernet0/0
CORE_EOUTER1#
```

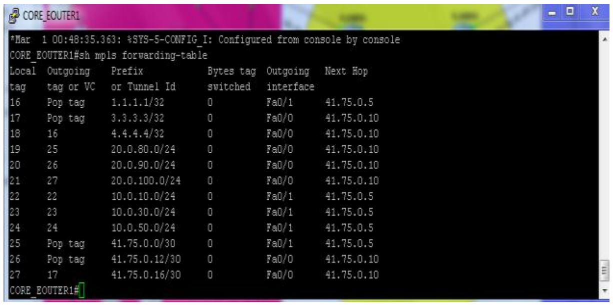
Figure 6: IP routes of the Core router

Figure 10 below shows all the working interfaces, and their IP addresses using the ‘show interface brief’ command.



```
!
!
!
interface Loopback0
 ip address 2.2.2.2 255.255.255.255
!
interface FastEthernet0/0
 ip address 41.75.0.9 255.255.255.252
 duplex auto
 speed auto
 mpls ip
!
interface FastEthernet0/1
 ip address 41.75.0.6 255.255.255.252
 duplex auto
 speed auto
 mpls ip
!
router eigrp 100
 network 2.2.2.2 0.0.0.0
 network 41.75.0.4 0.0.0.3
 network 41.75.0.8 0.0.0.3
 no auto-summary
 eigrp router-id 2.2.2.2
!
!
```

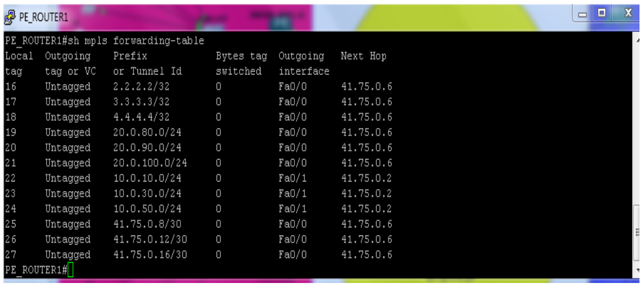
Figure 7: Configured interfaces on the core router



```
*Mar  1 00:48:35.363: %SYS-5-CONFIG_I: Configured from console by console
CORE_EOUTER1#sh mpls forwarding-table
Local  Outgoing    Prefix            Bytes tag  Outgoing   Next Hop
tag    tag or VC   or Tunnel Id      switched   interface
16     Pop tag     1.1.1.1/32        0          Fa0/1      41.75.0.5
17     Pop tag     3.3.3.3/32        0          Fa0/0      41.75.0.10
18     16          4.4.4.4/32        0          Fa0/0      41.75.0.10
19     25          20.0.80.0/24      0          Fa0/0      41.75.0.10
20     26          20.0.90.0/24      0          Fa0/0      41.75.0.10
21     27          20.0.100.0/24     0          Fa0/0      41.75.0.10
22     22          10.0.10.0/24      0          Fa0/1      41.75.0.5
23     23          10.0.30.0/24      0          Fa0/1      41.75.0.5
24     24          10.0.50.0/24      0          Fa0/1      41.75.0.5
25     Pop tag     41.75.0.0/30      0          Fa0/1      41.75.0.5
26     Pop tag     41.75.0.12/30     0          Fa0/0      41.75.0.10
27     17          41.75.0.16/30     0          Fa0/0      41.75.0.10
CORE_EOUTER1#
```

Figure 8: MPLS forwarding table on the Core router

6000 packets are also sent from PE-router 2 to PE-router 1 10000 times which keeps the network congested for the period.



```
PE_ROUTER1#sh mpls forwarding-table
Local  Outgoing    Prefix            Bytes tag  Outgoing   Next Hop
tag    tag or VC   or Tunnel Id      switched   interface
16     Untagged    2.2.2.2/32        0          Fa0/0      41.75.0.6
17     Untagged    3.3.3.3/32        0          Fa0/0      41.75.0.6
18     Untagged    4.4.4.4/32        0          Fa0/0      41.75.0.6
19     Untagged    20.0.80.0/24      0          Fa0/0      41.75.0.6
20     Untagged    20.0.90.0/24      0          Fa0/0      41.75.0.6
21     Untagged    20.0.100.0/24     0          Fa0/0      41.75.0.6
22     Untagged    10.0.10.0/24      0          Fa0/1      41.75.0.2
23     Untagged    10.0.30.0/24      0          Fa0/1      41.75.0.2
24     Untagged    10.0.50.0/24      0          Fa0/1      41.75.0.2
25     Untagged    41.75.0.8/30      0          Fa0/0      41.75.0.6
26     Untagged    41.75.0.12/30     0          Fa0/0      41.75.0.6
27     Untagged    41.75.0.16/30     0          Fa0/0      41.75.0.6
PE_ROUTER1#
```

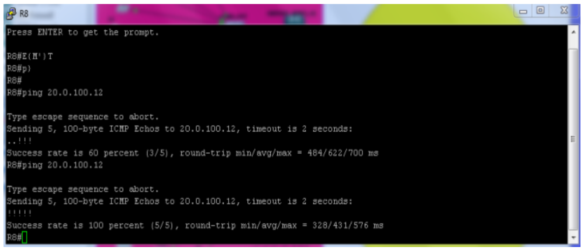
Figure 9: MPLS forwarding table on the Core router

4 Testing and results

Network tools like ping tests and trace route were used to measure latency by determining the time it takes a given network packet to travel from source to destination and back (RTT) then dividing the time by two (RTT/2), the most common technique used is the ping test.

4.1 Measuring network latency

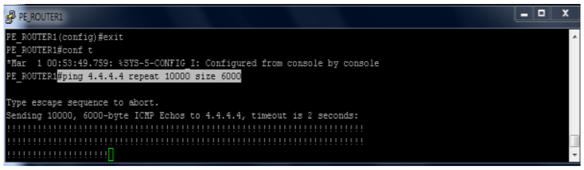
After the network is designed and simulated, a ping test was carried out from the headquarters to the branch office to confirm that the branch office can be reached from the headquarters via the service provider network and vice-versa.



```
Press ENTER to get the prompt.

R8#E(M')T
R8#p)
R8#
R8#ping 20.0.100.12

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
..!!!
Success rate is 60 percent (3/5), round-trip min/avg/max = 484/622/700 ms
R8#ping 20.0.100.12

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 328/431/576 ms
R8#
```

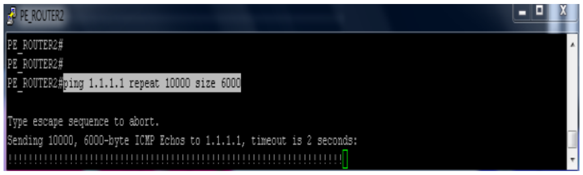
Figure 10: A ping test from headquarters to branch office

After ping test, 6000 packets are then sent from PE-router 1 to PE-router 2 10000 times with MPLS disabled. This is to keep the service provider network congested for the period it would take PE-router 1 to send 6000 packets to PE-router 2.



```
PE_ROUTER1(config)#exit
PE_ROUTER1#conf t
*Mar  1 00:53:49.759: %SYS-5-CONFIG_I: Configured from console by console
PE_ROUTER1#ping 4.4.4.4 repeat 10000 size 6000

Type escape sequence to abort.
Sending 10000, 6000-byte ICMP Echos to 4.4.4.4, timeout is 2 seconds:
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!
```

Figure 11: Traffic Build-up on PE router 1

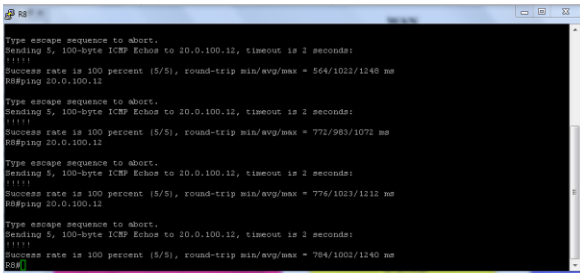


```
PE_ROUTER2#
PE_ROUTER2#
PE_ROUTER2#ping 1.1.1.1 repeat 10000 size 6000

Type escape sequence to abort.
Sending 10000, 6000-byte ICMP Echos to 1.1.1.1, timeout is 2 seconds:
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

Figure 12: Traffic build-up on PE-router 2

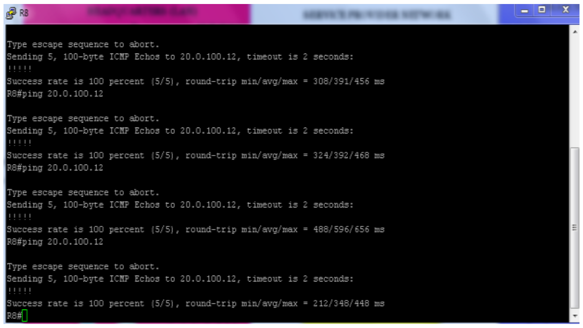
A ping test is then carried out from the headquarters to the branch office and latency is calculated from the round trip time (RTT). Latency = RTT/2.

The process is repeated with MPLS enabled and latency values were recorded. The simulated network is again tested with network congestion of 9000, 12000, 15000, 18000 packets sent 10000 times, recording the latency values when MPLS is disabled and when MPLS is enabled.



```
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 564/1022/1248 ms
R8#ping 20.0.100.12

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 772/983/1072 ms
R8#ping 20.0.100.12

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 776/1023/1212 ms
R8#ping 20.0.100.12

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 784/1002/1240 ms
R8#
```

Figure 13: Ping tests from HQ to branch office (MPLS disabled)



```
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 308/391/456 ms
R8#ping 20.0.100.12

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 324/392/468 ms
R8#ping 20.0.100.12

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 488/596/656 ms
R8#ping 20.0.100.12

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.100.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 212/348/448 ms
R8#
```

Figure 14: Ping test from HQ to branch office (MPLS enabled)

The results of the simulation and test are tabulated as shown below in Table 1. Figure 18 and Figure 19 are the graphical realization of the results. The graphical comparison shows a sharp rise in latency as the network is getting congested but the reverse is the case when MPLS is enabled showing a drastic reduction in latency. Also, it can be inferred from the graph (Figure 19) that the latency with MPLS enabled is decreasing with increasing core network congestion which implied that MPLS is a good technology for congested WAN core network.

Table 1 Results of Traffic Congestion with different test packets

Order of Ping Test/No. Packets	Delay (RTT/2) in ms (MPLS Disabled)					Delay (RTT/2) in ms (MPLS Enabled)				
	6000	9000	12000	15000	18000	6000	9000	12000	15000	18000
1st	331	401	543.5	749	607.5	157	142.5	188.5	158	208
2nd	301.5	382.5	511	558	554	151.5	157.5	195.5	159.5	163
3rd	420	394	491.5	742.5	530	246.5	244	196	144.5	144
4th	431.5	450	511.5	533.5	690	200	226	298	260	133.5
5th	390.5	380	501	649.5	550	190	169	174	142.5	131.5
Avg.Delay(D/5)	374.9	401.5	511.7	646.5	586.3	189	187.8	210.4	172.9	156

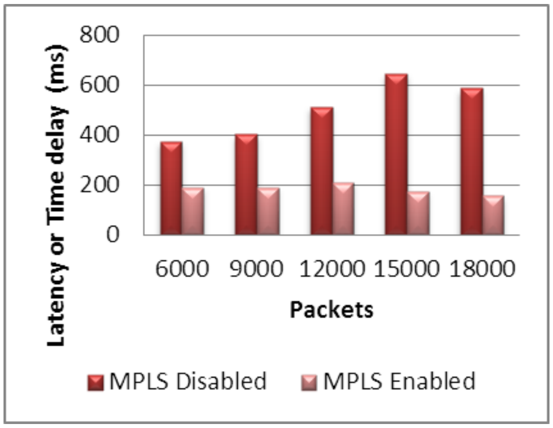


Figure 15: A Bar chart showing Latency per increase in number of packets with regards to (MPLS Enabled) and (MPLS Disabled

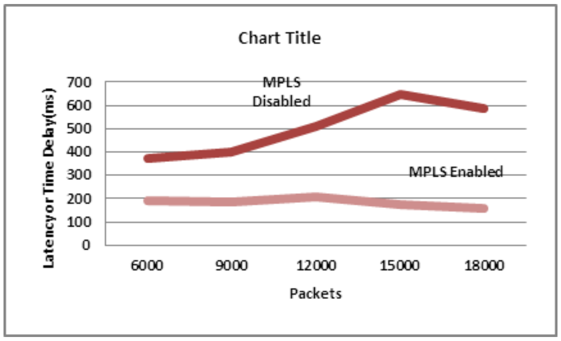


Figure 16: Graph of Latency per increase in number of packets with regards to MPLS Enabled and MPLS Disabled Scenarios

5 Conclusion

Results gotten from the simulation shows that the Latency in an IP network that is when MPLS is disabled, rises sharply as number of packets in the core of the service provider network increases while it drops drastically when MPLS is enabled, even as number of packets in the core of the service provider network increases. From the simulation, it is clear that MPLS is a better technique for improving latency when compared with the traditional IP network in that it takes less time to send data from a source to its destination. MPLS will therefore be more efficient if applied in the current internet architecture. Moreover, the potential of the MPLS technology is yet untapped in most parts of the world with respect to the services it can provide such as MPLS VPN (Virtual Private Network) and MPLS TE (Traffic Engineering) among others. Enterprises and service providers can experience an improvement in the rate of achievement of business targets by implementing and maximizing the capabilities of MPLS in their networks.

6 References

[1] Aliso Viejo, “Introduction to Ethernet Latency“ QLogic Corporation, 2011.

[2] Wikipedia. Retrieved March 30, 2014, http://en.wikipedia.org/wiki/Low_latency

[3] Horney, C., “Quality of Service and Multi-Protocol Label Switching”. Irvine, CA: Nuntius Systems Inc., 2014

[4] Mitchell, B. (n.d.). Network Bandwidth and Latency. About.com Wireless/Networking. Retrieved March 28, 2014, http://compnetworking.about.com/od/speedtests/a/network_latency.htm

[5] Smutz, E. (n.d.). Network Latency. Network Latency. Retrieved March 30, 2014, from http://smutz.us/techtips/NetworkLatency.html

[6] Latency versus Bandwidth - What is it? DSL FAQ | DSL Reports, ISP Information. (n.d.). Latency versus Bandwidth - What is it? DSL FAQ | DSL Reports, ISP Information. Retrieved March 30, 2014, from http://www.dslreports.com/faq/694

[7] Grigorik I., “Latency: The New Web Performance Bottleneck”, July, 2012. http://www.igvita.com/2012/07/19/latency-the-new-web-performance-bottleneck/

[8] Hoffman, B. “Bandwidth, Latency, and the Size of your Pipe”, December, 2011. http://zoompf.com/blog/2011/12/i-dont-care-how-big-yours-is

[9] Teng, Y., “Multi-Protocol Label Switching”, Computer Science and Engineering. Lecture conducted from University of Maryland Baltimore County, Baltimore, Maryland, March 2010.

[10] Traffic Engineering and QoS with MPLS and its applications, http://utta.csc.ncsu.edu/csc570_fall10/wrap/MPLS_Report.pdf , 2014

[11] Latency (engineering). http://en.wikipedia.org/wiki/Latency (engineering), 2014

[12] Ghein, L., “MPLS Fundamentals”. Indianapolis, Ind.: Cisco Press, 2007.

[13] MPLS Basic MPLS Configuration Guide, (2008). San Jose, CA: Cisco Systems, Inc.

[14] Welsh, C., “GNS3 network simulation guide”, Birmingham, UK: Packet Publishing,. October, 2013. .

A Heuristic Algorithm for Longest Simple Cycle Problem

Parveen Kumar
National Institute of Technology Hamirpur
Himachal Pradesh, India
parveen.csed@gmail.com

Nitin Gupta
National Institute of Technology Hamirpur
Himachal Pradesh, India
nitin3041@gmail.com

Abstract: **The longest simple cycle problem is the problem of finding a cycle of maximum length in a graph. As a generalization of the Hamiltonian cycle problem, it is NP-complete on general graphs and, in fact, on every class of graphs that the Hamiltonian cycle problem is NP-complete. The longest simple cycle problem may be solved in polynomial time on the complements of comparability graphs. It may also be solved in polynomial time on any class of graphs with bounded tree width or bounded clique-width, such as the distance-hereditary graphs. However, it is NP-hard even when restricted to split graphs, circle graphs, or planar graphs. In this paper a heuristic algorithm is proposed which can solve the problem in polynomial time. To solve the longest simple cycle problem using adjacency matrix and adjacency list by making a tree of given problem to find the longest simple cycle as the deepest path in tree following reconnect the leaf node of deepest path with root node. The result resolves the open question for the complexity of the problem on simple unweighted graphs. The algorithm is implemented on a simple labeled graph without parallel edges and without self-loop. The worst case time complexity for the proposed algorithm is O(V+E).**

Index Terms: **Adjacency List, Adjacency Matrix, Graph, Tree, NP-Completeness**

I. INTRODUCTION

The area of approximation algorithms for NP-hard optimization problems has received much attention over the past two decades [1], [2]. Although some notable positive results have been obtained, such as the fully polynomial approximation scheme for bin packing [3][4], it has now become apparent that even the approximate solution of a large class of NP-hard optimization problems remains outside the bounds of feasibility. For example, a sequence of results [5][6][7][8][9] established the intractability of approximating the largest clique in a graph. The optimization version of this problem is NP-hard since it includes the Hamiltonian path problem as a special case. Therefore, it is natural to look for polynomial-time algorithms with a small performance ratio, where the performance ratio is defined as the ratio of the longest path in the input graph to the length of the path produced by the algorithm. Our results attempt to pin down the best possible performance ratio achievable by polynomial-time approximation algorithms for longest paths (with same start and end vertices). In this paper, we address the problem of finding the longest Simple Cycle in an undirected graph G=(V, E). V is the set of n vertices and E is the set of m edges. The goal is to find for all u, v ϵ V, the longest path from u to u, using maximum number of edges. We consider simple paths, which do not have any repeated edges or vertices. This problem belongs to the NP-Complete class of problems, as it is a generalization of the Hamiltonian path problem and cannot be solved in polynomial time unless P=NP. For this reason, the algorithm proposed is approximation algorithm. The main objective of this work is to apply various approaches combined to solve this problem (i.e. adjacency list, adjacency Matrix, Tree). In Section 2, various proposed algorithms are discussed and work which is already done on this topic is also discussed. In Section 3, describe the proposed algorithm and its implementation steps. In Section 4, experimental results, and finally the conclusion.

II. PRELIMINARY AND BACKGROUND WORK:

The longest Simple Cycle problem, i.e. the problem of finding a simple path (with same start and end vertices) with the maximum number of vertices, is one of the most important problems in graph theory. The well-known NP-complete Hamiltonian path problem [10,11], i.e. deciding whether there is a simple path that visits each vertex of the graph exactly once, is a special case of the longest path problem. Only a few polynomial-time algorithms are known for the longest path problem for special classes of graphs. Trees are the first class of graphs that a polynomial-time algorithm for finding the diameter of an unweighted tree. Originally, this algorithm was proposed by Dijkstra around 1960 but Bulterman et al. [12] provided a proof for it, and later it was improved by Uehara and Uno [13] for the case of weighted trees. They also solved the problem for block graphs in linear time and for cacti in quadratic time. Recently, Ioannidou et al. [14] showed that the problem is polynomial for general interval graphs. Their algorithm is based on dynamic programming and runs in O (n4) time. More recently, Mertzios and Corneil [15] solved the problem in polynomial time for the larger class of graphs i.e. cocomparability graphs. Also, there is an O (n3)-time algorithm for the problem for complete m-partite digraphs proposed by Gutin [16]. Also, it has been shown that finding a path of length n − nϵ is not possible in polynomial time unless P = NP [17]. To our knowledge, the best-known approximation algorithm for the problem has the ratio of O (n (log log n/ log n)2) [17]. For more related results on approximation algorithms on general graphs see [18][19][20][21]. Monien [7] presented an 0 (k! n m)-time algorithm that finds paths of length k in a Hamiltonian graph with n vertices and m edges. since in polynomial time Monien's algorithm can only find paths of length O(log n/log log n). Furer and Raghavachari [22] present approximation algorithms for the minimum-degree spanning tree problem that delivered absolute performance guarantees (within an additive factor of 1). No hardness results for longest paths were known earlier, although a seemingly related problem has been studied by Berman and Schnitger [23]. They show that the hardness conjecture is true for the problem of approximating the longest induced path in an undirected graph. Note that the induced-path problem is strictly harder and their hardness result does not carry over to the problem under consideration here. Bellare [24] considers a generalization of the longest-paths problem called the longest color-respecting path problem. This involves graphs with 2-colored edges and labeled vertices, and a feasible path must have the property that at each vertex its label specifies whether the incident edges of the path are of the same color or not. He obtains essentially the same hardness results through different techniques.

III. REPRESENTATIONS OF GRAPHS

There are two standard ways to represent a graph G = (V, E): as a collection of adjacency lists or as an adjacency matrix. Either way applies to both directed and undirected graphs. Because the adjacency-list representation provides a compact way to represent sparse graphs—those for which |E| is much less than |V |2 — it is usually the method of choice. Most of the graph algorithms assume that an input graph is represented in adjacency list form. We may prefer an adjacency-matrix representation, however, when the graph is dense—|E| is close to |V|2—or when we need to be able to tell quickly if there is an edge connecting two given vertices.

The adjacency-list representation of a graph G = (V, E): consists of an array Adj of |V| lists, one for each vertex in V. For each u ∈ V, the adjacency list Adj[u] contains all the vertices v such that there is an edge (u,v) ϵ E. That is, Adj[u] consists of all the vertices adjacent to u in G. (Alternatively, it may contain pointers to these vertices.) Since the adjacency lists represent the edges of a graph,. If G is an undirected graph, the sum of the lengths of all the adjacency lists is 2|E|, since if (u,v) is an undirected edge, then u appears in v’s adjacency list and vice versa. For undirected graphs, the adjacency-list representation has the desirable property that the amount of memory it requires is ɵ(V+E). It can readily adapt adjacency lists to represent weighted graphs, that is, graphs for which each edge has an associated weight, typically given by a weight function w: E→R. For example, let G=(V,E) be a weighted graph with weight function w. The weight w(u,v) of the edge (u,v) ϵ E simply store with vertex v in u’s adjacency list. The adjacency-list representation is quite robust in a way that it can be modified to support many other graph variants. A potential disadvantage of the adjacency-list representation is that it provides no quicker way to determine whether a given edge (u,v) is present in the graph than to search for v in the adjacency list Adj[u]. An adjacency-matrix representation of the graph remedies this disadvantage, but at the cost of using asymptotically more memory.

For the adjacency-matrix representation of a graph G =(V,E), it is assumed that the vertices are numbered 1,2, ……,|V| in some arbitrary manner. Then the adjacency-matrix representation of a graph G consists of a |V|*|V| matrix A=(aij) such that

$$\text{aij} \begin{cases} 1 & \text{if}(i, j) \in E\,, \\ 0 & \text{otherwise} \end{cases}$$

Like the adjacency-list representation of a graph, an adjacency matrix can represent a weighted graph. For example, if G = (V, E) is a weighted graph with edge weight function w, it can simply store the weight w(u,v) of the edge (u,v) ∈ E as the entry in row u and column v of the adjacency matrix. If an edge does not exist, it can store a NIL value as its corresponding matrix entry, though for many problems it is convenient to use a value such as 0 or ∞. Although the adjacency-list representation is asymptotically at least as space efficient as the adjacency-matrix representation, adjacency matrices are simpler, and so they may preferred when graphs are reasonably small. Moreover, adjacency matrices carry a further advantage for unweighted graphs: they require only one bit per entry.

A. Complexity Classes

Mainly three classes of problems are referred: P, NP, and NPC, the latter class being the NP-complete problems. Which are described formally as:

Class P

The class P consists of those problems that are solvable in polynomial time. More specifically, they are problems that can be solved in time O(nk) for some constant k, where n is the size of the input to the problem. A lot of the problems considered in P. Any problem in P is also in NP, since if a problem is in P then it can be solved in polynomial time without even being supplied a certificate. For now it is considered P ⊂ NP. The open question is whether or not P is a proper subset of NP.

The Complexity Class NP

The complexity class NP is the class of languages that can be verified by a polynomial- time algorithm. More precisely, a language L belongs to NP if and only if there exist a two-input polynomial-time algorithm A and a constant c such that L = {

x ∈ {0,1} there exists a certificate y with |y| = O(|x|) such that A(x, y) = 1}.

It can be said that algorithm A verifies language L in polynomial time. It is unknown whether P = NP, but most researchers believe that P and NP are not the same class. Intuitively, the class P consists of problems that can be solved quickly. The class NP consists of problems for which a solution can be verified quickly. it is often more difficult to solve a problem from scratch than to verify a clearly presented solution.

Complexity Class NP complete

NP-complete problems arise in diverse domains: boolean logic, graphs, arithmetic, network design, sets and partitions, storage and retrieval, sequencing and scheduling, mathematical programming, algebra and number theory, games and puzzles, automata and language theory, program optimization, biology, chemistry, physics, and more. In this section, the reduction methodology used to provide NP completeness proofs for a variety of problems drawn from graph theory and set partitioning.

B. The Longest Simple Cycle Algorithm

In the Proposed Algorithm, the input graph considered to be a simple graph (i.e. without self-loop and without parallel edges), the algorithm for Longest Simple Cycle in simple graph is summarized below:

1. Enumerate all the nodes to calculate degree of each node to find the node with highest degree.
2. Assign the node with highest degree as the root for tree.
3. Construct a tree T of the given graph G considering the adjacent nodes as successor and predecessors accordingly for each vertex using adjacency matrix.
4. Do apply the proposed LSC algorithm to find the longest path.
5. Join the leaf node of the longest path with root and retrieve the path considering it as the longest cycle in graph.

From now on, we describe each step of the algorithm in more detail.

1. *Enumerate all the nodes to calculate degree of each node to find the node with highest degree.*

For a given Graph G=(V,E) where V is the set vertices n and E is the set of edges e, first to make an adjacency matrix for all vertices to find the maximum degree vertex, which would become the root of tree.

2. *Assign the node with highest degree as the root for tree.*

Now assign the MAX (which was returned by pseudo code) as root.

ROOT ← MAX

3. *Construct a tree of the given graph considering the adjacent nodes as successor and predecessors accordingly for each vertex.*

The vertex with maximum degree is taken as root and the adjacent vertices are considered as predecessor of root and taken as child of root vertex, now for child vertices their adjacent vertices are taken as predecessor of child vertex and so on. This process of converting graph into tree will go on till all vertices and their adjacent vertices are expended.

Figure 1 shown below the tree constructed from graph G.

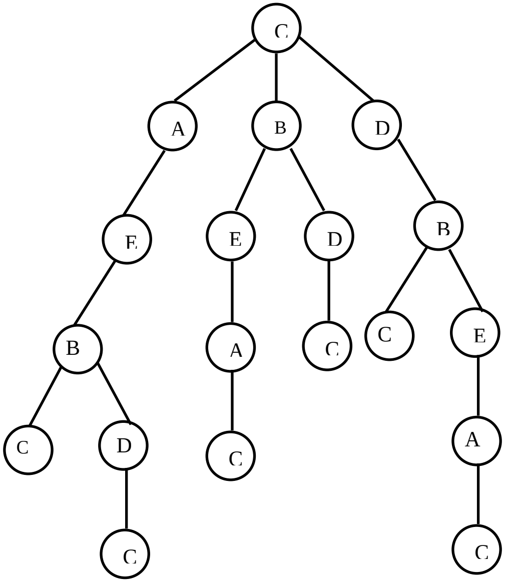


Figure 1. tree constructed by using adjacency matrix to represent and expand adjacent vertex for all vertices

4. *Do apply the proposed LSC algorithm to find the longest path.*

The Proposed algorithm for Longest Simple Cycle Problem is proposed as:

LSC(G)

```
for each vertex u ∈ G.[V- ROOT]
    {
      Color[u] ← white
      Pred[u] ← NIL
    }
count = 0
for each vertex u ∈ G.adj[ROOT]
    {
      if color[u] = white
        then LSC_TRAV (u)
      endif
    }
```

LSC_TRAV (u)

```
Color[u] ← pink   // vertex u has just discovered
Count ← count + 1
Discover[u] ← count
for each v ∈ G.adj(u)             //Explore edge (u,v)
      {
        if (v = ROOT) then
            Finish[u] ← count + 1
        else
         if (adj(u) = NIL) then
10.          Finish[u] ← count
11.       else
12.        if color[v] = white then
13.          Pred[v] ← u
14.          LSC_TRAV (G.v)
15.       }
16.  Color[u] ← RED        //make u RED ; it is finished
17.  count ← count + 1
18.  finish[u] ← count
```

Procedure LSC works as follows. Lines 1-4 paint all vertices white and initialize their PRES field to NIL. Line 5 resets the global counter. Line 7-12 check each vertex in V in turn and, when a white vertex is found, visit it using LSC_TRAV. Every Time LSC_TRAV (u) is called in line 10, vertex u becomes the root of a new tree T'. When LSC returns, every vertex u has been assigned a discovery time **Discover[u]** and a finish time **finish[u].**

In each call LSC_TRAV (u), vertex u is initially white. Line 1 paints u pink, line 2 increments the global variable count and line 3 records the new value of time as the discovery time discover[u]. Lines 4-15 examine each vertex v if it is white. As each vertex v ∈ adj[u] is considered in line 4 we say that each is explored by this step. Finally after each edge leaving u has been explored, line 15 paint u pink and record the finishing time finish[u]. Note that result of Longest simple Cycle Algorithm depend upon the order in which the vertices are examined in line 7 of LSC, and upon the order in which the neighbors of a vertex are visited in line 6,9 and 12 of LSC_TRAV. These different visitation orders tend not to cause problem in execution as any order of exploring result can usually be used effectively, with essentially equivalent results.

The running time of LSC is computed as follows:

The loop on lines 1-5 and lines 7-12 of LSC takes time O(V), exclusive of the time to execute the call LSC_TRAV. As the procedure LSC_TRAV is called exactly once for each vertex v ∈ V , since LSC_TRAV invoked only white vertices and the first thing it does it paint vertices pink. During an execution of LSC_TRAV (v), the loop on line 4-15 is executed |Adj[v]| times. Since

$$\sum_{v \in V} |\mathrm{Adj}[v]| = \Theta(E)$$

The total cost of executing lines 4-15 of LSC_TRAV is$\Theta(E)$. The running time of LSC is therefore $\Theta(V + E)$

5. *Join the leaf node of the longest path with root and retrieve the path considering it as the longest cycle in graph.*

The pseudo code to retrieve longest path from the above tree is:

```
     PRINT_CYCLE (G, ROOT, V)
1.   if (v == ROOT)
2.       then print "ROOT"
3.       else
4.         if (V.PRED = NIL)
5.           then print "no cycle exist"
6.           else
7.          PRINT_CYCLE(G, ROOT, V.PRED)
8.          print V
```

Procedure PRINT_CYCLE runs in time Linear in the number of vertices in the Cycle printed. Since each recursive call is for a path one vertex shorter.

The given graph after connecting deepest leaf with root is shown in figure 2:

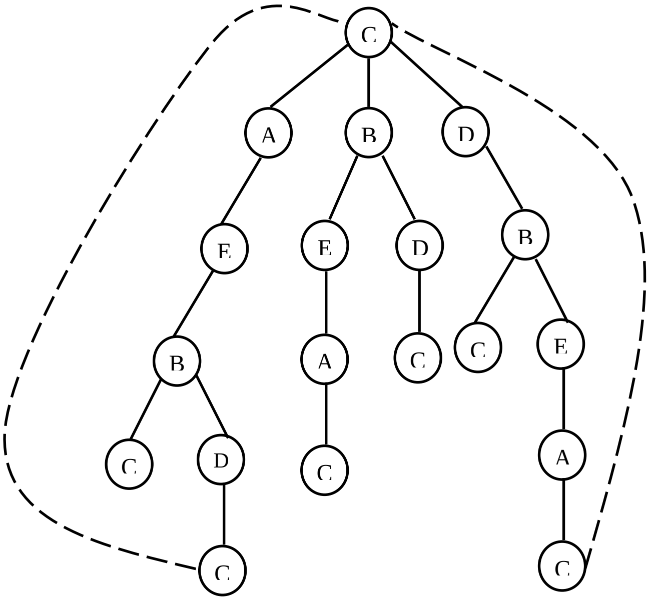


Figure 2. Connecting deepest vertex with root vertex to make a cycle

IV EXPERIMENTAL RESULTS:

For experimental research and proof of algorithm the problem graph is considered is shown in figure 3:

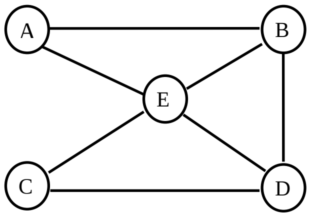


Figure 3. graph G = (V, E) considered to find Longest simple cycle

Now, after applying the proposed Longest Simple Cycle Algorithm the diagrammatic procedure is shown below in figures:

1. *The adjacency matrix for the given problem graph is shown in figure 4:*

	A	B	C	D	E
A	0	1	0	0	1
B	1	0	1	0	1
C	1	1	0	1	1
D	0	0	1	0	1
E	0	1	1	1	0

Figure 4. Adjacency matrix for Graph G

2. Assianina the node with hiahest dearee as the root for tree.

$$\begin{aligned}
D_1 &= e_{11} + e_{12} + e_{13} + e_{14} + e_{15} & 0 + 1 + 0 + 0 + 1 = 2\\
D_2 &= e_{21} + e_{22} + e_{23} + e_{24} + e_{25} & 1 + 0 + 1 + 0 + 1 = 3\\
D_3 &= e_{31} + e_{32} + e_{33} + e_{34} + e_{35} & 1 + 1 + 0 + 1 + 1 = 4\\
D_4 &= e_{41} + e_{42} + e_{43} + e_{44} + e_{45} & 0 + 0 + 1 + 0 + 1 = 2\\
D_5 &= e_{51} + e_{52} + e_{53} + e_{54} + e_{55} & 0 + 1 + 1 + 1 + 0 = 3
\end{aligned}$$

From the above adjacency matrix it is calculated that D3 has highest degree i.e. Vertex C. Now assign vertex C as Root Node for the tree.

ROOT $\leftarrow$ C

3. Construct a tree of the given graph considering the adjacent nodes as successor and predecessors accordingly for each vertex.

After assigning vertex C as ROOT node tree of adjacent vertices is shown in figure 5.

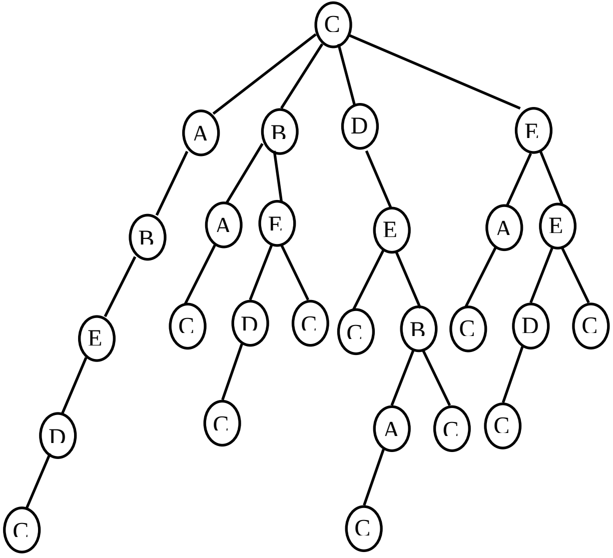
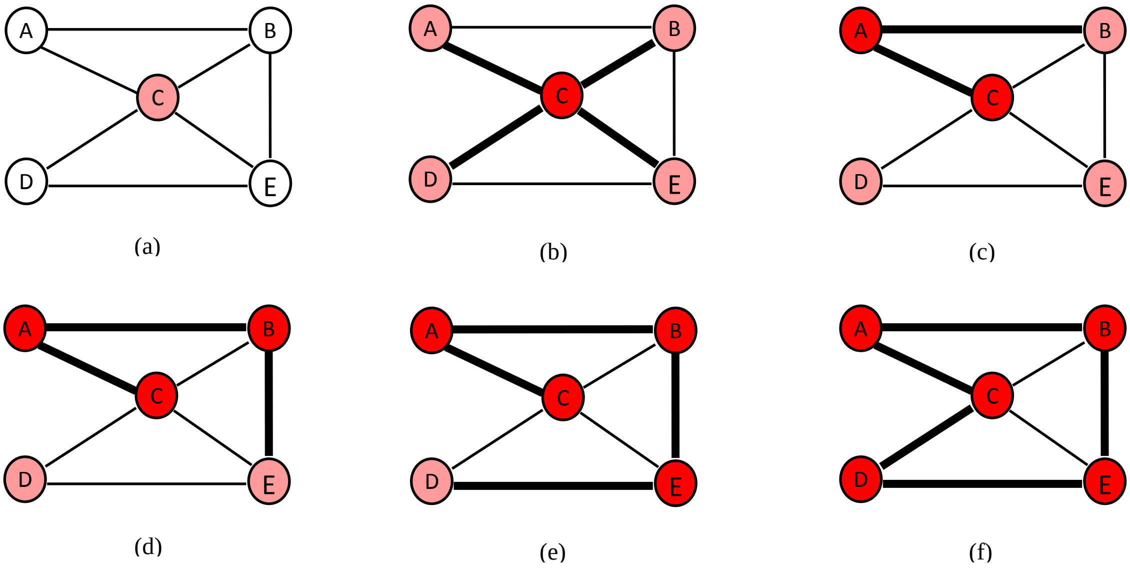


Figure 5. Tree constructed from adjacency matrix for Graph G

4. Do apply the proposed LSC algorithm to find the longest cycle.



(a) (b) (c)

(d) (e) (f)

Figure 6. Implementing Steps of Longest Simple Cycle Problem

5. Join the leaf node of the longest path with root and retrieve the path considering it as the longest cycle in graph.

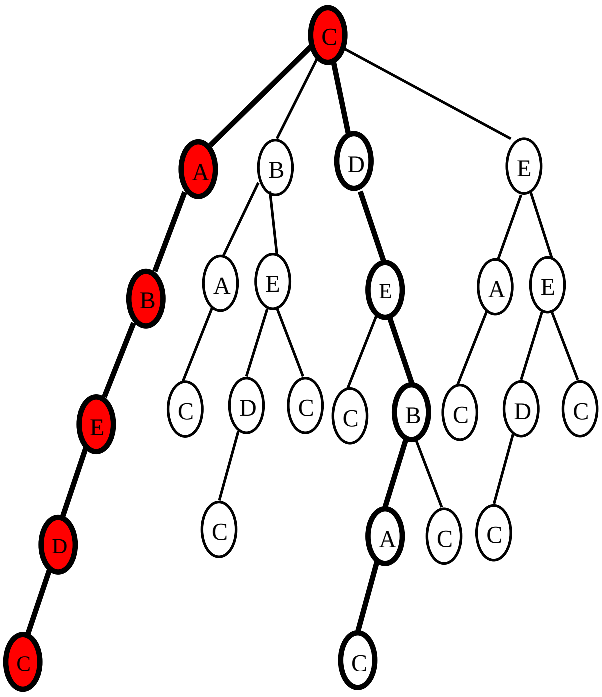


Figure 7. Tree showing the deepest node as the farthest vertex

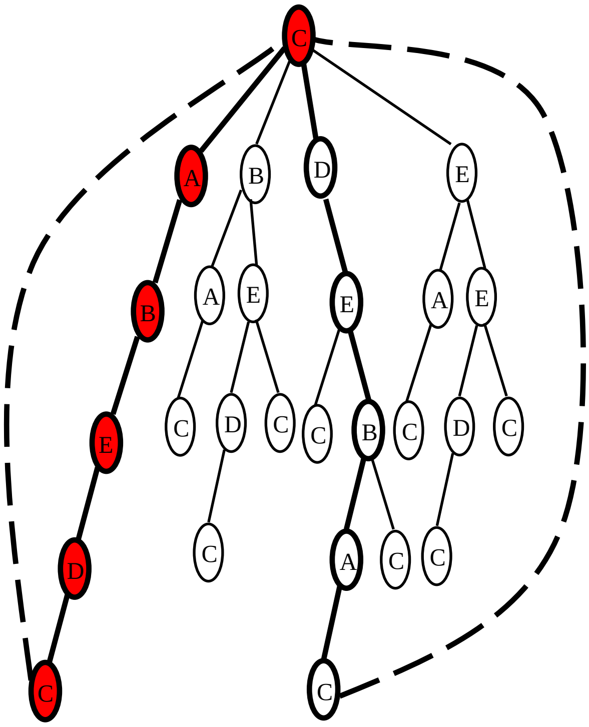


Figure 8. Deepest vertex connected with ROOT vertex to make a cycle

After joining the deepest node with the root the retrieved longest paths are C-A-B-E-D-C and C-D-E-B-A-C.
Now technically both cycles are same but retrieved in reverse directions, So only one cycle is considered to be as longest.

Complexity Computation:
Step 1: compute adjacency Matrix = $|V|^2$
Step 2: compute degree of each vertax = $|V|^2$
Step 3: Retrieve Node with maximum Degree = V
Step 4: Complexity of LSC Algorithm = O(V+E)

Total complexity for the experiment is:
$|V|^2 + |V|^2 + V + (V+E) \rightarrow V^2$

The above computed complexity is for the total result from computing adjacency matrix to finally retrieval of longest path. But the actual complexity of proposed algorithm i.e. Longest Simple Cycle Algorithm is O(V+E).which is quite less than V2.

CONCLUSION:

In this work we help to shed some light on the borderline between P and NP, since the longest simple cycle problem is known to be NP-complete on graphs. It would be interesting to study the complexity of the longest simple cycle problem on distance hereditary and bipartite distance-hereditary graphs, since they admit polynomial solutions for the Hamiltonian path problem, and also since the longest simple cycle problem has been proved to be NP-complete on chordal bipartite graphs, and parity graphs, while it is polynomial on ptolemaic graphs and trees. In this Paper we presented an approximation algorithm for solving the longest simple cycle problem on simple graphs, which find the maximum length cycle in a connected graph (if exist) with average case complexity O(V+E). Various techniques are used to implement the strategy such as adjacency matrix, adjacency List and simple rooted tree. Experimental results are shown in section 4.The work constitute a significant achievement on NP-complete problems to solve it in approximate time.

ACKNOWLEDGMENTS:

The authors would like to thank the referees and reviewers for a very careful reading and many helpful comments and suggestions, which have led to great improvement of the paper.

REFERENCE:

[1] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, CA, 1979.

[2] R. Motwani, Lecture Notes on Approximation Algorithms, Technical Report No. STAN-CS-92-1435, Department of Computer Science, Stanford University, 1992.

[3] W. E de la Vega and G. S. Lueker, Bin packing can be solved within 1 + e in linear time, *Combinatorica*, 1 (1981), 349-355.

[4] N. Karmakar and R. M. Karp, An efficient approximation scheme for the one-dimensional bin packing problem, *Proceedings of the 23rd Annual lEEE Symposium on Foundations of Computer Science*, 1982, pp_ 312-320.

[5] U. Feige, S. Goldwasser, L. Lovfisz, S. Safra, and M. Szegedy, Approximating clique is almost NPcomplete, *Proceedings of the 32nd Annual IEEE Symposium on Foundations of Computer Science*, 1991, pp. 2-12.

[6] A. Blum, Some tools for approximate 3-coloring, *Proceedings of the 31 st Annual IEEE Symposium on Foundations of Computer Science*, 1990, pp. 554-562.161

[7] B. Monien, How to find long paths efficiently, *Annals of Discrete Mathematics*, 25 (1984), 239-254.

[8] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, Proof verification and hardness of approximation problems, *Proceedings of the 33rd Annual llz'ls Symposium on Foundations of Computer Science*, 1992, pp. 14-23.

[9] V. Chvatal, Tough graphs and Hamiltonian circuits, *Discrete Mathematics*, 5 (1973). 215-228.

[10] R. Diestel, Graph Theory, Springer, New York, 2000.

[11] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.

[12] R.W. Bulterman, F.W. van der Sommen, G. Zwaan, T. Verhoeff, A.J.M. van Gasteren, W.H.J. Feijen, On computing a longest path in a tree, Inform. Process. Lett. 81 (2) (2002) 93–96.

[13] R. Uehara, Y. Uno, On computing longest paths in small graph classes, Internat. J. Found. Comput. Sci. 18 (5) (2007) 911–930.

[14] K. Ioannidou, G.B. Mertzios, S. Nikolopoulos, The longest path problem is polynomial on interval graphs, in: Proc. Of 34th Int. Symp. on Mathematical Foundations of Computer Science, vol. 5734, Springer-Verlag, Novy Smokovec, High Tatras, Slovakia, 2009, pp. 403–414.

[15] G.B. Mertzios, D.G. Corneil, A simple polynomial algorithm for the longest path problem on cocomparability graphs, in: Proc. of CoRR, 2010.

[16] G. Gutin, Finding a longest path in a complete multipartite digraph, SIAM J. Discrete Math. 6 (2) (1993) 270–273.

[17] D. Karger, R. Montwani, G.D.S. Ramkumar, On approximating the longest path in a graph, Algorithmica 18 (1) (1997) 82–98.

[18] A. Björklund, T. Husfeldt, Finding a path of superlogarithmic length, SIAM J. Comput. 32 (6) (2003) 1395–1402.

[19] T. Feder, R. Motwani, Finding large cycles in Hamiltonian graphs, in: Proc. 16th Annual ACM–SIAM Symp. on Discrete Algorithms, SODA, ACM, 2005, pp. 166–175.

[20] H.N. Gabow, Finding paths and cycles of superpolylogarithmic length, in: Proc. 36th Annual ACM Symp. on Theory of Computing, STOC, ACM, 2004, pp. 407–416.

[21] Z. Zhang, H. Li, Algorithms for long paths in graphs, Theoretical Computer Science 377 (1–3) (2007) 25–34

[22] F. Luccio, C. Mugnia, Hamiltonian paths on a rectangular chessboard, in: Proc. 16th Annual Allerton Conference, 1978, pp. 161–173.

[23] A. Itai, C. Papadimitriou, J. Szwarcfiter, Hamiltonian paths in grid graphs, SIAM J. Comput. 11 (4) (1982) 676–686.

[24] S.D. Chen, H. Shen, R. Topor, An efficient algorithm for constructing Hamiltonian paths in meshes, J. Parallel Comput. 28 (9) (2002) 1293–1305.

[25] Yancai Zhaoa, Liying Kang , Moo Young Sohn " The algorithmic complexity of mixed domination in graphs" *Theoretical Computer Science 412 (2011) 2387–2392.*

[26] Christian Glaßer , Christian Reitwießnera , Victor Selivanov "The shrinking property for NP and coNP" *Theoretical Computer Science 412 (2011) 853–864.*

[27] Ferdinando Cicalese , Tobias Jacobs , Eduardo Laber , Marco Molinaro "On the complexity of searching in trees and partially ordered structures" *Theoretical Computer Science 412 (2011) 6879–6896.*

[28] James K. Lana, Gerard Jennhwa Chang "On the mixed domination problem in graphs" *Theoretical Computer Science 476 (2013) 84–93.*

[29] Van Bang Le, Ragnar Nevries "Complexity and algorithms for recognizing polar and monopolar graphs"*Theoretical Computer Science 528 (2014) 1–11.*

[30] Lei Chena , Weiming Zeng , Changhong Lub "NP-completeness and APX-completeness of restrained domination in graphs" *Theoretical Computer Science 448 (2012) 1–8.*

[31] Kirishima T, Shiono Y, Sugimoto F, Kato C, Yaku T Optimality and Complexity for Drawing Problems of Tree-Structured Diagrams *14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2013, pp. 484 – 489, Honolulu, HI*

SESSION

WIRELESS NETWORKS AND ENERGY EFFICIENCY ISSUES

Chair(s)

TBA

A New Power Saving Routing Algorithm For Delay Tolerant Networks

A. Omidvar, and Dr. K. Mohammadi
Electrical Engineering Department, Iran University of Science and Technology, Tehran, Iran

Abstract - *Delay tolerant networks (DTNs) are wireless networks with intermittent connections. Node mobility, limited radio range and power cause intermittency. These networks have different applications such as space searching, wild life tracking, military battlefields and etc. Conventional routing approaches are not practical in DTN because of intermittency. So, routing is a challenging matter. This paper proposes a new routing method called geographic destination (GD). It is a power saving method which tries to find a neighbor which is geographically closest to destination. Among the neighbors, one closer to destination will receive a number of copies from the source. This approach continues until the message reaches destination. After that, extra copies are excluded. This method reduces transitions for conveying data to decrease energy usage. Message drop and overhead have also decreased. Comparing GD to well known protocols such as PROPHET and ER, proves GD success in power saving and preserving network resources.*

Keywords: Delay tolerant network (DTN); energy; message drop; average hops.

1 Introduction

DTNs are also known as disruption tolerant networks. Intermittent connections, high error rate, long delays, variable data rates and etc are general characteristic of DTNs [3]. Thus, store-carry-and-forward mechanism (SCF) is used in these networks [4,5,6]. SCF stores the data while next hop is not available for forwarding message. So, it carries the message while moving and forwards it to node which has better opportunity to transmit message to destination.

DTNs have different applications such as wild life tracking [1], deep space searching [7], vehicular networks [8], and government services [9].

Due to DTN intermittency, conventional routing protocols in mobile ad hoc networks (MANETs) cannot be applied to DTNs. Energy, buffer, message delivery ratio, message drop, overhead and etc are important factors in designing routing approaches.

Message delivery ratio shows the ratio of delivered messages to all produced messages.

Up to now, different studies have been done on designing routing protocols. These routing protocols are based on SCF to overcome long delay, disconnections, queuing delays and limited resources. These methods can be classified from different viewpoints [10]. A very general categorization can be flooding and non-flooding.

Flooding protocols give a copy of the message to every node they encounter which wastes node resources. Other approaches try to limit number of copies spread in networks. Since some parameters maintenances are in contrast, routing protocols should consider the desired application.

This paper proposes a new approach, which is called geographic destination (GD), to limit energy usage.

This paper is organized as follows: Section (II) reviews related works and section (III) discusses the proposed method. GD evaluation is discussed in section (IV). Section (V) concludes the paper.

2 Related works

Recently there have been many researches finding routing protocols for DTNs. Limited sources make researches consider different factors in designing routing methods such as energy, buffer, message delivery ratio, message drop number, overhead, and etc. Regarding these conditions, existing approaches can be categorized in 3 main categories [10,11]: flooding approaches [12,13], prediction based methods [14,15,16,17] and forwarding [18].

Flooding methods try to increase message delivery ratio by forwarding messages to encountered nodes. This helps to decrease message delivery delay while increasing overhead. Epidemic routing (ER) is one of well known flooding approaches [12]. In addition to flooding approaches, controlled flooding methods try to reduce overhead. Spray and Wait (SAW) [13] is one of these algorithms. It sprays messages to connected nodes. It waits for these connected nodes to eventually deliver message to destination.

Prediction based approaches use history of node encounters to find the appropriate node. These methods try to increase message delivery, by finding nodes which have recently encountered destination. Probabilistic Routing Protocol using History of Encounters and Transitivity (PROPHET) [14], MaxProp [15], Spray and Focus [16] and etc are good instances for this category. In PROPHET, vectors indicating the probability of delivering its message to destination for each node are exchanged between nodes when they met each other.

MaxProp considers the meeting chance between nodes which is not necessary the destination.

BUBBLE [17] and SimBet [18] use information about social community structures, contact history and communication patterns to find appropriate nodes for forwarding.

Forwarding category, forwards a single message through a carefully chosen path. MEED [19], direct transmission, randomized routing, and utility based routing [20] are good example of this approach.

Based on this works, this paper proposes an approach based on reducing node energy. The method finds the geographic distance of each connected node to destination. Finding the minimum distance helps to decrease number of transitions. This will be discussed in next section.

3 Proposed method

Network resources such as buffer, energy, and etc are limited. It makes researchers find ways which preserve sources while conveying information. GD tries to reduce energy usage of nodes by decreasing number of transitions in the network.

GD calculates the distance between each message destination and connected host neighbors. If the source node has geographic position of (x_0, y_0) and destination node is placed in (x_f, y_f), GD finds the neighbor which has minimum distance to destination. Each message is sent to closest neighbor to destination. This algorithm iterates until each message reaches the destination. In order to improve GD, some copies of each message are sent to connected node which is nearest to destination. Number of copies can be fixed or dynamically chosen regarding message delivery ratio and overhead. After delivering the message to its destination, extra copies are deleted. Simulation proves GD success in power saving in comparison to ER and PROPHET.

Each node has limited energy. Message sending, receiving and relaying uses energy. GD helps to reduce energy consumption by fewer transitions. In order to better understanding of matter, some statements have to be explained.

Dropped messages refer to messages drop from node buffers.

Delivered messages are those successfully delivered to message destination.

Overhead ratio is found by (1):

$$\frac{(\text{Number of Relayed Messages - Number of Delivered Messages})}{\text{Number of Delivered Messages}} \qquad (1)$$

Hop counts show average hop count between source and destination.

GD uses greedy forwarding to reduce number of transmissions and replicated messages.

4 GD evaluation

Simulations are done in opportunistic network simulator environment (ONE) which is based on java. It was first developed in Helsinki University of Finland [21-23]. Simulations have been done for 40 times and the average results have been considered. Simulation settings are shown in Table 1. Mobility model used in simulations is shortest path map based movement (SPMBM). Mobility model shows node movement on the predefined maps of Helsinki downtown area.

In order to evaluate GD, two scenarios were implemented. In first one, buffer size varies from 10 k to 100 k. Message time to live (TTL) is set to 300 s. As it can be seen in Fig. 1, energy level of nodes in GD is greater than ER and PROPHET. Due to limited energy of nodes, message replication, relaying and etc waste node battery. Since the number of transitions and message copies are reduced in GD, energy usage has reduced. GD tries to find nearest nodes to destination for forwarding messages, and also reduces number of message replication. So, number of message drop decreases in GD. As shown in Fig.2, GD has reduced message drop by 96% compared to ER and PROPHET.

TABLE I. SIMULATION SETTINGS

Network Area	$5000\times5000m^2$
Number of Nodes	100
Simulation duration	43200 s= 12 h
Message size	100B-200B
Mobility Model	Shortest Path Map Based Movement (SPMBM)

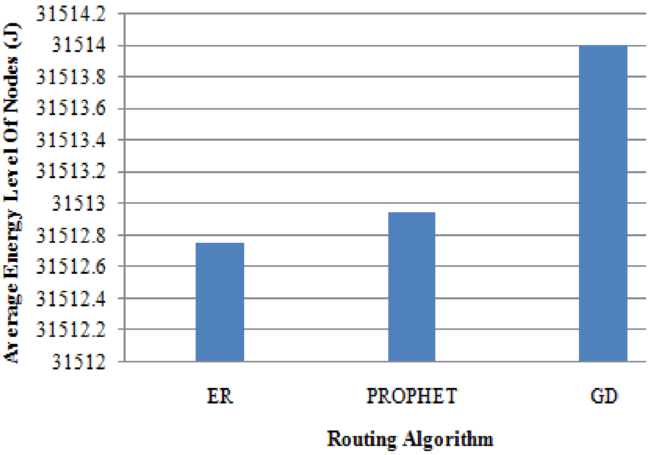


Fig. 1. Average energy level of nodes comparison among ER, PROPHET, and GD regarding buffer size variation

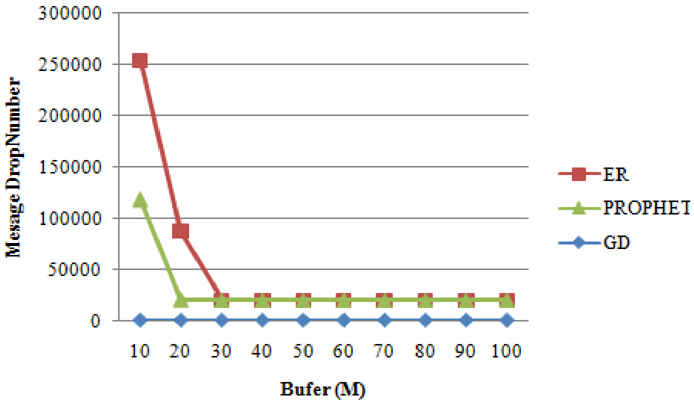


Fig. 2. Message drop number comparison in ER, PROPHET and GD regarding buffer size variation

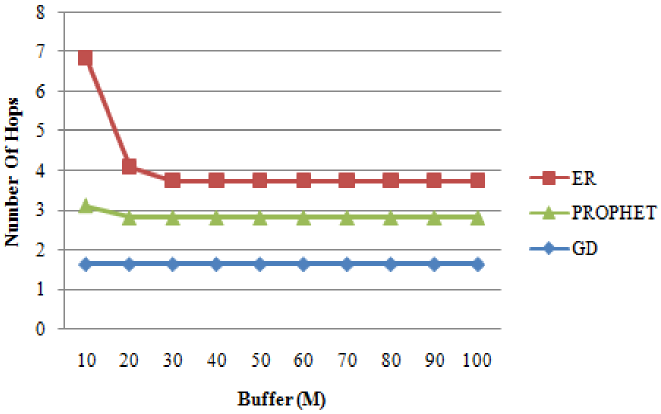


Fig. 3. Average number of hops messages pass to reach destination regarding Buffer change

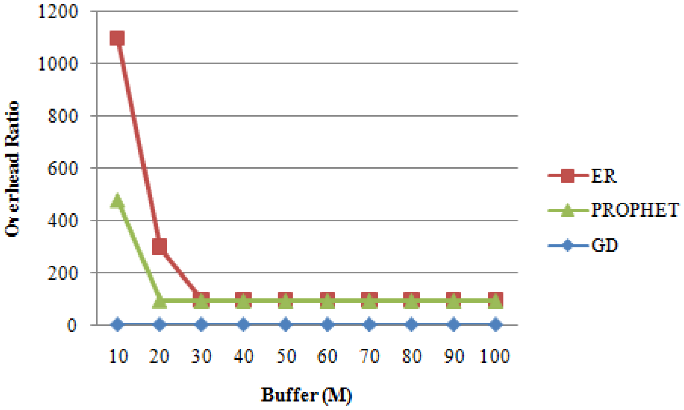


Fig. 4. Overhead ratio comparison among ER, PROPHET and GD regarding Buffer size variation

Finding nearest hops to destination for forwarding messages also helps to reduce number of hops message has to pass to reach destination. GD, as shown in Fig. 3, has reduced number of hops by 74% compared to ER and 66% compared to PROPHET.

As shown in Fig. 4, GD overhead has reduced 97% compared to ER and PROPHET. Message delivery ratio has decreased because number of copies scattered in the network has reduced. Fig. 5 shows message delivery ratio. ER and PROPHET which have more delivery ratio, has greater overhead and message drop compared to GD regarding Fig. 3 and Fig. 4.

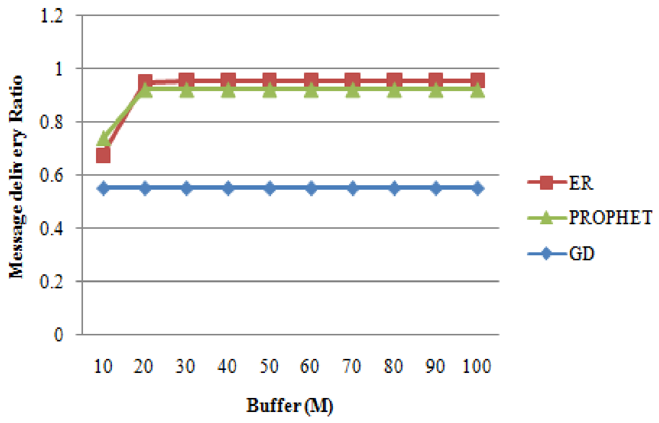


Fig. 5. Message delivery ratio comparison among ER, PROPHET and GD regarding Buffer size variation

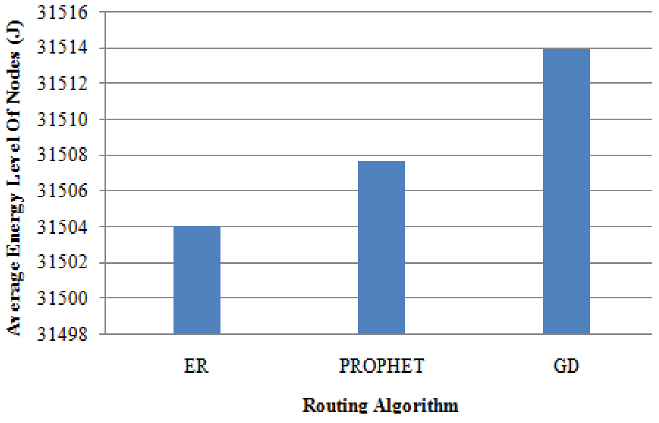


Fig. 6. Average energy level of nodes comparison among ER, PROPHET, and GD regarding TTL variation

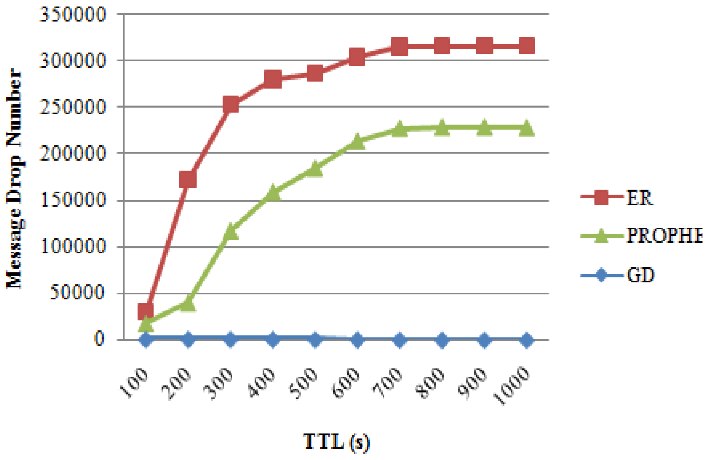


Fig. 7. Message drop number comparison in ER, PROPHET and GD regarding TTL variation

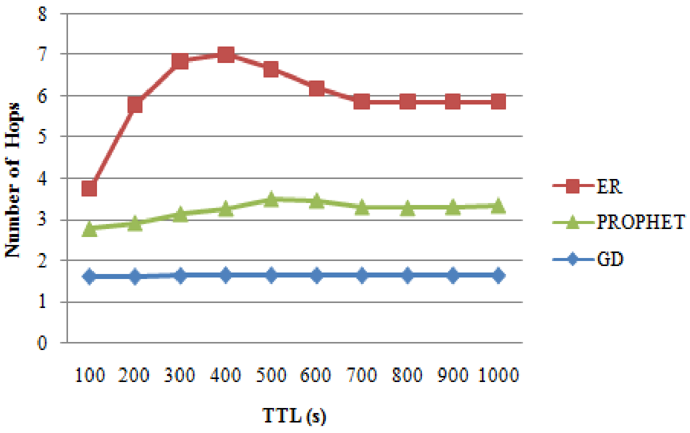


Fig. 8. Average number of hops messages pass to reach destination comparison in ER, PROPHET and GD regarding TTL variation

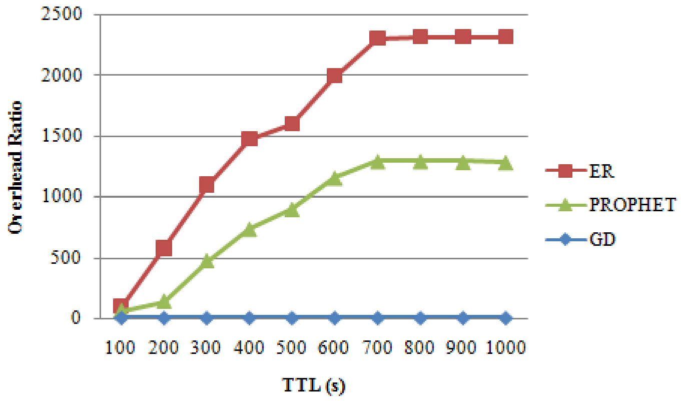


Fig. 9. Overhead Ratio comparison in ER, PROPHET and GD regarding TTL variation

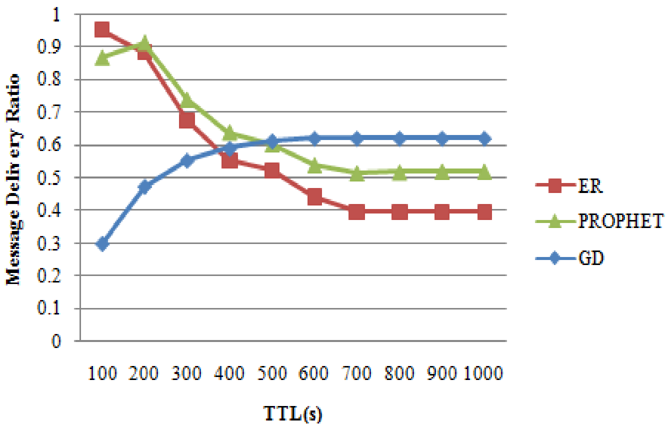


Fig. 10. Message delivery ratio comparison in ER, PROPHET and GD regarding TTL variation

In second scenario, buffer size is set to 10 k and TTL is changed from 100 s to 1000 s. As it can be observed in Fig. 6, average energy level of nodes in GD is greater than ER and PROPHET. Since message replication and forwarding is controlled, nodes energy is preserved. According to Fig. 7, message drop number in GD is 0.001 of ER and 0.002 of PROPHET. Limited number of message copies helps to reduce message drop from buffers.

In addition to reducing message drop and energy, average hops passed by message have also reduced. As it can be seen in Fig. 8, average hop has reduced by 53% compared to ER and 48% compared to PROPHET. GD message overhead ratio is 0.001 of ER and 0.002 of PROPHET as shown in Fig. 9.

As shown in Fig. 10, for TTL less than 500 s, GD delivery ratio is 18% below ER and PROPHET. After 500 s, GD message delivery ratio is 37% better than ER and 25% greater than PROPHET. According to simulations, GD has successfully reduced energy usage, message drop number, overhead ratio and average hop count. Message delivery ratio is not disturbed considerably.

These characteristics suggest GD as a powerful approach.

5 Conclusions

Delay tolerant networks (DTNs) are wireless networks which suffer intermittency. This makes routing a challenging matter. This paper proposes a method named GD which helps to save node energy while reducing message drop number. GD calculates distance between connected neighbor geographic location and message destination for each message. It forwards messages to nodes which are closest to destination. After delivering the message to destination, extra copies are deleted. This operation greatly reduces energy consumption by reducing number of transmissions and copies. GD also reduces message drop number and average hops. Comparing GD to ER and PROPHET, GD shows success over these famous methods.

Future studies will consider using evolutionary algorithm (EA) in finding appropriate nodes for forwarding messages.

6 References

[1] K. Fall, “A delay-tolerant network architecture for challenged internets,” in *Proc. ACM SIGCOMM*, pp. 27-34, 2003.

[2] S. Jain, K. Fall, R. Patra, “Routing in a delay tolerant network,” in *Proc. ACM SIGCOMMS*, 2004.

[3] V. Rodoplu, T. H. Meng, “Core capacity region of energy-limited, delay-tolerant wireless networks,” *IEEE Trans. Wireless Commun.* vol. 6, no.5, pp. 1844-1853, 2007.

[4] Z. Zhang, “Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: Overview and challenges,” *IEEE Commun. Surveys,* vol. 8, no. 1, pp. 24–37, 2006.

[5] T. Spyropoulos, T. R. N. B. Rais, T. Turletti, K. Obraczka, A. Vasilakos, “Routing for disruption tolerant networks: Taxonomy and design,” *Wireless Network*, vol. 16, no. 8, pp. 2349–2370, 2010.

[6] K. Fall, S. Farrell, “DTN: An architectural retrospective,” *IEEE J. Select. Areas Commun*. vol. 26, no. 5, pp. 828–836, 2008.

[7] S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, “Delay-tolerant networking: an approach to interplanetary internet,” *IEEE Commun. Mag*. vol. 41, pp. 128-136, 2003.

[8] J. Lebrun, C. N. Chuah, D. Ghosal, M. Zhang, “Knowledge-based opportunistic forwarding in vehicular wireless ad hoc networks,” in *Proc. IEEE Vehicular Technology Conference*, pp. 2289-2293, 2005.

[9] E. Brewer, M. Demmer, B. Du, M. Ho, M. Kam, S. Nedevschi, J. Pal, R. Patra, S. Surana, k. Fall, “The case for technology in developing regions,” *IEEE Computer*, vol. 38, no. 6, pp. 25-38, 2005.

[10] W. Moreira, P.Mendes, “Survey on opportunistic routing for delay/disruption tolerant networks,” *SITI Tech. Rep. SITI-TR-11-02*, 2010.

[11] D. Hua, X. Du, Y. Qian, S. Yan, “A dtn routing protocol based on hierarchy forwarding and cluster control,” in *Proc. International Conference on Computational Intelligence and Security*, pp. 397-401, 2009.

[12] A.Vahdat, D. Becker, “Epidemic routing for partially-connected ad hoc networks,” Uni. Of Duke, Tech. Rep. CS-200006, 2000.

[13] T. Spyropoulos, K. Psounis, C. S. Raghavendra, “Spray and wait: An efficient routing scheme for intermittently connected mobile networks,” in *Proc. ACM SIGCOMM WDTN*, pp. 252–259, 2005.

[14] A. Lindgren, A. Doria, O. Schelen, “Probabilistic routing in intermittently connected networks,” in *Proc. ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 7, no. 3, pp. 19–20, 2003.

[15] J. Burgess, B. Gallagher, D. Jensen, B. N. Levine, “MaxProp: Routing for vehicle-Based disruption-tolerant networks,” in *Proc. IEEE INFOCOM*, 2006.

[16] T. Spyropoulos, K. Psounis, C. S. Raghavendra, “Spray and focus: efficient mobility-assisted routing for heterogeneous and correlated mobility,” in *Proc. IEEE PerCom Workshop on Intermittently Connected Mobile Ad Hoc Networks*, pp. 79-85, 2007.

[17] P. Hui, J. Crowcroft, E. Yoneki, “BUBBLE Rap: Social-based forwarding in delay tolerant networks,” in *Proc. 9th ACM international symposium on Mobile ad hoc networking and computing*, 2008.

[18] E. Daly, M. Haahr, “Social network analysis for information flow in disconnected delay tolerant MANETs,” *IEEE Trans. Mobile Comput*. vol. 8, no. 5, pp. 606–621, 2009.

[19] E. Jones, L. Lily, J. K. Schmidke, P. Ward, “Practical routing in delay-tolerant networks,” *IEEE Trans. Mobile Comput.* Vol. 6, no. 8, pp. 943–959, 2007.

[20] T. Spyropoulos, K. Psounis, C. S. Raghavendra, “Efficient routing in intermittently connected mobile networks: The single-copy case,” *IEEE/ACM Trans. Networking*, vol. 16, no. 1, pp. 63–76, 2008.

[21] The Opportunistic Network Environment (ONE) Simulator. http://www.netlab.tkk.fi/tutkimus/dtn/theone/.

[22] A. Keränen, J. Ott, T. Kärkkäinen, “The ONE simulator for DTN protocol evaluation,” *in Proc. SIMUTools’09: 2nd International Conference on Simulation Tools and Techniques,* Rome, 2009.

[23] A. Keränen, T. Kärkkäinen, J. Ott, “Simulating mobility and DTNs with the ONE,” *J. Commun*. vol. 5, no. 2, pp. 92-105, 2010.

Energy-efficient Heuristics for Multi Hop Routing in People-centric Environments

Antonio Oliveira-Jr, Rosario Ribeiro, Tercio Filho, Dalton Matsuo, Marcos Batista, Claudio Souza
University Federal of Goias (UFG), Brazil
Email: {antoniojr, rosarioribeiro, tercioas, dalton_tavares, marcos.batista, clsouza}@ufg.br
José-Ricardo Ribeiro
Centro Universitário de Goiás - Uni-ANHANGUERA, Brazil
Email: jricardo@anhanguera.edu.br

***Abstract*—This paper explore different heuristics that may be applied to provide a node-based cost for energy-aware multihop routing for wireless environments which integrate heterogeneous devices that are carried or owned by Internet end-users. We analyzed based on simulations of the different heuristics when applied to distance-vector approach, namely the Ad Hoc On Demand Distance Vector (AODV) routing protocol.**

***Index Terms*—Multihop routing; energy-efficiency; user-centric networks; AODV.**

I. Introduction

People-centric wireless environments integrate a highly dynamic behavior of mobile nodes, in particular of nodes that are owned or carried by humans. Examples of such environments and dynamism are the need to autonomously start a network based on end-user devices after a disaster of some nature (e.g., disaster networks) or even the need to assist emerging markets in remote areas, sometimes highly populated. Such people-centric environments attain specific requirements, of which *energy efficiency* is one of them.

Albeit being spontaneously deployed, people-centric environments rely on traditional multihop routing approaches. Multihop routing has been extensively analyzed and optimized in terms of resource management, but in terms of energy efficiency there is a lack of a thorough analysis in particular in what concerns people-centric environments such as *User-provided Networks (UPNs)* or *Mobile Ad-hoc Networks (MANETS)*. On the other hand, there is considerable related work in the fields of energy efficiency and energy awareness for sensor networks.

Even though it is relevant to consider the results achieved in such networks, there are specific requirements of people-centric environments which makes energy awareness and efficiency problems not trivial to be solved. Firstly, nodes in people-centric networks are expected to be heterogeneous in terms of resources such as battery capacity. Secondly, such nodes exhibit frequent movement and are also expected to frequently join and leave a network. We refer heterogeneous for the different nodes regarding mobile devices, such as the technology itself (e.g., laptop, smart phone), battery capacity, energy consumption, energy parameters and processing. Regarding movement, we consider a social mobility model which the parameters, such as frequency of movements, are characterized by the mobility pattern.

The main goal of this work is focused on making current multihop routing approaches, i.e., *shortest-path based*, adequate for people-centric environments. In such environments there are several requirements to be met in terms of energy awareness, by exploring new routing metrics that take into consideration the state of a node, i.e., *node-based perspective*, or not only the originating node's perspective, but also the potential of successors, i.e., *link-based perspective*, in terms of energy awareness. Our expectations are to optimize network utilization by optimizing the energy-awareness of multi hop routing approaches.

In this paper we explore different energy-efficient heuristics as node-based cost to apply in multi hop routing with heterogeneous mobile devices. We evaluate based on simulations the heuristics in a multi hop distance vector approach, namely the AODV [1] routing protocol. We show the heuristics applyed as a cost function improve the network lifetime without penalizing the end to end delay and throughput.

The rest of this paper is organized as follows. Section II describes selected related work focused on multihop energy efficiency. Section IV presents the notions, parameters and the current energy-aware routing metrics for multi hop routing. Section V is our proposed heuristics with discussions regarding network lifetime. Then, in section VI, we present the performance evaluation based on simulations with statistically rigorous results. Conclusions and future work are presented in section VII.

II. Related Work

There are few approaches [2], [3], [4] that have surveyed multihop proposals focused on energy efficiency, considering both the energy spent when nodes are engaged in active communication or inactive communication (e.g., in idle mode). Such work has as underlying scenarios homogeneous environments, and many proposals combine a different energy-aware metrics to maximize the network lifetime.

Attempting to make multihop routing adaptive, some proposals [5], [6], [7] have explored new metrics having in mind different types of optimization, e.g., reduction of energy spent across a path or avoiding nodes with low residual energy, on the global network.

C. K. Toh provides a relevant overview [8] of different routing properties to consider in multihop routing that one of them is efficient utilization of battery capacity. In this work, the author also addresses the performance of power efficiency in ad-hoc mobile networks by analyzing four approaches which have as common goal to select an optimal path, being the optimum the minimization of the total power required on the network (across all nodes) and also the maximization of the lifetime of all nodes in the network.

The cost function of MRPC (Maximum Residual Packet Capacity) protocol [9] comprises a node-parameter (battery power of node) and a link-parameter (packet transmission energy in a link) across the link between nodes. MRPC identifies the capacity of a node not just by its residual battery energy, but also by the expected energy spent in reliably forwarding a packet over a specific link. However, such formulation better captures scenarios where link transmission costs depend on physical distances between nodes and the link error rates, which does not consider energy as a prime metric.

Considering power constraint as a metric, Senouci et. al. [10] propose three routing algorithms unless the shortest path routing. However, the routing algorithms was devised as three different routing protocols based on AODV modifying the routing process. Our work consider a node-based cost as a metric to use any distance vector or link state multihop routing protocols.

A recent work [11] proposed a multi-objective approach which consider three routing metrics (delay, energy and link lifetime) in a prediction way. The methods are predicting queuing delay and energy consumption, and predicting residual link lifetime using a heuristic of the distributions of the link lifetimes. However, the energy resource is combined with another metrics that is hard to find a trade off considering the energy-efficient routing.

The Working Group ROLL (Routing Over Low Power and Lossy Networks) of the IETF (Internet Engineering Task Force) have working on routing metrics to consider in this type of networks, which energy-aware is one of them [12]. A node energy object is used to provide information related to node energy and may be used as a metric or as constraint.

We emphasize that our proposed heuristics is to consider routing metrics that can be coupled to any multihop routing protocol, i.e., distance vector or link state approaches, to provide multihop routing with better energy-awareness.

III. Energy awareness in People-centric Networks

People-centric Networks (PCN) integrate the end-user connected to the Internet by means of a variety of broadband access technologies, which the final segment is provided by a number of short-range technologies, among which Wireless Fidelity (Wi-Fi) is a solution.

In this scenario, the end-user (or a community of end-users) is a micro-provider in the sense that he/she shares his/her subscribed broadband Internet access based on some incentive scheme. The way people interact and move is a behavior that we can root on social network theory, due to the fact that humans are not only carriers but also the decision makers for the operation of nodes that form the topology.

We provide an example of a generic scenario, where groups of mobile nodes are depicted by a dotted line. Within each group, nodes may move in an independent way according to human movement behavior (social mobility). Furthermore, nodes may also move in groups, also mimicking human social behavior. Groups have a spatial-temporal correlation, e.g., a group at an instant in time may dissolve in a different instant in time and space. The illustrated nodes can be either static or mobile. In addition, nodes may behave as a regular node, or a micro-provider node. A micro-provider node is basically a node that provides Internet access to other nodes. It should be noticed that in contrast to the notion of gateway in MANETs, a micro-provider may simply relay Internet access from a gateway to a group of nodes. In addition, a micro-provider node may be completely mobile. Therefore, the topology shows a highly dynamic behavior, where not only links are bound to frequent changes, but also where the nodes that provide Internet access can also change on-the-fly, e.g., due to congestion of the micro-provider(s) in the group, due to better network conditions. Adding to the variability due to node movement, for instance, another key aspect is that some devices are multimedia capable with strong limitations in terms of energy capabilities.

IV. Energy Awareness in Multi Hop Routing

A *node* i represents a wireless heterogeneous device with a single or with multiple network interfaces. Edges interconnecting nodes are represented as *links* (i, j) with a cost which is a measure of energy expenditure. Such energy expenditure can be obtained from a single node, a link, or network utilization perspective. From a single node perspective, there are three main modes of operation which depend on the node status. A node is in *Transmit mode* when transmitting information. Hence, *Transmit Power (Tx Power)* for a node corresponds to the amount of energy (in Joules) spent when the node transmits a unit (bit) of information. A node is in *Receive mode* if it is receiving data. Hence, *Reception Power (Rx Power)* for a node corresponds to the amount of energy (in Joules) spent when the node receives a unit (bit) of information. Particularly for the case of 802.11, there are two additional states a node may be at. When not receiving or transmitting, the node is still listening the shared medium (*overhearing*) and is said to be in *Idle mode*. When the node is not overhearing, then it is said to be in *Sleep mode*. In this mode, no communication is possible but there is still a low-power consumption.

Another relevant parameter to consider from an energy-awareness perspective is a node's degree, N_i as the surrounding nodes impact the transmission channel heavily, as well as on energy consumption. We use the node degree definition where N_i corresponds to the amount of neighbors that a node i has at an instant in time. More relevant than the number of neighbors, is the history of variation of N_i through time.

The main energy-aware metrics for people-centric environments are the residual energy and drain rate. The *Residual Energy (RE)* of a node i, RE_i [13] is defined as the amount of energy units that the battery of node i has at an instant in time. The *Drain Rate (DR)* of a node i, DR_i [14] is defined as the amount of energy being spent by node i through time, due to the activities the node is performing. $DR(i)$, can be computed by applying an *Exponential Weighted Moving Average (EWMA)*. The DR alone simply provides a way to measure energy being spent by nodes.

For heterogeneous environments, a combination of the DR with the RE of a node is significant to capture both the expenditure and the resources still available. Such combination can be provided in several ways. Kim et. al. consider also the ratio between RE and DR as C_i defined as the node lifetime. However, all of these metrics still is not sufficient to people-centric environment which there are mobile devices in different capacities and states affecting path robustness.

V. Proposed Heuristics

This section provides an overview on the heuristics that we are currently testing, to provide multihop routing with better energy-awareness. As mentioned our proposal is to consider routing metrics that can be coupled to any multihop routing protocol. In other words, the proposed heuristics are not expected to be tied to a specific protocol.

A. Heuristic 1: Energy-awareness Ranking of Node Based on Idle Times

In this first heuristic we take into consideration the periods over time where i is in idle mode. In other words, over time we estimate how much time of its lifetime has node i been in idle mode, to then provide an estimate on a potential behavior in the future, as this will for sure impact the node's lifetime. Such periods are the ones the most expensive to i in terms of energy, and in those periods, the node degree becomes highly relevant as the more nodes surround node i, the worst the energy expenditure of i. So we consider the total period in idle time, t_{idle} over the past period together with the estimated lifetime of the node, as provided in equation 1.

$$E_{1_i} = \frac{t_{idle}}{T + C_i},\ E_{1_i} \in [0, 1] \tag{1}$$

E_1 is therefore a node weight which provides a ranking in terms of the node robustness, from an energy perspective, and having as goal to optimize the lifetime.

B. Heuristic 2: Energy-awareness Ranking of Node Based on Idle Times and Node Degree History

This second heuristic considers also the potential impact that the node degree may have in the energy expenditure of a node. Surrounding nodes impact the conditions of the wireless media and as such, the node degree history, in particular the variability of the node degree is one additional aspect that may impact node lifetime. Hence, still following a simplistic approach, we consider ways to combine the history of the node degree with E1, having derived as a first approach E2, provided in equation 2.

$$E_{2_i} = \frac{t_{idle}}{T + C_i} * N_i', \tag{2}$$

For instance, let us assume that node i has, at a specific instant in time, a lifetime that seems to be long. If the node has an history of a low number of neighbors as happens in the case of less dense networks, then in contrast to a node that has the same lifetime but a larger number of nodes around, we can decide on which node to opt. Deciding for a node that has a higher node degree implies having more alternate paths being the flip-side to this the possibility of seeing an abrupt change in the time left until the node exhausts energy. Opting for a node with a lower node degree may provide more robustness at the cost of having less alternate paths. Depending on the situation of the nodes around (e.g. movement; short lifetimes), there is a variability associated.

The node degree history, N_i', is provided by an *Exponential Moving Average (EMA)* as provided in equation 3.

$$N_i' = \alpha \times N_{i_{t-1}} + (1 - \alpha) \times N_i' \tag{3}$$

C. Discussion

The ranking of a node considering the different heuristics can be seen from an energy-wise point of view on the global network and the impact of the heuristic considering the slope variations of the cost functions as shown in Table I.

A ranking of the node is based on the values of t_{idle} and C_i for the heuristic E1. In case of high idle time and high lifetime is a good candidate to opt the nodes on the path. On the other hand, nodes with low idle time and low lifetime is a node that can be avoided to select on the path. Then, we want to favor the inactive nodes with long lifetime since they are spend energy but not too much like an active node.

Table I
Ranking the node cost

t_{idle}	C_i	Ranking E_{1_i}	N_i'	Ranking E_{2_i}
high	high	candidate	high	low potential
			low	candidate
high	low	low potential	high	low potential
			low	good potential
low	high	good potential	high	good potential
			low	candidate
low	low	avoid	high	avoid
			low	low avoid

For the case of heuristic E2, which we consider the node degree history, the ranking of a node is based on both the values of E1 and N_i'. In this case, a node with high idle time, high lifetime and low node degree history is the best candidate to opt on the path. Depending of the lifetime of node, a node with low idle time and low node degree is also a good candidate, while a node with low lifetime, low idle time and high node degree is a node ranking that be avoided to be selected. We want to favor nodes with low node degree, but finding a balance having less alternate paths.

Next sections presents the performance evaluation of the proposed heuristics applying as cost function energy-aware metrics based on simulations with different energy metrics when applied to distance vector approach.

VI. Performance Evaluation

This section covers the tools and scenarios to evaluate the proposed heuristics. The simulator considered is the NS-2 simulator (version 2.34) [15], a discrete event networking simulator. We have used a realistic physical layer including a radio propagation model, radio network interfaces and the IEEE 802.11 MAC protocol using the *Distributed Coordination Function (DCF)*. To simulate adequately the MAC layer we have considered the 802.11g parameters, namely, a data rate of 54 Mbps and a radio range of 250 meters.

We simulate a static network with 25 nodes distributed in a flat grid topology. For the traffic models, we use CBR sources as VoIP standard with the source-destination pairs randomly chosen over the scenario. There are 5, 10 and 15 connections pairs to represent different degrees of traffic load in different sets of simulations.

The nodes are static and have been simulated to hold different energy characteristics, in order to represent heterogeneous portable devices, e.g. laptop, PDA, a device with continuous power.

A. Routing Mechanism: AODV

Our heuristics are being developed to be applicable to any shortest-path based protocol. In this work we evaluate the heuristics with AODV protocol as distance vector approach. In this section we explain how we have implemented the routing protocol, what has been changed to accommodate our heuristics.

We have considered the native AODV, in NS-2 simulator referenced in this work as *AODV-native*. Native AODV considers hop count as the metric to compute a shortest-path. Moreover, the original C_i has been developed to be applied to DSR [16]. The original specification of C_i therefore selects a best path based on a min-max approach, where the best path is the one that has the lowest bottleneck in terms of energy. So, we adapt the protocol to select the path in a min-max way as the original specification of the C_i . The modifications is only regarding using the energy metric instead of hop count by change the control messages of the AODV. We refer as *AODV-minmax-Ci* for this implementation.

To be as realistic as possible, we consider the native AODV with our proposed heuristic which we call *AODV-SP-E1* and *AODV-SP-E2* to represent a shortest path (SP) node cost applying our heuristics as a metric.

B. Simulation Results

The heuristics are being analyzed from a perspective where the purpose is to increase network lifetime. As such, the results that are being extracted, are: (i) Average end-to-end (e2e) delay, (ii) Average throughput and (iii) Average aggregate node lifetime.

To generate statistical sound results to attend the credibility aspects on simulations analysis, we are currently using the Akaroa2 [17] tool which can be integrated to NS-2 to provide credible and efficient simulations. Akaroa2 can assist us in adequately devising results to extract statistically independent results, which it provides heuristics to detect the beginning of the steady-state and eliminates the correlation by means of the spectral analysis method. The simulations were carried out with infinite time horizon, where for each run, there are about 2500 to 30000 samples and a confidence interval of 95%.

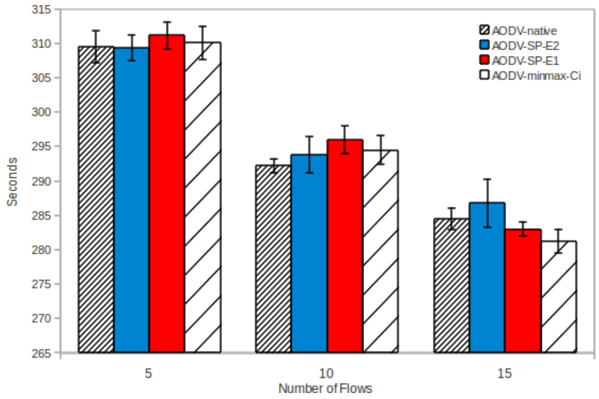


Figure 1. Average Aggregate Node Lifetime

Figure VI-B show the average aggregate node lifetime, i.e. average network lifetime, of the E1 and E2 heuristics, native AODV and AODV in a min-max way with Ci cost function. The average lifetime is represented in seconds in X axis while in Y axis we represent the number of connections according to the degree of load in the network.

We can see the heuristic E1 and E2 outperforms the native AODV and AODV in a min-max way with Ci cost function. This results show that a node ranking considering the idle time and node degree history can select a more robust path in terms of energy prolonging the network lifetime. The higher traffic load favor the heuristics since more robust paths are selected. The worst performance of the native AODV is expected since uses the shortest path hop count as metric, which does not consider the energy resources of the nodes. The more traffic load is the worst performance of the native AODV. The AODV-minmax-Ci is expected to have a better performance than native since this mechanism consider the best path is the one that has the lowest bottleneck in terms of energy. However, nodes with low energy ranking is still selected on the path.

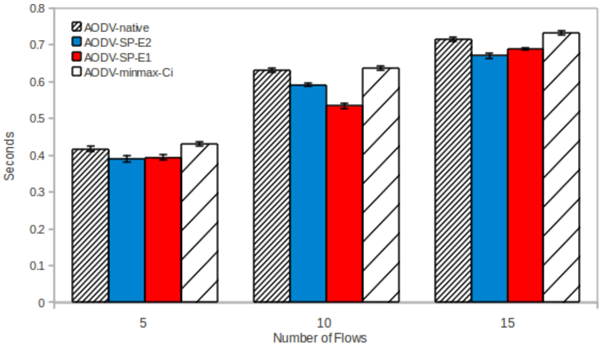


Figure 2. Average End-to-end Delay

Figure 2 show the average end-to-end delay of the E1 and E2 heuristics, native AODV and AODV in a min-max way with Ci cost function. The X axis represents the average end-to-end delay in seconds while in Y axis we represent the number of connections according to the degree of load in the network.

According to the results, our E1 and E2 heuristics are not penalized regarding the average end-to-end delay, unless the higher traffic load the gain is more since the node ranking is favor to more robust paths. The AODV native and AODV in a min-max way with Ci cost function have around the same delay values. It is surprise since we expect the min-max way should have higher delay than others because the mechanism selects path with excessive hop count depending the scenario and node energy costs.

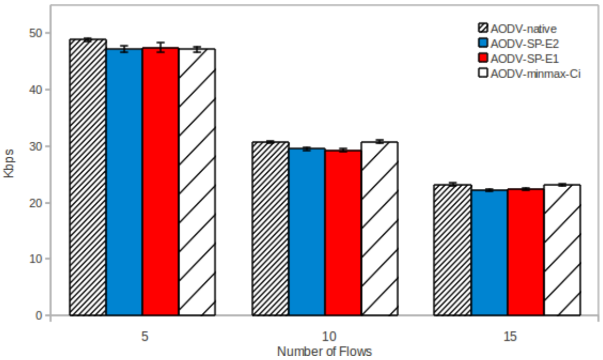


Figure 3. Average throughput

Figure VI-B show the average throughput of the E1 and E2 heuristics, native AODV and AODV in a min-max way with Ci cost function. The X axis represents the average throughput in Kbps while in Y axis we represent the number of connections according to the degree of load in the network.

The results show our E1 and E2 heuristics are not penalized regarding the average throughput for all traffic load. It is expected due to robust paths selected according to the node ranking. The heuristics, AODV native and AODV in a min-max way with Ci cost function have around the same throughput values. It is important to emphasize since the our goal is optimize the network lifetime without penalize the other network performance metrics.

VII. Conclusions and Future Work

Energy efficiency is a key aspect to consider in people-centric routing environments. We proposed a energy-awareness ranking of node based on idle times, which a node provides a ranking in terms of the node robustness to optimize the node lifetime as well as the global network lifetime. Then we consider the impact of node degree history for ranking the node to extend the lifetime.

We evaluated both heuristics in a distance vector multihop routing protocol, namely AODV, showing that a more robust in terms of energy is selected allowing to preserve the energy resources and selecting a path robust too.

As a future work, we are working on providing an analysis based on simulations of the different metrics and heuristics for link-based cost when applied to distance vector approach. For the link state approach, i.e., OLSR [18] routing protocol, we will provide analysis of the heuristics for node-based cost and link-based cost. We also are providing analysis with different networks scenarios with different load traffic and also with a social mobility pattern regarding mobility.

Acknowledgment

This work is supported by Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) in the context of the project entitled "Redes Oportunistas Sem Fio Energeticamente Eficientes como Suporte a Inclusão Digital e Fácil Acesso a Informação no Sudeste Goiano" number 201210267001156.

References

[1] C. E. Perkins, E. M. Belding-Royer, and S. R. Das, “Ad hoc on-demand distance vector (aodv) routing,” RFC Experimental 3561, Internet Engineering Task Force, July 2003.

[2] C. Yu, B. Lee, and H. Y. Youn, “Energy efficient routing protocols for mobile ad hoc networks,” *Wireless Communications and Mobile Computing*, vol. 3, no. 8, pp. 959–973, 2003.

[3] C. E. Jones, K. M. Sivalingam, P. Agrawal, and J. C. Chen, “A survey of energy efficient network protocols for wireless networks,” *Wirel. Netw.*, vol. 7, no. 4, pp. 343–358, 2001.

[4] S. Mahfoudh and P. Minet, “Survey of energy efficient strategies in wireless ad hoc and sensor networks,” in *Seventh International Conference on Networking, ICN 2008*, pp. 1–7, April 2008.

[5] K. Scott and N. Bambos, “Routing and channel assignment for low power transmission in pcs,” in *5th IEEE International Conference on Universal Personal Communications*, vol. 2, pp. 498–502, Oct 1996.

[6] J.-H. Chang and L. Tassiulas, “Energy conserving routing in wireless ad-hoc networks,” in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1, pp. 22–31 vol.1, 2000.

[7] Q. Xie, C.-T. Lea, M. Golin, and R. Fleischer, “Maximum residual energy routing with reverse energy cost,” in *Global Telecommunications Conference, IEEE GLOBECOM ’03.*, vol. 1, pp. 564–569, Dec. 2003.

[8] C.-K. Toh, “Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks,” *IEEE Communications Magazine,*, vol. 39, pp. 138–147, Jun 2001.

[9] A. Misra and S. Banerjee, “MRPC: maximizing network lifetime for reliable routing in wireless environments,” in *Wireless Communications and Networking Conference, IEEE WCNC*, pp. 800–806, Mar 2002.

[10] S.-M. Senouci and G. Pujolle, “Energy efficient routing in wireless ad hoc networks,” in *ICC 2004 - IEEE International Conference on Communications*, vol. 7, pp. 4057–4061, June 2004.

[11] Z. Guo, S. Malakooti, S. Sheikh, C. Al-Najjar, and B. Malakooti, “Multi-objective olsr for proactive routing in manet with delay, energy, and link lifetime predictions,” *Applied Mathematical Modelling*, vol. 35, no. 3, pp. 1413 – 1426, 2011.

[12] K. P. M. Kim, JP. Vasseur and H. Chong, “Routing Metrics used for Path Calculation in Low Power and Lossy Networks.” draft-ietf-roll-routing-metrics-19 (working in progress), 2011.

[13] S. Singh, M. Woo, and C. S. Raghavendra, “Power-aware routing in mobile ad hoc networks,” in *MobiCom ’98: Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*, (New York, NY, USA), pp. 181–190, ACM, 1998.

[14] D. Kim, J. J. Garcia-Luna-Aceves, K. Obraczka, J.-C. Cano, and P. Manzoni, “Routing mechanisms for mobile ad hoc networks based on the energy drain rate,” *IEEE Transactions on Mobile Computing*, vol. 2, no. 2, pp. 161–173, 2003.

[15] “The Network Simulator NS-2.” http://www.isi.edu/nsnam/ns/.

[16] D. Johnson, Y. Hu, and D. Maltz, “The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4,” RFC 4728, Internet Engineering Task Force, February 2007.

[17] G. C. Ewing, K. Pawlikowski, and D. Mcnickle, “Akaroa2: Exploiting network computing by distributing stochastic simulation,” in *International Society for Computer Simulation*, pp. 175–181, 1999.

[18] T. Clausen and P. Jacquet, “Optimized Link State Routing Protocol (OLSR),” RFC 3626, Internet Engineering Task Force, October 2003.

QoS-Based Power-Efficient Resource Management for LTE-A Networks with Relay Nodes

Kai-Ten Feng, Tzu-Hao Su, and Tain-Sao Chang
Department of Electrical and Computer Engineering
National Chiao Tung University, Hsinchu, Taiwan
ktfeng@mail.nctu.edu.tw, lycosidae.cm99g@nctu.edu.tw, and tschang06@nctu.edu.tw

***Abstract*—This paper investigates the resource management problem in the relay-enhanced long term evolution advanced (LTE-A) systems. The challenges of this resource allocation problem arise from the complication of assigning transmission links to the multicast broadcast single frequency network (MBSFN) subframe within the numerous physical resource block (PRB) pair. Existing research work does not fully consider all the influential factors to achieve power efficiency for the LTE-A networks. In this paper, a power-efficient QoS-based resource management (PERM) scheme is proposed to allocate MBSFN subframes, PRB pairs, and transmission power. The proposed PERM scheme is targeting at power efficiency with the consideration of both QoS requirements of user equipment (UE) and direct/two-hop communications. Moreover, the heuristic PERM (H-PERM) scheme is designed to provide efficient resource allocation for the LTE-A systems compared to the original PERM scheme. Simulation results show that the proposed schemes can provide power efficiency with consideration of UE's QoS requirements for the LTE-A systems.**

I. Introduction

Due to the rise of smart phones in recent years, the demand for high-speed mobile networks rapidly increase. The 3GPP long term evolution Advanced (LTE-A) [1] is a standard for next generation wireless communication systems to provide higher data rate services. The orthogonal frequency division multiplexing access (OFDMA) technology is a mutli-user version of original OFDM for LTE-A downlink, which divides the wideband channel into numerous subchannel in order to both provide high spectral efficiency and alleviate frequency-selective fading. Based on the OFDMA technique, multiuser diversity can be achieved by opportunistic scheduling which appropriately allocates the subsets of subchannels to individual user equipment (UE). However, excessive power consumption can be induced when the wireless network operator intends to provide quality-of-service (QoS) for UEs. In particular, UEs may inevitably be assigned to operate under a worse channel for data transmission which can result in additional power consumption. Therefore, relay node (RN) is introduced to provide an alternative path between evolved nodeB (eNB) and the UEs. Data transmission will have the flexibility to be conducted in either the original direct path from eNB to UE or via the RN. Therefore, it is important to provide feasible resource management for relay-enhanced communications in order to both fulfill the QoS requirements of UEs and preserve network energy.

Related research in [2] focuses on subchannel assignment and path selection by comparing the effective data rates between relay-based and direct transmissions. A void filling algorithm is proposed in [3] as a heuristic joint path selection and subchannel allocation scheme for throughput enhancement. However, these two schemes are designed with constant power allocation for UEs. The suboptimal solutions are obtained by jointly considering the subchannel and power allocation with QoS consideration for direct transmission [4] and relay-based network [5]. Furthermore, the two transmission phases within a subchannel are assigned to a single UE by these [2; 3; 5], i.e., the RN receives data from eNB in the first transmission phase and utilizes the same subchannel to forward the data in the second phase. On the other hand, heuristic switching assignments between the two transmission phases are considered in [6] for power allocation and in [7] for opportunistic power scheduling. However, QoS constraints from UEs have not been addressed in these two schemes. Moreover, the coexistence of both direct and relay-based communications in the network has not been investigated in most of existing research. A QoS-based resource allocation scheme is proposed in [8] which provides optimal assignments of two transmission phases considering both direct and relay-based links. Note that most of the existing studies are designed based on either generic OFDMA networks or IEEE 802.16 systems. For LTE-A standard, a heuristic resource allocation scheme has been proposed in [9] that utilizes the relay zone to ensure the transmission of RN-UE links to a specific portion of the entire resource allocation. Feasible throughput performance of the entire network can be achieved.

However, it is noticeable to observe that most of the existing research focuses on maximization of network throughput. Since most UEs in a wireless network are battery-powered, power efficiency is considered one of the principal issues to prolong the lifetime of UE. Moreover, energy conservation of network components, including RNs and eNBs is crucial from green energy perspective. Therefore, a power-efficient QoS-based resource management (PERM) scheme is proposed in this paper to solve the resource allocation problem for LTE-A systems. According to the further advancemet of LTE-A standard [10], the multicast broadcast

[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, and the Telecommunication Laboratories at Chunghwa Telecom Co. Ltd, Taiwan.

single frequency network (MBSFN) subframes are defined to specifically reserve for transmissions of eNB-RN links. All the remaining transmission links can only be allocated to those subframes other than MBSFN. Moreover, instead of adopting two transmission phases in most of existing research work, total of ten subframes can be individually allocated for each UE in the LTE-A systems. Note that each subframe consists of two physical resource block (PRB), i.e., denoted as a PRB pair. Therefore, the resource allocation problem for the LTE-A network is considered more challenging with both the constraint from MBSFN subframes and the additional degree of freedom to allocate the PRB pairs for UEs. An optimization problem is formulated by the proposed PERM scheme to acquire resource allocation for MBSFN subframes, PRB pairs, and transmission power. Moreover both the QoS requirements from UEs and direct/two-hop communications are considered in the PERM scheme. Owing to the NP-hard nature of the original optimization problem for resource allocation, the Lagrangian formulation is adopted to obtain the suboptimal solution based on the continuous relaxation [11] for PRB pair and MBSFN subframe assignment. However, intensive computation is required for solving the proposed PERM scheme due to the complication of resource allocation for the ten PRB pairs. Therefore, a heuristic PERM (H-PERM) scheme is proposed to efficiently resolve the resource allocation problem for LTE-A systems. Hungarian algorithm [12] is adopted to heuristically perform resource allocation for MBSFN, PRB pair, and transmission power. Simulation results show that the proposed PERM scheme can provide better power efficiency than conventional direct transmission. With slightly sacrificing power saving performance, the proposed H-PERM method effectively reduces computation complexity of the original PERM scheme.

II. System Model and Problem Formulation

As shown in Fig. 1, a downlink scenario of relay-enhanced LTE-A system is considered. There exists an eNB, R fixed RNs, and U UEs in a single cellular network. The total channel bandwidth is equally divided into N subcarriers each with B Hz. The downlink transmission frame is equally divided into $T = 10$ subframe as shown in Fig. 2. According to the LTE-A specification [10], a PRB pair is the basic unit of resource which consists of two time slots in the time domain and N_c consecutive subcarriers in the frequency domain. Note that each PRB pair can only be allowed to allocate one transmission link. In the relay-based LTE-A system, a two phase half-duplex transmission mode is adopted. To facilitate the operations of RNs in the network, each subframe is classified as either MBSFN subframe or normal subframe. Following relay-based LTE-A specification [13], the MBSFN subframe is possible to be assigned at number 1, 2, 3, 6, 7, and 8 subframe. The MBSFN subframe can be only assigned to the eNB-RN links and the normal subframe can be assigned to either eNB-UE links or RN-UE links, as shown in Fig. 2. Noted that the eNB-UE link represents direct communication between the eNB and the UE where the RNs are not involved in data transmission.

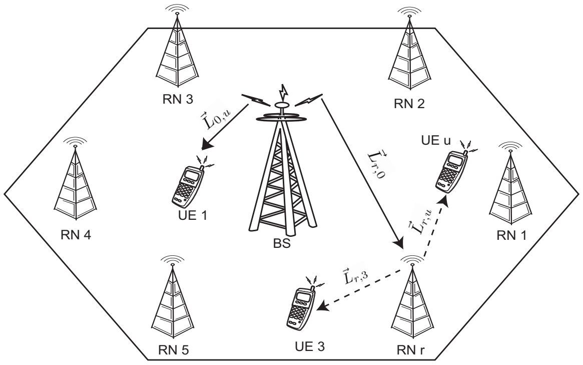


Fig. 1. Downlink relay-based LTE-A system.

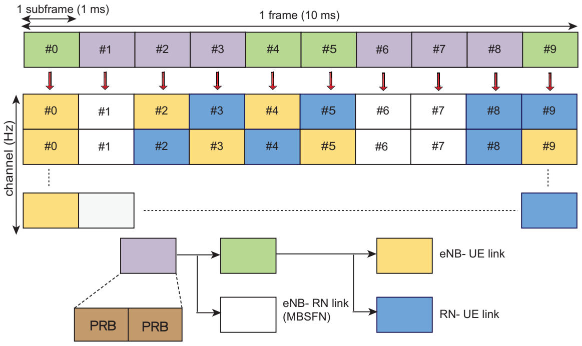


Fig. 2. Schematic diagram of ten subframes transmission for downlink relay-based LTE-A system.

Let $\vec{L}_{r,u}$ be denoted as the transmission link from relay r to UE u. $\vec{L}_{r,0}$ and $\vec{L}_{0,u}$ respectively denote transmission links of eNB to RN r link and eNB to UE u link. The relay selection function is defined as $\Omega(u) = r$ if a UE u is served by the RN r; while $\Omega(u) = 0$ if a UE u is operated in direct transmission. The parameter θ^{τ} is defined as the MBSFN subframe binary assignment variable for assigning MBSFN subframe at $\tau \in \{1, \ldots, T\}$th subframe, i.e.,

$$\theta^{\tau} = \begin{cases} 1, & \text{if } \tau\text{th subframe is the MBSFN subframe.} \\ 0, & \text{if } \tau\text{th subframe is the normal subframe.} \end{cases} \quad (1)$$

Moreover, $\rho_{r,u}^{n,\tau} \in \{0, 1\}$ denotes the PRB pair assignment indicator for $\vec{L}_{r,u}$ on the τth subframe of PRB pair n as either assigned ($\rho_{r,u}^{n,\tau} = 1$) or not assigned ($\rho_{r,u}^{n,\tau} = 0$). The other two PRB pair assignment indicators $\rho_{0,u}^{n,\tau} \in \{0, 1\}$ and $\rho_{r,0}^{n,\tau} \in \{0, 1\}$ can also be defined in a similar manner.

Before each downlink transmission, the eNB can obtain all the channel state information (CSI), e.g., the channel gain, of both the RNs and UEs based on their corresponding feedback mechanisms. It is also assumed that the channel gains of all the communication links remain constant in one downlink transmission frame. The normalized data rate $C_{r,u}^{n,\tau}$ of $\vec{L}_{r,u}$ on PRB pair n of τth subframe can be acquired as

$$C_{r,u}^{n,\tau} = (1 - \theta^{\tau})\rho_{r,u}^{n,\tau} \log_2(1 + p_{r,u}^{n,\tau} g_{r,u}^{n,\tau}), \quad (2)$$

where $p^{n,\tau}_{r,u}$ is the transmission power, $g^{n,\tau}_{r,u} = \frac{|H^n_{r,u}|^2}{\Gamma B N_0}$ with $H^n_{r,u}$ as the channel gain of $\vec{L}_{r,u}$, and N_0 is the power spectral density of additive white Gaussian noise (AWGN). The parameter $\Gamma = -\ln(5BER)/1.5$ is obtained from [14] given the target bit error rate BER and the continuous-rate M-ary quadrature amplitude modulation. Based on the normalized data rate $C^{n,\tau}_{r,u}$ obtained from (2), an optimization problem with the objective to minimize the sum power of entire system can be formulated as

$$\min_{(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{p})} \sum_{n=1}^{N}\sum_{\tau=1}^{T}\sum_{(r,u)\in L_{r,u}} p^{n,\tau}_{r,u} \tag{3a}$$

$$\text{s. t.} \quad \theta^\tau, \rho^{n,\tau}_{r,u} \in \{0,1\}, \forall\, n, \forall\, \tau, \forall\, r, \forall\, u; \tag{3b}$$

$$(1-\theta^\tau)\sum_{n=1}^{N}\sum_{u=1}^{U} p^{n,\tau}_{0,u} + \theta^\tau \sum_{n=1}^{N}\sum_{r=1}^{R} p^{n,\tau}_{r,0} \le P^{max}_{eNB},\ \forall\,\tau; \tag{3c}$$

$$(1-\theta^\tau)\sum_{n=1}^{N}\sum_{u=1}^{U} p^{n,\tau}_{r,u} \le P^{max}_{r},\ \forall\, r, \forall\,\tau; \tag{3d}$$

$$\theta^\tau \sum_{r=1}^{R} \rho^{n,\tau}_{0,u} + (1-\theta^\tau)[\sum_{r=1}^{R}\sum_{u=1}^{U} \rho^{n,\tau}_{r,u} + \sum_{u=1}^{U} \rho^{n,\tau}_{0,u}] \le 1,$$

$$\forall n, \forall\,\tau; \tag{3e}$$

$$\sum_{\tau=1}^{T}\sum_{n=1}^{N} C^{n,\tau}_{r,u} \ge R^{QoS}_{r,u},\ \forall L_{r,u}; \tag{3f}$$

where $\boldsymbol{\theta}$, $\boldsymbol{\rho}$ and $\boldsymbol{p}$ are defined as the sets of θ^τ, $\rho^{n,\tau}_{r,u}$ and $p^{n,\tau}_{r,u}$ for all n, τ, r, and u, respectively. The expression in (3b) states each communication link can be either assigned with a PRB pair, i.e., $\rho^{n,\tau}_{r,u} = 1$, or not assigned, i.e., $\rho^{n,\tau}_{r,u} = 0$. The constraints in (3c) indicate that the transmission power of eNB should be smaller than maximum transmission power of the eNB, i.e., P^{max}_{eNB}. The constraint in (3d) ensures that RN r's transmission power cannot exceed P^{max}_{r} for all r, where P^{max}_{r} is the maximum transmission power of RN r at each subframe. (3e) is utilized to denote that each PRB pair is allocated with at most one communication link in each subframe. The condition in (3f) indicates that each UE is required to satisfy its QoS constraint, where the parameter $R^{QoS}_{r,u}$ is the minimum required transmission rate of UE u and RN r according to its QoS requirement.

It can be observed that the optimization problem in (3a) is a mixed integer programming which is in general considered as an NP-hard problem and does not exist efficient algorithm to acquire the optimal solution except by adopting the exhausted search algorithm. Note that the complexity of exhausted search algorithm to allocate PRB pairs is $O\left(T\left[(U + R\cdot U)^N\right]\right)$. Moreover, The optimization problem in (3a) is not considered as convex due to the discrete manner of θ^τ and $\rho^{n,\tau}_{r,u}$ assignment indicators, i.e., can only be assigned with either 0 or 1 value. In the case that the constraint can be released as stated in [11], the indicators θ^τ and $\rho^{n,\tau}_{r,u}$ will be allowed to be any value within the interval $[0,1]$. Let $\varepsilon^{n,\tau}_{r,u}$ be defined as the effective transmission power as $\varepsilon^{n,\tau}_{r,u} = p^{n,\tau}_{r,u}\rho^{n,\tau}_{r,u}$, the normalized data rate $C^{n,\tau}_{r,u}$ of $\vec{L}_{r,u}$ in (2) can be rewritten as

$$C^{n,\tau}_{r,u} = (1-\theta^\tau)\rho^{n,\tau}_{r,u}\log_2\left(1 + \frac{\varepsilon^{n,\tau}_{r,u} g^{n,\tau}_{r,u}}{\rho^{n,\tau}_{r,u}}\right), \tag{4}$$

By defining $\boldsymbol{\varepsilon}$ as the set of $\varepsilon^{n,\tau}_{r,u}$ for all n, τ, r, and u, the set $(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{p})$ in (3a) will be replaced with $(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{\varepsilon})$, and the constraints in (3b)-(3d) can be modified as

$$\theta^\tau, \rho^{n,\tau}_{r,u} \in [0,1], \forall\, n, \forall\,\tau, \forall\, r, \forall\, u, \tag{5a}$$

$$(1-\theta^\tau)\sum_{n=1}^{N}\sum_{u=1}^{U} \varepsilon^{n,\tau}_{0,u} + \theta^\tau \sum_{n=1}^{N}\sum_{r=1}^{R} \varepsilon^{n,\tau}_{r,0} \le P^{max}_{BS}; \tag{5b}$$

$$(1-\theta^\tau)\sum_{n=1}^{N}\sum_{u=1}^{U} \varepsilon^{n,\tau}_{r,u} \le P^{max}_{r}. \tag{5c}$$

As a result, the optimization problem for resource management in (3a) can be converted into a convex optimization problem by replacing (3b)-(3d) with (5a)-(5c). Moreover, the convex optimization problem can be solved by applying the Lagrange Multiplier method.

III. Proposed Power-Efficient QoS-based Resource Management (PERM) Scheme

In this section, the proposed PERM scheme will be described. Let λ^τ_1 and $\lambda^\tau_{2,r}$ be defined as the Lagrangian multipliers of (5b) and (5c), respectively. The parameters $\lambda_{2,r}$, $\eta^{n,\tau}$, and $\mu^\tau_{r,u}$ are the Lagrangian multipliers of the rth RN's constraint in (5c), the τth subframe of nth PRB pair's constraint in (3e), the uth UE's constraint and the rth RN's constraint in (3f), respectively. Moreover, $\boldsymbol{\Phi}$ is defined as the set of all Lagrangian multipliers. Hence, the Lagrangian function $L(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{\varepsilon},\boldsymbol{\Phi})$ of modified optimization problem in (3a) along with the convex properties as in (5a)-(5c) can be formulated as

$$\begin{aligned} L(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{\varepsilon},\boldsymbol{\Phi}) &= \sum_{n=1}^{N}\sum_{\tau=1}^{T}\sum_{(r,u)\in L_{r,u}} \varepsilon^{n,\tau}_{r,u} \\ &+ \sum_{\tau=1}^{T}\lambda^\tau_1\left((1-\theta^\tau)\sum_{n=1}^{N}\sum_{u=1}^{U}\varepsilon^{n,\tau}_{0,u} + \theta^\tau\sum_{n=1}^{N}\sum_{r=1}^{R}\varepsilon^{n,\tau}_{r,0} - P^{max}_{eNB}\right) \\ &+ \sum_{\tau=1}^{T}\sum_{r=1}^{R}\lambda^\tau_{2,r}\left((1-\theta^\tau)\sum_{n=1}^{N}\sum_{u=1}^{U}\varepsilon^{n,\tau}_{r,u} - P^{max}_{r}\right) \\ &+ \sum_{\tau=1}^{T}\sum_{n=1}^{N}\eta^{n,\tau}\left(\theta^\tau\sum_{r=1}^{R}\rho^{n,\tau}_{0,u} + (1-\theta^\tau)[\sum_{r=1}^{R}\sum_{u=1}^{U}\rho^{n,\tau}_{r,u}\right. \\ &\left. + \sum_{u=1}^{U}\rho^{n,\tau}_{0,u}] - 1\right) - \sum_{\tau=1}^{T}\sum_{(r,u)\in L_{r,u}}\mu^\tau_{r,u}\left(\sum_{n=1}^{N}C^{n,\tau}_{r,u} - R^{QoS}_{r,u}\right). \end{aligned} \tag{6}$$

Furthermore, the Karush-Kuhn-Tucker (KKT) conditions of modified convex optimization problem for obtaining the optimal solution are given by

$$\frac{\partial L(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{\varepsilon},\boldsymbol{\Phi})}{\partial\varepsilon^{n,\tau}_{r,u}} \begin{cases} \le 0, & \text{if } \varepsilon^{n,\tau}_{r,u} = 0 \\ = 0, & \text{if } \varepsilon^{n,\tau}_{r,u} > 0 \end{cases} \tag{7a}$$

$$\frac{\partial L(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{\varepsilon},\boldsymbol{\Phi})}{\partial \rho_{r,u}^{n,\tau}} \begin{cases} \leq 0, & \text{if } \rho_{r,u}^{n,\tau}=0 \\ =0, & \text{if } \rho_{r,u}^{n,\tau}>0 \end{cases} \tag{7b}$$

$$\frac{\partial L(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{\varepsilon},\boldsymbol{\Phi})}{\partial \theta^{\tau}} \begin{cases} \leq 0, & \text{if } \theta^{\tau}=0 \\ =0. & \text{if } \theta^{\tau}>0 \end{cases} \tag{7c}$$

Equation (7a) can further be expressed as

$$\frac{\partial L(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{\varepsilon},\boldsymbol{\Phi})}{\partial \varepsilon_{r,u}^{n,\tau}} = \lambda_r^{\tau} - \frac{\mu_{r,u}^{\tau}\theta^{\tau}\rho_{r,u}^{n,\tau}g_{r,u}^{n,\tau}}{\rho_{r,u}^{n,\tau}+\varepsilon_{r,u}^{n,\tau}g_{r,u}^{n,\tau}}, \tag{8}$$

where

$$\lambda_r^{\tau} = \begin{cases} 1+\lambda_1^{\tau}\left(1-\theta^{\tau}\right), & \text{if } r=0 \text{ and } u\neq 0 \\ 1+\lambda_{2,r}^{\tau}\left(1-\theta^{\tau}\right), & \text{if } r\neq 0 \text{ and } u\neq 0 \\ 1+\lambda_{1,r}^{\tau}\theta^{\tau}. & \text{if } r\neq 0 \text{ and } u=0 \end{cases} \tag{9}$$

Therefore, according to (7a) and (8), the effective transmission power $\varepsilon_{r,u}^{n,\tau}$ can be written as

$$\varepsilon_{r,u}^{n,\tau} = \rho_{r,u}^{n,\tau}\left(\frac{\mu_{r,u}^{\tau}\theta^{\tau}}{\lambda_r^{\tau}} - \frac{1}{g_{r,u}^{n}}\right)^{+}. \tag{10}$$

It is noted that the expression $(x)^+$ in (10) indicates $(x)^+ = x$ if $x \geq 0$ and $(x)^+ = 0$ if $x < 0$. Furthermore, (7b) can also be similarly derived as

$$\begin{aligned} &\frac{\partial L(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{\varepsilon},\boldsymbol{\Phi})}{\partial \rho_{r,u}^{n,\tau}} = \eta^{n,\tau}\theta^{\tau} \\ &-\mu_{r,u}^{\tau}\left[\theta^{\tau}\log_2\left(1+\frac{\varepsilon_{r,u}^{n,\tau}g_{r,u}^{n,\tau}}{\rho_{r,u}^{n,\tau}}\right) - \frac{\theta^{\tau}\varepsilon_{r,u}^{n,\tau}g_{r,u}^{n,\tau}}{\rho_{r,u}^{n,\tau}+\varepsilon_{r,u}^{n,\tau}g_{r,u}^{n,\tau}}\right]. \end{aligned} \tag{11}$$

By substituting (10) into (11), the parameter $D_{r,u}^{n,\tau}(\mu_{r,u}^{\tau},\lambda_1^{\tau},\lambda_{2,r}^{\tau})$ can be defined as

$$\begin{aligned} &D_{r,u}^{n,\tau}(\mu_{r,u}^{\tau},\lambda_1^{\tau},\lambda_{2,r}^{\tau}) \\ &=\mu_{r,u}^{\tau}\left[\theta^{\tau}\left(\log_2\left(\frac{\theta^{\tau}\mu_{r,u}^{n,\tau}g_{r,u}^{n}}{\lambda_r^{\tau}}\right)\right)^{+} - \left(1-\frac{\lambda_r^{\tau}}{\theta^{\tau}\mu_{r,u}^{n,\tau}g_{r,u}^{n}}\right)\right] \\ &\begin{cases} \leq \eta^{n,\tau}, & \text{if } \rho_{r,u}^{n,\tau}=0, \\ =\eta^{n,\tau}, & \text{if } \rho_{r,u}^{n,\tau}>0. \end{cases} \end{aligned} \tag{12}$$

As a consequence of above formulation, given the $\hat{n}$th PRB pair and the $\hat{\tau}$th transmission subframe, there exists a link $\vec{L}_{r^*,u^*}$ such that

$$(r^*,u^*) = \arg\max_{(r,u)} D_{r,u}^{\hat{n},\hat{\tau}}(\mu_{r,u}^{\hat{\tau}},\lambda_1^{\hat{\tau}},\lambda_{2,r}^{\hat{\tau}}). \tag{13}$$

If there exists a unique $\vec{L}_{r^*,u^*}$ to achieve the maximal value of $D_{r,u}^{n,\tau}(\mu_{r,u}^{\tau},\lambda_1^{\tau},\lambda_{2,r}^{\tau})$, the optimal resource allocation can be obtained such that

$$\rho_{r^*,u^*}^{\hat{n},\hat{\tau}} = 1,\ \rho_{r,u}^{\hat{n},\hat{\tau}} = 0, \forall r \neq r^* \text{ or } \forall u \neq u^*. \tag{14}$$

As mentioned before, the PRB pair assignment indicator is relaxed from the original two distinct values, i.e., $\rho_{r,u}^{n,\tau} \in \{0,1\}$, into a continuous set of values in the interval of $[0,1]$. Therefore, the resulting optimal solution can happen if the indicator $\rho_{r,u}^{n,\tau}$ is a fractional value between $[0,1]$. In such case, suboptimal and non-unique solutions with link $\vec{L}_{r^*,u^*}$ that satisfy (13) can be acquired by constraining $\rho_{r,u}^{n,\tau}$ to be either 0 or 1. As a result, the τth subframe of nth PRB pair will be assigned with the link that has the largest value of $D_{r,u}^{n,\tau}(\mu_{r,u}^{\tau},\lambda_1^{\tau},\lambda_{2,r}^{\tau})$. The allocation for all the subframes can also be determined in a similar manner. Similarly, (7c) can be derived as

$$\begin{aligned} &\frac{\partial L(\boldsymbol{\theta},\boldsymbol{\rho},\boldsymbol{\varepsilon},\boldsymbol{\Phi})}{\partial \theta^{\tau}} = \lambda_1^{\tau}\left(\sum_{n=1}^{N}\sum_{u=1}^{U}\varepsilon_{0,u}^{n,\tau} - \sum_{n=1}^{N}\sum_{r=1}^{R}\varepsilon_{r,0}^{n,\tau}\right) \\ &+\sum_{r=1}^{R}\lambda_{2,r}^{\tau}\left(\sum_{n=1}^{N}\sum_{u=1}^{U}\varepsilon_{r,u}^{n,\tau}\right) \\ &+\sum_{n=1}^{N}\eta^{n,\tau}\left[\sum_{u=1}^{U}\left(\left(\sum_{r=1}^{R}\rho_{r,u}^{n,\tau}\right)+\rho_{0,u}^{n,\tau}\right)-\sum_{r=1}^{R}\rho_{r,0}^{n,\tau}\right] \\ &-\sum_{r=1}^{R}\sum_{u=1}^{U}\mu_{r,u}^{\tau}\sum_{n=1}^{N}\rho_{r,u}^{n,\tau}log_2\left(1+\frac{\varepsilon_{r,u}^{n,\tau}g_{r,u}^{n}}{\rho_{r,u}^{n,\tau}}\right) \\ &\begin{cases} \leq 0, & \text{if } \theta^{\tau}=0 \\ =0, & \text{if } \theta^{\tau}>0 \end{cases} \end{aligned} \tag{15}$$

Moreover, in order to obtain the suboptimal solution for (13), the values of Lagrangian multiplier are required to be obtained. An iterative approach that exploits the subgradient method as in [15] is utilized to update the value of each Lagrangian multiplier. For example, considering that $\mu_{r,u}^{\tau,i}$ is defined as the ith iteration of $\mu_{r,u}^{\tau,i}$, its updating process can be expressed as

$$\mu_{r,u}^{\tau,(i+1)} = \left(\mu_{r,u}^{\tau,(i)} - s^{(i)}\left(R_{r,u}^{QoS} - \sum_{n=1}^{N}C_{r,u}^{n,\tau}\right)\right)^{+}, \tag{16}$$

where $s^{(i)} = \alpha/\sqrt{i}$ is the step size and α is a tunable constant. The updating processes for the other Lagrangian multipliers can also be obtained similarly. The complexity of proposed PERM algorithm can be obtained as $O(NURTI)$ where I denotes the number of total iterations. It can be seen that the proposed scheme can provide more efficient computation than the exhaustive search algorithm especially under large number of PRB pairs.

IV. Proposed Heuristic Power-Efficient QoS-based Resource Management (H-PERM) Scheme

In this section, a low complexity H-PERM scheme which adopts the Hungarian algorithm [12] is designed to heuristically solve the optimization problem in (3a). The Hungarian algorithm is an optimal algorithm for solving the one-to-one assignment problem. In other words, a square-matrix relationship is required between input and output by adopting the Hungarian algorithm. The proposed H-PERM scheme is to heuristically form square matrix from the original formulation in order to apply the Hungarian method.

First of all, the required number of MBSFN subframes can be calculated by

$$M_N = \left\lceil \frac{\sum_{u=1}^{U} R_{\Omega(u)\neq 0,u}^{QoS}}{NC_{\tau}} \right\rceil, \tag{17}$$

where

$$C_{\tau} = \log_2(1+\kappa(N)), \tag{18}$$

$$\kappa(N) = 2^{\frac{R^{QoS}_{r,u}}{BN}} - 1. \tag{19}$$

Note that $\kappa(N)$ is a pre-defined rate adjustment threshold, which depends on the both uth UE's required data rate and the number of PRB pairs N. Moreover, the number of normal subframes which are assigned to $\vec{L}_{r,u}$ for $u \neq 0$ links can be obtained as $(T - M_N)$. In order to formulate the square matrix for Hungarian algorithm, the matrix $\mathbf{H}^{\tau}_{N\times N'}$ is defined for the τth subframe as

$$\mathbf{H}^{\tau}_{N\times N'} = \\ \begin{pmatrix} g^{1,1}_{\Omega(1),1} & g^{1,2}_{\Omega(1),1} & \cdots & g^{1,m_{r,1}}_{\Omega(1),1} & \cdots & g^{1,m_{r,U}}_{\Omega(U),U} \\ g^{2,1}_{\Omega(1),1} & \vdots & \ddots & g^{2,m_{r,1}}_{\Omega(1),1} & & \vdots \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ g^{N,1}_{\Omega(1),1} & g^{N,2}_{\Omega(1),1} & \cdots & g^{N,m_{r,1}}_{\Omega(1),1} & \cdots & g^{N,m_{r,U}}_{\Omega(U),U} \end{pmatrix}, \tag{20}$$

where $N' = \sum_{u=1}^{U} m_{\Omega(u),u}$ and $m_{\Omega(u),u}$ is denoted as the number of PRB pairs that are allocated to uth UE which is served by $\Omega(u)$th RN. For example, the elements in $\mathbf{H}^{\tau}_{N\times N'}$ matrix $g^{N,1}_{\Omega(1),1}$ represents the channel gain of the Nth assigned PRB pair in the first subframe from the $\Omega(1)$th relay to UE 1. Similarly, $m_{r,0}$ represents the number of PRB pairs that are assigned to rth RN which is served by eNB. In the proposed H-PERM scheme, same data rate across all the PRB pairs is assumed for each subframe.

The operations of proposed H-PERM algorithm for normal subframes is described as follows.

1) Initialize $m_{\Omega(u),u} = \lfloor N/U \rfloor$, $\forall u$.
2) Apply the Hungarian algorithm to $\mathbf{H}^{\tau}_{N\times N'}$, and then obtain the allocation matrix $\mathbf{I}_{N\times N'}$. Note that the ith row element of $\mathbf{I}_{N\times N'}$ represents that the corresponding UE is assigned with the ith PRB pair if its value is equal to 1.
3) For $\tau = 1$

   While $\sum_{u=1}^{U} m_{\Omega(u),u} < N$

$$\bar{g}_u = \frac{1}{m_{\Omega(u),u}} \sum_{j=u}^{m_{\Omega(u),u}} \sum_{i=1}^{N} (\mathbf{I}_{N\times N'})_{i,j} (\mathbf{H}^{\tau}_{N\times N'})_{i,j} \tag{21}$$

$$\triangle p \left(m_{\Omega(u),u}\right) = \frac{m_{\Omega(u),u}}{\bar{g}_u} \kappa(m_{\Omega(u),u}) - \frac{m_{\Omega(u),u} + 1}{\bar{g}_u} \kappa(m_{\Omega(u),u} + 1) \tag{22}$$

$$u' = \arg\max_{\forall u} (\triangle p \left(m_{\Omega(u),u}\right)) \tag{23}$$

$$m_{\Omega(u'),u'} = m_{\Omega(u'),u'} + 1 \tag{24}$$

   End while

   Update $\mathbf{H}^{\tau}_{N\times N'}$ and repeat 2).

   End for
4) For $\tau = 2 : T - M_N$

   While $\sum_{u=1}^{U} m_{\Omega(u),u} < N \cdot \tau$

      Calculate $\bar{g}_u$ and $\triangle p \left(m_{\Omega(u),u}\right)$ by (21) and (22), respectively.

      Obtain u' by (23) and update $m_{\Omega(u'),u'}$ by (24).

   End while

   Update $\mathbf{H}^{\tau}_{N\times N'}$ and repeat 2).

   End for

Note that the $\bar{g}_u$ in (21) is defined as the average channel gain for UE u over the $m_{\Omega(u),u}$ allocated PRB pairs. The parameter $\Delta p(m_{\Omega(u),u})$ in (22) represents the net power reduction while the UE u is assigned with an additional PRB pair. Equation (23) is utilized to acquire the UE that can result the maximum power reduction with an additional PRB assignment. Moreover, since the complexity of Hungarian algorithm is $O(N^3)$, that of the proposed H-PERM scheme can be obtained as $O(TN^3)$. Compared to the PERM algorithm as described in previous section, it can be observed that smaller computational complexity can be obtained by adopting the H-PERM scheme if $N^2 \leq URI$, which is considered valid since significant amount of iteration number I is required for obtaining Lagrangian multipliers.

V. Performance Evaluation

In this section, the performances of proposed PERM and H-PERM schemes will be compared with direct transmission. Considering a relay-enhanced network, there exists one eNB with radius of transmission range equal to 1732 meters, several numbers of RNs are located at the cell edge which are $1732 \times 2/3$ from the eNB, and 10 UEs which are randomly distributed within the transmission range of eNB. The large-scale fading model composed of both path loss and shadowing is adopted from urban macro scenario. Moreover, for small-scale fading model, Rician distributions for line-of-sight (LOS) environments are considered in eNB-RN links, and Rayleigh distributions for non-LOS scenarios are adopted in RN-UE as well as eNB-UE links. The system parameters and configurations are listed in Table 1 as below.

TABLE 1 : SYSTEM PARAMETERS

Parameter	Value
Number of PRB pairs (N)	50
Number of subframes (T)	10
Bandwidth of PRB (B)	180 KHz
Channel noise density	-174 dBm/Hz
Maximum transmission power of eNB (P^{max}_{eNB})	46 dBm
Maximum transmission power of RN (P^{max}_{r})	37 dBm
Target bit error rate (BER)	10^{-5}

Fig. 3 shows power consumption in eNB-UE, eNB-RN, and RN-UE links versus the number of MBSFN subframes under the minimum required data rate $R^{QoS}_{r,u} = 2$Mbps (left plot) and 5Mbps (right plot) for proposed H-PERM scheme with 6 RNs in the network. It can be observed that power consumption in eNB-RN links decreases as the number of MBSFN subframes is augmented since channel capacity is a non-linear logarithmic function. In other words, more power consumption in eNB-RN links is obtained when less MBSFN subframes are provided. On the other hand, the power consumption in eNB-UE and RN-UE links increases with larger number of

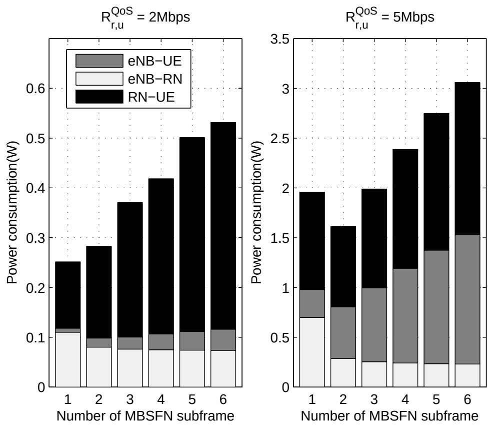


Fig. 3. Performance evaluation of proposed H-PERM scheme: power consumption in eNB-UE, eNB-RN, and RN-UE links versus the number of MBSFN subframes under the minimum required data rate $R^{QoS}_{r,u}$ = 2Mbps (left plot) and 5Mbps (right plot).

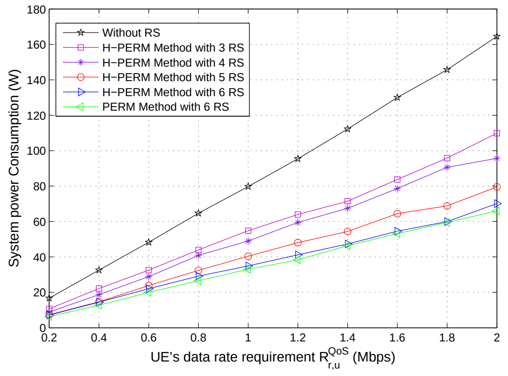


Fig. 4. Performance comparison: system power consumption versus the minimum required data rate of each UE $R^{QoS}_{r,u}$.

MBSFN subframes. Moreover, it can be seen that the system power consumption is minimum when the number of MBSFN subframes is set to one under $R^{QoS}_{r,u} =$ 2Mbps. For the case of $R^{QoS}_{r,u} =$ 5Mbps, setting the number of MBSFN subframes as two can achieve the optimal system performance. It is intuitive to explain the results since more MBSFN subframes are required for the data transmission in eNB-RN links in order to acquire larger $R^{QoS}_{r,u}$.

As shown in Fig. 4, system power consumption versus the minimum required data rate of each UE $R^{QoS}_{r,u}$ is demonstrated for performance comparison between proposed H-PERM/PERM schemes and direct transmission. Note that the cases with different numbers of RNs are conducted for the H-PERM scheme. Intuitively, system power consumption increases as $R^{QoS}_{r,u}$ is augmented. It can be seen that the proposed PERM scheme can provide better performance compared to both H-PERM scheme and direct transmission owing to its optimal formulation. By observing the H-PERM scheme, system power consumption can be further reduced by setting more RNs to improve UEs' channel qualities. The power consumption of H-PERM mechanism is almost the same as that of PERM scheme in the case that the number of RNs is equal to 6. The proposed PERM scheme is served as a performance upper bound; while the H-PERM scheme can be implemented in a computation-efficient manner. The merits of both PERM and H-PERM schemes can be observed.

VI. Conclusion

In this paper, power-efficient quality-of-service-based (QoS-based) resource management (PERM) scheme is proposed for the relay-enhanced long term evolution advanced (LTE-A) networks. The proposed PERM scheme is formulated as an optimal resource allocation for multicast broadcast single frequency network (MBSFN) subframes, physical resource block (PRB) pairs, and transmission power considering QoS requirements of user equipment. Compared with existing works, the factors of MBSFN subframe and PRB pairs are considered in order to meet the practical LTE-A systems. For the purpose of implementation, the simplified version of PERM named as heuristic PERM (H-PERM) scheme is proposed to reduce computational complexity. Numerical results show that the proposed H-PERM scheme with lower computational complexity can achieve almost the same performance as PERM mechanism from the perspective of power consumption.

References

[1] 3GPP TS 36.300 V9.2.0, "Overall Description Stage 2 (Release 9)," Dec. 2009.

[2] B. Can, H. Yanikomeroglu, F. A. Onat, E. De Carvalho and H. Yomo, "Efficient Cooperative Diversity Schemes and Radio Resource Allocation for IEEE 802.16j," in *Proc. IEEE WCNC*, Mar. 2008.

[3] L. Wang, Y. Ji, and F. Liu, "Resource Allocation for OFDMA Relay-Enhanced System with Cooperative Selection Diversity," in *Proc. IEEE WCNC*, Apr. 2009.

[4] N. U. Hassan and M. Assaad, "Resource Allocation in Multiuser OFDMA System: Feasibility and Optimization Study," in *Proc. IEEE WCNC*, Apr. 2009.

[5] Z. Tang and G. Wei, "Resource Allocation with Fairness Consideration in OFDMA-Based Relay Networks," in *Proc. IEEE WCNC*, Apr. 2009.

[6] L. Huang, M. Rong, L. Wang, Y. Xue, and E. Schulz, "Resource Allocation for OFDMA Based Relay Enhanced Cellular Networks," in *Proc. IEEE VTC*, Apr. 2007.

[7] B. G. Kim and J. W. Lee, "Opportunistic Power Scheduling for OFDMA Cellular Networks with Scheduling at Relay Stations," in *Proc. IEEE WCNC*, Apr. 2009.

[8] W. P. Chang, J. S. Lin, and K. T. Feng, "QoS-based Resource Allocation for Relay-Enhanced OFDMA Networks," in *Proc. IEEE WCNC*, Apr. 2011.

[9] B. G. Choi, S. J. Bae, K-y. Cheon, A-S. Park, and M. Y. Chung, "Relay Selection and Resource Allocation Schemes for Effective Utilization of Relay Zones in Relay-Based Cellular Networks," *IEEE Commun. Lett.*, Apr. 2011.

[10] 3GPP TR36.814 v9.0.0, "Further Advancements for E-UTRA Physical Layer Aspects (Release 9)," Mar. 2010.

[11] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1747 – 1758, Oct. 1999.

[12] H. W. Kuhn, "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1, pp. 83–97, 1955.

[13] 3GPP TS 36.211 V10.0.0, "Physical Channels and Modulation (Release 10)," Dec. 2010.

[14] M. S. Alouini and A. Goldsmith,, "Adaptive Modulation over Nakagami Fading Channels," *Wireless Personal Communications*, vol. 13, pp. 119–143, 2000.

[15] N. Z. Shor, K. C. Kiwiel, and A. Ruszcayǹski, *Minimization Methods for Non-differentiable Functions*. Springer-Verlag, 1985.

SESSION

POSTER PAPERS

Chair(s)

TBA

A New Routing Protocol for Improving Path Stability in Mobile Ad-hoc Networks

Hyung-jik Kim, Hae-il Choi, and Sunwoong Choi
School of Electrical Engineering, Kookmin University, Seoul, Korea

Abstract - *Since mobile ad-hoc network usually consists of battery-operated nodes, load balancing for balanced energy consumption is important. In this paper, we propose a new routing protocol to find the highest minimum node residual energy path among shortest paths for improving path stability in mobile ad-hoc networks. Using ns-3 simulation, we show that the proposed routing protocol provides more stable routing path than AODV and EA-AODV.*

Keywords: MANET, Routing Protocol, Path Stability, AODV, EA-AODV

1 Introduction

Traditional ad-hoc routing protocols such as DSDV [1], DSR [2], and AODV [3] aim to find the shortest path from a source node to a destination node. However, load balancing for balanced energy consumption is also important since mobile ad-hoc network usually consists of battery-operated nodes. In this paper, we propose a new routing protocol to find the highest minimum node residual energy path among shortest paths. A path with higher residual energy is expected to have longer lifespan, so it can improve routing path stability. Using ns-3 simulation, we show that the proposed routing protocol provides more stable routing path than other protocols, AODV and EA-AODV.

2 Related Work

Many routing protocols have been proposed for ad-hoc networks. Traditional routing protocols aim to find the shortest path from a source node to a destination node. Proactive protocols such as DSDV maintain routes between all source-destination pairs regardless of the use of such routes. On the other hand, reactive protocols such as DSR and AODV, try to find routes on demand, that is, when a source node requests them.

Mobile ad-hoc network usually consists of battery-operated nodes. So, the energy efficiency is important in mobile ad-hoc networks. Power-aware protocols consider the node residual energy determining the routing path for energy efficiency and load balancing. Some nodes may transmit and/or relay more traffic than others. This unbalanced battery power consumption may cause an early battery exhaustion of a node and network partitioning as a result.

M. Tamilarasi and T. G. Palanivelu proposed an energy aware protocol, EA-AODV [4] which operates similarly to AODV but selects a route based on the minimum energy availability and the energy consumption per packet of the routes. A source node initiates a route discovery by broadcasting the RREQ (Route Request) packet. The destination node replies back to the source node using the RREP (Route Reply) packet. When an intermediate node forwards the RREP packet, it records its residual energy in the packet. Then the source node selects the path with the highest minimum node residual energy thought AODV select the shortest path.

3 New Routing Protocol for Improving Path Stability

Note that the RREQ packet is broadcasted but the RREP packet is unicasted to the source node. The RREP packet is forwarded through the path of the first RREQ packet received by the destination node in the reverse direction. Thus, the RREP forwarding path is determined by each node’s random delay before forwarding the RREQ packet.

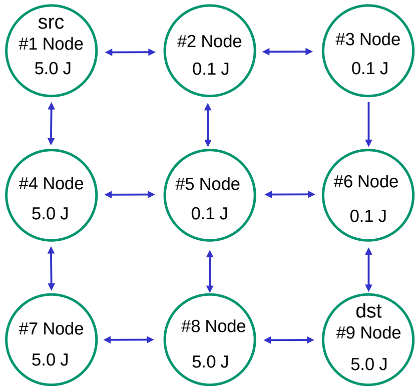


Fig. 1. Example topology with node residual energy.

Consider the simple topology in the Fig. 1. The source node is node 1 and the destination node is node 9. Each node’s residual energy is also given. There are 6 different shortest paths. Among them, the best routing path, that is, the path with the longest lifetime is 1 – 4 – 7 – 8 – 9 since the

Corresponding author: Sunwoong Choi (schoi@kookmin.ac.kr) minimum node residual energy is the highest, 5.0J. However, EA-AODV may not find the best routing path. If node 8 received the RREQ from node 5 earlier than from node 7, node 8 will forward the replied RREP packet to node 5.

In this paper, we propose to record the node residual energy information in the RREQ packet instead of RREP packet. Fig. 2 shows the flowchart of the relay nodes of the proposed protocol. When a node receives a RREQ packet, it determines whether the packet is broadcasted or discarded. First, the hop count in the RREQ packet is compared with the routing table entry. And then the minimum node residual energy is compared. Through this process, the routing path to the destination node is updated to the higher minimum node residual energy path, that is, the most stable path out of the shortest paths.

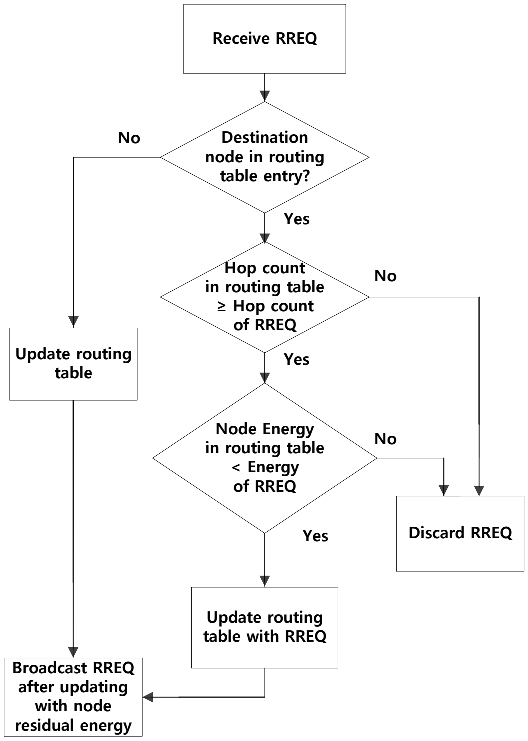

Receive RREQ
Destination node in routing table entry?
No
Yes
Hop count in routing table ≥ Hop count of RREQ
No
Yes
Update routing table
Node Energy in routing table < Energy of RREQ
No
Yes
Discard RREQ
Update routing table with RREQ
Broadcast RREQ after updating with node residual energy

Fig. 2. End-to-end delay.

4 Performance Evaluation

We evaluate the performance of the proposed routing protocol using ns-3 simulator [5]. We compare AODV, EA-AODV, and the proposed routing protocol in Fig. 1. The source node generates UDP packet to the destination node every 1 second. The simulation time is 30 seconds.

Fig. 3 shows the simulation result. We can observe a few packets suffer huge end-to-end delay with AODV and EA-AODV. Those delays result from a battery exhaustion of a node on the routing path. As a result, it takes considerable time to find a new routing path.

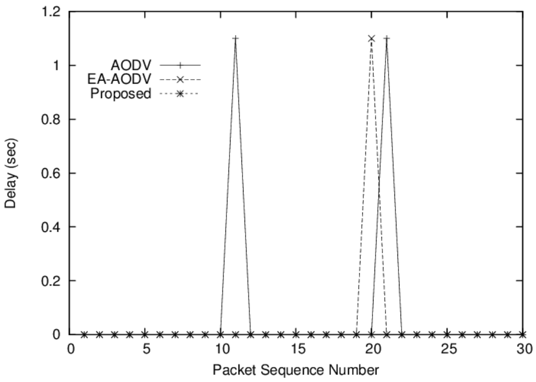


Fig. 3. Flowchart of the proposed protocol.

5 Conclusions

In this paper, we propose a new routing protocol for improving path stability in mobile ad-hoc networks. RREQ packets convey the node's residual energy information and the routing path is determined with the highest minimum node residual energy path. From computer simulation, we showed the proposed routing protocol provides more stable routing path than AODV and EA-AODV.

6 ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2011-0023856), and in part by the MSIP(Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013.

7 References

[1] C. Perkins and P. Bhagwat, "Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers," ACM SIGCOMM Computer Communication Review, vol. 24, no. 4, pp. 234–244, 1994.

[2] D. Johnson, D. Maltz, and J. Broch, "DSR: the dynamic source routing protocol formultihop wireless ad hoc networks," in Ad Hoc NetworkIng, pp. 139–172, Addison-Wesley, Boston, Mass, USA, 2001.

[3] C. Perkins, E.Belding-Royer, S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing", RFC 3561, July 2003.

[4] M. Tamilarasi and T. G. Palanivelu, "Integrated Energy-Aware Mechanism for MANETs using On-demand Routing", International Journal of Computer and Information Engineering, vol. 2 Issue 3, pp212-216, 2008.

[5] http://www.nsnam.org/

Developing Mobile Raspberry Pi Base Station Cloud for Wireless Sensor Networks

Andrew K. Wilson, Hwajung Lee
Department of Information Technology
Radford University
Radford, Virginia, U.S.A.
{awilson8, hlee3}@radford.edu

***Abstract*—A base station (BS) in a traditional wireless sensor network (WSN) is a single stationary node, which makes a decision based on the collected data from the WSN, However, the institution of a mobile cloud BS gives way to much greater computational power than that of a single node BS through way of distributed computing, extends the lifetime of the network by moving each node to the optimal location, and the ability to physically restore the sensor motes of the network. Thus, in this paper we propose a mobile BS cloud.**

Keywords-wireless sensor network; mobile; cloud; base station; optimization

I. INTRODUCTION

In a wireless sensor networks (WSN), a base station is essentially the brain of a network. It handles data aggregation and processing, intercommunication of sensor motes and the end user, as well as optimization strategies to extend the lifetime of the network.

The necessary numbers of hops required for message passing are decreased when instituting multiple base station node. A problem arises in this structure when focusing on network lifetime as a base station nodes closest mote takes on not only its own messages but also those of the motes in the near vicinity. The institution of a base station cloud that periodically moves to new locations based on an optimization problem solves this because the "routing" mote(s) continually change. Additionally, the institution of a WiTricity charging protocol will also greatly increase the network lifetime.

Another issue is the loss of sensor motes that can move or can be moved. A multiple node mobile base station can recover lost motes based on last known location and signals sent from the mote. Since each base station node is mobile it can move to the last known location of said mote and physically return it to the network and reestablish its connectivity.

Previously, a base station in a wireless sensor network has been a single, stationary machine with limited capabilities. However, the institution of a mobile cloud base station gives way to much greater computational power than that of a single node base station through way of distributed computing, extends the lifetime of the network by moving each node to the optimal location, and the ability to physically restore the sensor motes of the network. Thus, in this paper we propose a mobile base station cloud comprised of multiple Raspberry Pi powered Lego Mindstorms NXT robots.

II. BENEFITS OF BASE STATION CLOUD

There are many benefits to the institution of a base station cloud which include the following:

1. For Large Data Collection
2. For Decision Making Capabilities
3. To Solve Optimization Problems

These benefits cannot be easily implemented without the formation of a cloud. Figure 1 shows the proposed cloud network and the communication of its nodes. The benefits of this are better defined below:

II.1. For Large Data Collection

Data collection is the cornerstone of a WSN. The institution of a cloud base station makes data collection much easier as the large computations required can be distributed across the cloud to the many nodes for quicker, more efficient work. For this purpose we have proposed the incorporation of Hadoop/MapReduce.

II.2. For Decision Making Capabilities

The base station cloud allows for the ability of decision-making distributed across the multiple nodes. For instance, in a wireless sensor network, when large amounts of data are gathered from motes by the base station cloud it takes a large amount of computational power to sort and become useable. Distributing this work across the cloud, again, fulfills this need. For this purpose we have proposed the incorporation of Hadoop/MapReduce.

II.3. To Solve Optimization Problems

Decision-making can be done by solving an optimization problem. Again, solving optimization problems needs a large amount of computational power. The base station cloud again solves this issue. Solving an optimization problem, such as the most efficient path when a mote shouts for help (in need of power), is solved much more efficiently through distributed computing across the cloud.

II.4. Related Work

According to Gandham *et al.*, by employing multiple base stations "we have effectively either reduced or

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

retained the hop count of each sensor node in the network" [1].

This is most important from the standpoint of energy consumption since the energy expenditure of a message from its sensor node to its nearest base station is directly proportional to the number of hops it must travel. Thus, the institution of multiple base stations reduces the energy consumption per message.

III. Benefits Of A Mobile Base Station

There are many benefits of a mobile base station in a mobile or non-mobile wireless sensor network. The following three are most pertinent to this research: (1) communication; (2) to build a mobile cloud; and (3) for decision making capabilities

The addition of a mobile base station gives way to the ability of the nodes of the base station to physically move to different areas throughout the network. These benefits are better defined below:

III.1. Communication

Communication between motes and the base station(s) is essential in a WSN. The proposed mobile base stations increases the communication due to the ability to physically move to the most beneficial position.

For instance, when a mote gets disconnected from the network and the network recognizes this, a node of the base station can move to the motes last known location (as determined by the network) and search – based on a signal emitted by the mote – and physically bring the mote back in range of the network and reestablish the connection.

III.2. To Build A Mobile Cloud

With the introduction of multiple mobile nodes comes the ability of a mobile cloud. In the case of a Mobile Wireless Sensor Network (MWSN) where the motes of the network can move it is necessary for the base station, in this case the network of mobile nodes that creates a cloud, to also be mobile.

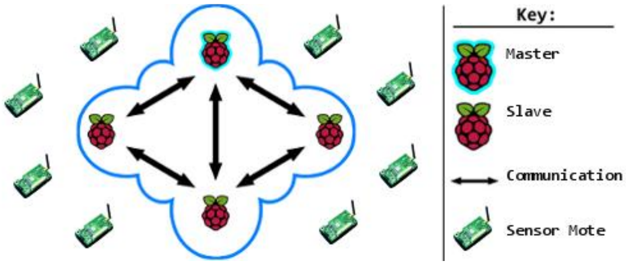


Figure 1. Proposed Cloud Network

As the motes of the MWSN move the mobile cloud base station would move with it. Also, the cloud base station could move to the more beneficial position (in terms of efficiency).

For instance, it might be more beneficial for more nodes of the cloud base station to be in one area of the network for more efficient data collection.

III.3. For Decision Making Capabilities

The institution of a mobile cloud base station can increase the network lifetime by making decisions based on the location of the base station nodes and their ability to move. By determining the location of each node we can determine the most efficient path of movement to assist throughout the network.

We have proposed the institution of nodes that will work in conjunction with a WiTricity charging protocol to solve an optimization strategy to extend the lifetime of the network.

III.4. Related Work

It has been show by [2] that sensor nodes that are only one hop away from a base station consume more energy as they essentially become the main routing node – forwarding not only their own messages but also those of the others belonging to that base station. This depletion of energy renders the node and others like it in the network in operational that in turn does the same to the network.

[1] suggests two strategies for periodically changing the locations of the base station nodes to increase the lifetime of the sensor network.

IV. How To

Each node of the mobile base station cloud is most significantly comprised of one Raspberry Pi and one Lego Mindstorms NXT robot connected through the NXT's provided USB A-to-B cord.

We first set up the Raspberry Pi with a fresh distribution of Raspbian and install the Java Development Kit. We then begin developing the cloud network by installing and setting up `hostapd` and `isc-dhcp-server`. We then assign a static IP address to the Wi-Fi adapter. Next we configure the Access Point and configure Network Address Translation. Finally, we set these up as daemons to run automatically on boot.

We also flash the Lego Mindstorms NXT brick so that we can program it in Java. The NXT is connected to the Raspberry Pi through USB that allows us to control it from the Pi.

V. Conclusions

In this paper, we proposed a mobile base station cloud and developed it with multiple Raspberry Pi powered Lego Mindstorms NXT robots.

As a next step, we will institute Hadoop and MapReduce to let the mobile BS cloud solve a linear problem, which will be useful to find out an optimal solution on its decision making.

References

[1] S.R. Gandham, M. Dawande, R. Prakash and S.Venkatesan. Energy Efficient Schemes for Wireless Sensor Networks With Multiple Mobile Base Stations. *Globecom 2003*, pages 377-381.

[2] A. Manjeshwar and D.P. Agrawal. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. *Intl. Proc. of 15th Parallel and Distributed Processing Symp.*, pages 2009 – 2015, 2001.

SESSION

LATE BREAKING PAPERS AND POSITION PAPERS: WIRELESS NETWORKS AND RELATED ISSUES

Chair(s)

Prof. Hamid R. Arabnia

Copyright © 2014 CSREA Press, ISBN: 1-60132-278-X; Printed in the United States of America

Simulation of Reactive Routing Protocols in Wireless Mesh Networks: a Systematic Literature Review

S. J. Bachega[1], **and D. M. Tavares**[2]
[1]Production Engineering Department, Federal University of Goiás, Catalão, Goiás, Brazil
[2]Computer Science Department, Federal University of Goiás, Catalão, Goiás, Brazil

Abstract - *Wireless Mesh Network (WMN) has been considered as an important architecture for future wireless communications. WMN uses routing protocols to provide communications in wireless environment. Among the categories of routing protocols are proactive (table-driven), reactive (on-demand) and hybrid. The aim of this paper is to initiate a systematic review in order to collect and compare existing experimental evidences in simulation of WMN routing protocols. This study focuses in reactive (on demand) routing protocols. We identified 11 studies that contribute to the objective of this research. This review does establish a baseline, which we wish will be used by other researchers. For future work, it is envisaged the simulation of reactive routing protocols in a real WMN deployment.*

Keywords: routing protocols; reactive routing; wireless mesh network; simulation; systematic literature review.

1 Introduction

Wireless networks currently are part of our life. Short-range radio technologies (e.g. Wi-Fi) and Bluetooth are widespread; the number of cell (mobile) phone subscriptions; innovative technologies based on short-range wireless communication and miniaturized sensor devices have lately been standardized (or are in its final steps towards standardization) and so on [1].

The provision of mobile terminals with versatile communication capacity can be achieved by an effective solution: wireless mesh network (WMN). WMN has emerged as an important architecture for the future development of wireless communications [2].

A WMN is a self-configuring network of nodes interconnected using wireless links. WMNs represent the logical extension of WLANs. They provide high-speed seamless connectivity to nomadic and mobile users [3]. The paradigm of WMN arose from the concept of ad hoc wireless networks [4].

Wireless mesh, with respect to its Basic Service Set (BSS) and Extended Service Set (ESS), is a functionally similar to the IEEE 802.11 infrastructure network standard. Therefore, the mesh Access Points (APs) are called Mesh Points (MPs) and the stations are called relay competence. The Wireless Distribution System (WDS) uses an extension of the IEEE 802.11 MAC/PHY to provide a protocol for auto configuring paths between MPs in a multi-hop topology, supporting unicast, multicast and broadcast traffic [5].

WMNs can use the following routing protocol categories: proactive (table-driven), reactive (on-demand) and hybrid. In proactive routing protocols, nodes attempt to compute a priori and maintain consistent, up-to-date routing information to all nodes in the network, regardless of whether the routes are being used for carrying packets. A prerequisite in reactive protocols is an on-demand process for discovering routes. A path discovery is triggered asynchronously when there is a need for data packet and no path to an intended node is known. The hybrid routing approach mixes the behavior of proactive and reactive routing protocols. This hybridization is done in amounts that best match the efficient operation over a “wide range of conditions” [6].

Among the routing protocols designed to provide communication in wireless environments are Landmark Ad Hoc Routing (LANMAR), Optimization Link State Routing (OLSR), Destination Sequenced Distance Vector Routing (DSDV), Dynamic MANET On-demand (DYMO), Ad hoc On-demand Distance Vector (AODV), Dynamic Source Routing (DSR), Zone Routing Protocol (ZRP) etc. Researchers have been comparing some set of these routing protocols to evaluate their performance [7, 8, 9, 10, 11].

In order to map the existing knowledge in the field, the purpose of this paper is to initiate a systematic literature review in order to collect and compare existing experimental evidences in simulation of WMN routing protocols. This study focuses in reactive (on demand) routing protocols.

A Systematic Literature Review (SLR) is a way of identifying, evaluating and interpreting available researches significant to a research question, or phenomenon of interest, or topic area [12].

This paper is organized as follows: Section 2 describes the method, in section 3 we present the results (answers to the research questions) and the conclusions are shown in section 4.

2 Method

This paper has been undertaken as a systematic literature review based on the guidelines proposed by [12]. Other authors that developed systematic literature reviews are [13, 14, 15, 16, 17]. The steps followed are described below.

2.1 Research questions

The goal of this review is to find studies that contribute to understand how to conduct a simulation study to evaluate reactive routing protocols. The research questions stem from a research project called Mobile mEsh Network to Aid in CountEring drug TRAffiCKing (M.E.N.A.C.E-TRACK), which aims for the creation of a dynamic mesh network, intended to interconnect field personnel (e.g. in vehicles or on foot) to a base of operations (e.g. a police station) whenever possible. The questions are:

RQ1) How much simulation studies have been done to evaluate only reactive routing protocols in wireless mesh networks?

RQ2) What simulation techniques are used to evaluate reactive routing protocols?

RQ3) What performance measures are used to evaluate the reactive protocols in simulation studies?

R4) Can protocol A be shown to be superior to protocol B based on experimental results?

2.2 Search process

We searched widely in electronic sources, as recommended by [12], in order to gain a broad perspective. The advantage of searching databases is motivated by [18]. The following databases were covered: IEEE Xplore, Compendex, ScienceDirect, ACM Digital Library and Springer LNCS. These databases, as confirmed by [14, 19], cover the most relevant journals and proceedings of conferences and workshops within computer science.

The search covered the period from 1990 to 2014 to ensure that most relevant research within the field would be included. The key words used were: wireless mesh network, mesh networking, simulation, evaluation, reactive routing, routing protocols and reactive routing protocols. These key words were used alone and with possible combination.

2.3 Inclusion and exclusion criteria and quality assessment

We adopted the following inclusion criteria in making decisions about whether to include a paper in this study:

- Primary studies published in journals and proceedings of conferences and workshops, representing simulation of reactive routing protocols in WMN.
- The discussion results and analysis must be focused only on reactive routing protocols in WMNs.

The exclusion criteria were:

- Overlapped paper issued by the searches in the selected sources.
- Secondary Literature Reviews.
- Duplicate publications of the same study.
- Papers without full text on the databases.
- Technical reports, works in progress and some workshop reports ('grey' literature).

We excluded 'grey' literature from analysis because the volume of the studies included in the first searches would have grown unreasonably and the quality of these literatures is more difficult to assess [14].

Also, we consider only the most complete version of a study when several reports of the study exist in different journals (duplicate publications of the same study). The exclusion based on languages should be avoided [12], however, only studies written in English are included [14].

2.4 Data Extraction and Synthesis

The data collected from each study were:

- The source and full reference.
- The type of simulated scenario: real application or hypothetical scenario.
- The reactive protocols in WMNs simulated in the study.
- The simulation technique and the performance measures used.
- The main results about the comparison of the protocols simulated.

The data extraction was conducted by one research and verified by another. We discussed the issues when there was a point of disagreement. Allocation was based on the time availability of each researcher. This procedure is advocated by [20, 21]. The data was tabulated to show the items described above.

3 Results

This section summarizes the results of this research.

3.1 Answer to Research Question 1 (RQ1)

We found 11 papers that evaluate only reactive routing protocols in wireless mesh network. TABLE I shows Publication ID, authors, date and source of the publications. The earliest study actually included was published in 1999. Fig. 1 presents the frequency of papers per year. As can be seen, most of the papers in this issue were published in 2011 (5 studies).

TABLE I. DATA FOR RESEARCH QUESTION 1

ID	Author	Date	Source
P1	Ashraf, Abdellatif and Juanole [22]	2011	Computer Communications
P2	Ashraf [23]	2012	IEEE Symposium on Computers & Informatics
P3	Chen, Xiang, Dong and Lu [24]	2011	Wireless Pers Commun
P4	Fu and DaSilva [25]	2007	IEEE Military Communications Conference
P5	Ghahremanloo [26]	2011	International Siberian Conference on Control and Communications
P6	Guo, Peng, Wang, Jiang and Yu [27]	2011	Computers and Electrical Engineering

P7	Lee, Gerla and Chiang [28]	1999	IEEE Wireless Communications and Networking Conference
P8	Lima, Silva, Santos and Pujolle [29]	2008	IEEE Global Communications Conference, Exhibition & Industry Forum
P9	Loscrì [30]	2008	Journal of Networks
P10	Pal and Nasipuri [31]	2011	Pervasive and Mobile Computing
P11	Rahman, Azad, Anwar and Abdalla [32]	2009	International Conference on Future Networks

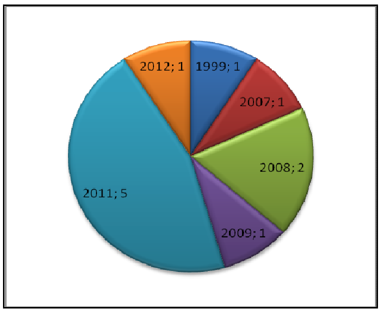


Fig. 1. Frequency of Papers per Year

3.2 Answer to Research Question 2 (RQ2)

It was found that in approximately 73% (8 papers) of the analyzed papers, the simulator used for modeling, simulating and evaluating the WMNs is NS-2. Also, we observed that all papers constituted hypothetical scenarios. In other words, the scenarios were not modeled based on real data for real deployment of WMNs. The most widely used networking standard was the IEEE 802.11 (TABLE II).

TABLE II. DATA FOR RESEARCH QUESTION 2

ID	Type of simulated scenario	Network standard	Simulator
P1	Hypothetical	IEEE 802.11	NS-2
P2	Hypothetical	IEEE 802.11	NS-2
P3	Hypothetical	IEEE 802.11	NS-2
P4	Hypothetical	IEEE 802.11, TDMA	NS-2
P5	Hypothetical	IEEE 802.11s	NS-2
P6	Hypothetical	IEEE 802.11	OPNET
P7	Hypothetical	IEEE 802.11	GloMoSim
P8	Hypothetical	IEEE 802.11, IEEE 802.11b	NS-2
P9	Hypothetical	IEEE 802.16	NS-2
P10	Hypothetical	IEEE 802.11	NS-2
P11	Hypothetical	IEEE 802.11s	QualNet/GloMoSim

3.3 Answer to Research Question 3 (RQ3)

The most simulated reactive routing protocol in WMN was AODV, which was analyzed on 9 studies (approximately 82% of the papers). The AOMDV and DSR protocols were the second most analyzed (Fig.2).

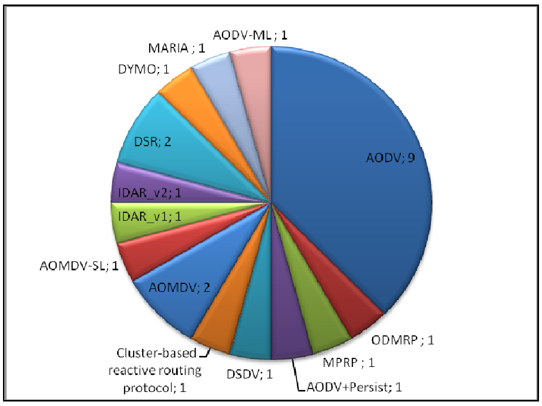


Fig. 2. Frequency of Simulation of the Reactive Routing Protocols

Throughput, end-to-end delay, jitter, packet delivery ratio and number of breakages were the most used among the performance measures to evaluate the reactive protocols in WMN (TABLE III). It is noteworthy that among the studies analyzed, there was a proposal for the use of multi-fractal characteristics as a performance measure.

TABLE III. DATA FOR RESEARCH QUESTION 3

ID	Reactive protocols in WMN simulated	Performance measures
P1	AODV, AODV-ML (Multi-link extension to AODV)	Throughput, Delay, Route Lifetime, Routing Overhead, No. of Breakages, Packet Delivery Ratio
P2	AODV, AODV+Persist	Total Route Breakages, Average Route Duration, Throughput, End-to-end delay
P3	AODV, DSDV	Multi-fractal characteristics
P4	AODV, Cluster-based reactive routing protocol	Routing overhead, route failure, and throughput
P5	AODV, AOMDV	Network Throughput, End to End Delay, Data Drop
P6	AODV, DSR	Routing discovery time, Average number of hops per route, Network delay, Network throughput
P7	ODMRP (On-Demand Multicast Routing Protocol)	Packet Delivery Ratio, Number of Control Bytes Transmitted per Data Byte Delivered, Number of Data and Control Packets Transmitted per Data Packet Delivered
P8	AODV, AOMDV (On-demand Multipath Distance Vector Routing in Ad Hoc Networks), AOMDV-SL	Misbehavior drop ratio, Packet delivery ratio, End-to-end delay of data packets

P9	MPRP (Multipath parallel routing protocol MPRP)	Throughput, Average end-to-end data packet delay, Data Packets Lost
P10	AODV, IDAR_v1, IDAR_v2 (Interference and delay aware routing), MARIA (Mesh Admission control and qos Routing with Interference Awareness)	Packet Delivery Ratio, Traffic Delay, Jitter
P11	AODV, DSR, DYMO	Packet Delivery Ratio, Jitter, Average End-to-end delay, Throughput

3.4 Answer to Research Question 4 (RQ4)

Through the analysis of the selected papers, we observe that it is possible to infer conclusions about the performance of reactive routing protocols in WMN compared. Therefore, it is possible to define a protocol A is greater than a protocol B based on the experimental results. TABLE IV shows the summary of the main results of the analyzed studies.

TABLE IV. DATA FOR RESEARCH QUESTION 4

ID	Main results
P1	Substantial performance improvement compared to classical AODV and local route repair schemes.
P2	AODV+Persist brings a high degree of routing stability to the network and improves performance significantly.
P3	Compared with AODV protocol, traffic with DSDV exhibits more multi-fractal characteristics.
P4	The cluster-based reactive routing protocol focus on improving the routing robustness, handling rapid topology changes and tolerating intermittent links, in contrast to conventional ad hoc routing protocols.
P5	AOMDV protocol is more efficient in the rate of successfully delivered packets in high mobility scenarios. The Routing Protocols work much better in less mobile scenarios.
P6	DSR protocol based on the dynamic source routing is not suitable for wireless transmission, while AODV routing protocol based on the purpose-driven routing is suitable for wireless transmission with rapid change of network topology, since the routing information of DSR will be significantly more than that of AODV.
P7	ODMRP is effective and efficient in dynamic environments and scales well to a large number of multicast members. The advantages are: low channel and storage overhead; usage of up-to-date and shortest routes; robustness to host mobility; maintenance and exploitation of multiple redundant paths; scalability to a large number of nodes; exploitation of the broadcast nature of wireless environments; unicast routing capability.
P8	The increase in the number of paths used by AOMDV and AOMDV-SL results in slight differences for all metrics. Results show a decrease in the impact of routing attacks with minimal performance loss.
P9	MPRP allows better performance to be achieved in terms of throughput and delay in all the schemes, but some of the schemes considered permits to have better performance than the unipath version.
P10	IDAR is effective in improving both the packet delivery and delay performance in multi-hop environments where the route selection is done by a gateway node that has global activity information of the network.
P11	On an average AODV perform better than DYMO and DSR. However, the overall performance of the three protocols in WMNs is not quite good.

4 Conclusions

In this paper, a systematic literature review was done to collect and compare existing experimental evidences in simulation of reactive routing protocols in WMN. Thus, the intended objective was met.

Among the results, we identified 11 studies that contribute to the objective of this research. We found that it is possible and reliable to compare the performance of reactive routing protocols in WMN via simulation. The most simulated protocol was AODV and the most used simulator was NS-2. Performance measures used in the studies were also described. However, the selection of which performance measure to use depends on the expected contribution to the research objective.

We add some further observations about how this review might be interpreted. The first observation is in order to emphasize that so far no systematic literature reviews on the topic covered in this paper were found. This fact confirms the importance of this research. This review does provide such a baseline, which we wish will be used by other researchers.

WMN is still considered an emerging area in which practical issues are still little explored. We note that the simulated scenarios in the analyzed articles are all hypothetical. According to [33], we note that simulation studies need to be complemented by more realistic testing. This testing is important to advance the understanding of WMNs. Also, it allows researchers to observe phenomena that may not be evident in simulator settings.

It is noteworthy that this research is part of a project that aims for the creation of a dynamic mesh network, intended to interconnect field personnel to the base of operations whenever possible. For future work, it is envisaged the simulation of reactive routing protocols in a real deployment of a WMN. Thus, it will be possible to check whether results obtained in simulated hypothetical cases converge with practical results.

Acknowledgment

The authors would like to thank the Research Assistance Foundation of the State of Goiás (FAPEG - Fundação de Amparo a Pesquisa do Estado de Goiás) for the full sponsorship of this research - edict number 006/2012.

References

[1] P. Santi. “Mobility Models for Next Generation Wireless Networks: Ad hoc, Vehicular and Mesh Networks”. JohnWiley & Sons, 2012.

[2] I.F. Akyildiz, X. Wang. “A Survey on Wireless Mesh Networks”. IEEE Communication Magazine 43(9), (2005) S23–S30.

[3] N. Scalabrino. “Performance of Wireless Mesh Networks: Experimental Studies of IEEE 802.11 and IEEE 802.16 solutions”. Verlag Dr. Müller, 2009.

[4] P. H. Pathak, R. Dutta. “Designing for Network and Service Continuity in Wireless Mesh Networks”. New York: Springer, 2013.

[5] A. Rahman, S. Azad and F. Anwar. “Performance Analysis of on-demand routing protocols in wireless mesh networks”, Informatica Economică, 13 (2) (2009) 120-127.

[6] G. Aggélou. “Wireless Mesh Networking: with 802.16, 802.11 and ZigBEE”. The McGraw-Hill Companies, Inc, 2008.

[7] A. Boukerche. “Performance Comparison and Analysis of Ad Hoc Routing Algorithms”, in International Professional Communication Conference (IPCCC 2001), 2001, pp. 171-178.

[8] A. Boukerche. “Performance Evaluation of Routing Protocols for AdHoc Ad hocNetworks”, Mobile Net-works and Applications, vol. 9, Kluwer Academic Publishers, 2004, pp. 333-342.

[9] S. Azad, A. Rahman and F. Anwar. “A Performance Comparison of Proactive and Reactive Routing Protocols of Mobile Ad-hoc NETwork (MANET)”, Journal of Engineering and Applied Sciences, vol. 2, no. 5, 2007, pp 891-896.

[10] J. G. Naragund and R. M. Banakar. "Analysis of HWMP-ETX Routing in Wireless Mesh", in. International Conference on Advanced Computing, Networking and Security, pp. 208-213, 2013.

[11] A. Aggarwal, S. Gandhi, N. Chaubey and K. A. Jani. "Trust Based Secure on Demand Routing Protocol (TSDRP) for MANETs", in. International Conference on Advanced Computing & Communication Technologies, pp. 432- 438, 2014.

[12] B. Kitchenham. “Procedures for Performing Systematic Reviews”, Joint Technical Report, Department of Computer Science, Keele University (TR/SE-0401) and National ICT Australia Ltd. (0400011T.1), 2004.

[13] G. S. Walia and J. C. Carver. “A systematic literature reviem to identify ans classify software requirement errors”, Information and Software Technology, 51 (2009) 1087-1109, doi: 10.1016/j.infsof.2009.01.004

[14] E. Engström, P. Runeson and M. Skoglund. “A Systematic Review on Regression Test Selection Techniques”, Information and Software Technology, 52 (1) (2010) 14-30, doi: 10.1016/j.infsof.2009.07.001

[15] D. Budgen, A. J. Burn, O. P. Brereton, B. A. Kitchenham and R. Pretorius. “Empirical evidence about the UML: a systematic literature review”, Softw. Pract. Exper., 41 (2011) 363-392, doi: 10.1002/spe.1009

[16] E. Hossain, M. A. Babar and H-y. Paik. "Using Scrum in Global Software Development: A Systematic Literature Review", in. IEEE International Conference on Global Software Engineering, pp. 174-184, 2009.

[17] T. Dybå and T. Dingsøyr. “Empirical studies of agile software development: A systematic review”, Information and Software Technology, vol 50, n. 9-10, pp. 833-859, 2008.

[18] O. Dieste and A. G. Padua. “Developing Search Strategies for Detecting Relevant Experiments for Systematic Reviews”, in 1st International Symposium on Empirical Software Engineering and Measurement, pp. 215-224, 2007.

[19] T. Dybå, T. Dingsøyr and G. K. Hanssen. “Applying Systematic Reviews to Diverse Study Types: An Experience Report”, in 1st International Symposium on Empirical Software Engineering and Measurement, pp. 225-234, 2007.

[20] O. P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner and M. Khalil. “Lessons from applying the systematic literature review process within the software engineering domain”, Journal of Systems and Software, 80 (4) (2007) 571-583.

[21] B. A. Kitchenham et al. “Systematic literature reviews in software engineering – A systematic literature review”, Information and Software Technology, 51 (2009) 7-15.

[22] U. Ashraf, S. Abdellatif and G. Juanole. “Route Maintenance in IEEE 802.11 wireless mesh networks”, Computer Communications 34 (2011) 1604–1621.

[23] U. Ashraf. "Persist: Mitigating Route Breakages in Wireless Mesh Networks", in IEEE Symposium on Computers & Informatics, pp. 97-100, 2012.

[24] Y. Chen, Z. Xiang, Y. Dong and D. Lu. “Multi-Fractal Characteristics of Mobile Node’s Traffic in Wireless Mesh Network with AODV and DSDV Routing Protocols”, Wireless Pers Commun, (2011) 58:741–757.

[25] B. Fu and L. A. DaSilva. "A Mesh in the Sky: a Routing Protocol for Airborne Networks", in IEEE Military Communications Conference (MILCOM), pp. 1-7, 2007.

[26] P. Ghahremanloo. "Multi-Path Routing Challenging Single-Path Routing in Wireless Mesh Networks: Network Modeling of AODV and AOMDV", in International Siberian Conference on Control and Communications (SIBCON), pp. 12-15, 2011.

[27] L. Guo, Y. Peng, X. Wang, D. Jiang, Y. Yu. "Performance evaluation for on-demand routing protocols based on OPNET modules in wireless mesh networks". Computers and Electrical Engineering, 37 (2011) 106–114.

[28] S-J. Lee, M. Gerla and C-C. Chiang. "On-Demand Multicast Routing Protocol", in IEEE Wireless Communications and Networking Conference (WCNC), 1298 - 1302 vol.3, 1999.

[29] M. N. Lima, H. W. Silva, A. L.Santos, G. Pujolle. "An Architecture for Survivable Mesh Networking", in. IEEE Global Communications Conference, Exhibition & Industry Forum (GLOBECOM), pp. 1-5, 2008.

[30] V. Loscrì. "On the interaction between multiple paths and Wireless Mesh Networks scheduler approaches", Journal of Networks, vol. 3, no. 7, 2008.

[31] A. Pal, A. Nasipuri. "A quality based routing protocol for wireless mesh networks", Pervasive and Mobile Computing, 7 (2011) 611–626.

[32] A. Rahman, S. Azad, F. Anwar and A. H. Abdalla. "A Simulation Based Performance Analysis of Reactive Routing Protocols in Wireless Mesh Networks", International Conference on Future Networks (ICFN), 268 - 272, 2009.

[33] H. Lundgren et al. "Experiences from the Design, Deployment, and Usage of the Ucsb Meshnet Testbed", IEEE Wireless Communications, vol. 13, no. 2, pp. 18-29, april 2006.

Intelligent and Secure Routing in Heterogeneous Wireless Networks

[1]Suman , [2] Parvinder Singh, [3]R. B. Patel

[1,2] Deptt. of Computer Sc. & Engg.,
Deenbandhu Chhotu Ram University of Sc. & Tech., Murthal, Sonipat, Haryana (INDIA).
[3] Deptt. of Computer Sc. & Engg.,
G. B. Pant Engineering College, Ghurdauri-246194 (Pauri-Garhwal), Uttarakhand, India

Abstract - *Dynamically changing environment of heterogeneous wireless networks (HWN) is more vulnerable to security threats. Finding effective solutions to deal with attacks in HWN is a challenging task. To reduce packet capturing and ensure greater protection from outside attacks in HWN, a multi parameter algorithm PAIRS (Periodic Adaptive and Intelligent Route Selection) is presented in this paper. PAIRS is intelligent and adaptive and it selects best route using ANFIS (adaptive neuro-fuzzy inference system) periodically. It effectively balance overall load of network and prevents denial of service (DoS) attack by providing improved resource management.Results obtained and evaluated based on a number of parameters. The results show that PAIRS successfully prevent packet capturing and provide better network performance at the same time.*

Keywords: Security, heterogeneous wireless networks (HWN), intelligent, adaptive, route selection, adaptive neuro-fuzzy inference system (ANFIS).

1 Introduction

Heterogeneous wireless networks (HWN) require new concepts and approaches to deal with the challenges posed by integration of technologies. Due to this reason, HWN have gained much attention of researchers over last few years. HWN are inherently more prone to packet capturing. This feature of HWN can be exploited by attackers to make network incapable of providing normal services.Many security threats can cause unexpected service interruption and disclosure of information in heterogeneous 4G network. The reality of ever increasing security threats such as intrusion detection, secure routing, key establishment and distribution, and authentication could affect the feasibility of HWN. Although many researchers are designing new security architectures for 4G, but still much more need to be done.

Neural network classifiers and fuzzy logic systems are good candidates for pattern classifiers due to their non-linearity and generalization capabilities. Fuzzy logic can represent human knowledge as fuzzy rules and can be used to develop cost-effective approximate solutions [1]. A neural network (NN) is a massively parallel distributed processor made up of simple processing units, which can store experimental knowledge for future use.

In this paper, we present a multi parameter based algorithm - PAIRS (Periodic Adaptive and Intelligent Route Selection) which is robust against a variety of challenges posed by HWN. It uses designed and trained ANFIS (adaptive neuro-fuzzy inference system) to select the best available route periodically. PAIRS is intelligent, adaptive and reduces congestion in network. This algorithm effectively balances overall load of the network and prevents DoS attack by offering improved resource management. Performance of PAIRS has been analyzed in terms of network throughput. Results are compared with existing algorithms to emphasize the significance of PAIRS.

2 Literature Review

Many secure routing techniques have been proposed so far but none of them provides a comprehensive solution for handling adaptive and ever increasing security threats in HWN. Previous research on security issues in HWN can be divided into three main categories, namely, authentication, collaboration incentives, and denial of service (DoS) prevention (summarized in [2]).

DoS prevention falls into two categories: improved resource management [3] and avoidance mechanisms [4]. For low-capacity networks (eg. ad-hoc and sensor) improving resource management mechanisms can help reduce the disparity when they are connected to high-capacity networks (eg. wired and fiber-optic), thus leading to increased heterogeneity in the network. Increased congestion and saturation in heterogeneous networks leads to more and more DoS disruptions, which increases with increase in disparity of resources between different parts of the network. Therefore, DoS prevention must be addressed explicitly in context of the next generation networks.

To deal with DoS, a probabilistic route selection algorithm that traces attacker's real origin is presented in [5]. Quality of service (QoS) aware path selection scheme to estimate required bandwidth ratio is presented in [6]. This ratio is based on the QoS requirements of target service and SINR of each path. The proposed scheme can select multiple/single optimal path to satisfy the QoS requirements among dynamically changing HWN.

Two distinct models to prevent wireless DoS and replay attacks by protecting the control packets are presented in [7].

In [8], the concepts of Learning Automata (LA) are applied on on the existing Optimized Link State Routing protocol (OLSR) to protect the network from distributed DoS attack. However nothing is said about performance of the proposed protocol when the number of nodes in the network is large.

In [9], NN based optimal path selection algorithm is presented for managing multihomed hosts attached to HWN. In [10], an adaptive and efficient routing protocol for integrated cellular and ad-hoc heterogeneous network with flexible access (iCAR-FA) is presented. [10] also presents detailed numerical analysis on route request rejection rate.

In [11], novel load-aware route selection algorithm, (LARS) is presented. In this approach each mesh node is allowed to distribute the traffic load among multiple gateways for uniform utilization of Internet connections leading to improved network capacity. However LARS design does not consider end-to-end delays in the selection of feasible network paths.

In [12], heterogeneity of nodes and delay is considered during route discovery to discover resource-rich routes. Paths that contain more capable nodes are utilized, thereby avoiding resource-poor nodes. This paper does not focus on the issue of choosing proper delay value for unknown network.

In [13], a generic and scalable security management service which scales with increasing diversity of network techniques and applications is proposed for inter domain communication of heterogeneous networks.

In [14], a flexible cryptography-based approach is presented for establishing trustworthiness between multi-hop mobile nodes using infrastructure supported authentication. Generalized multi-hop security protocol (GMSP) combines mobile IP and ad hoc security schemes to achieve an effective route discovery protection in accordance with anti-integrity, impersonation for generalized heterogeneous multi-hop networks. This protocol implements security in four steps namely registration for single hop mobile nodes, registration for multi-hop mobile nodes, routing security between base station and any mobile node, and security of the routing between any two mobile nodes.

In [15], modified destination sequenced distance vector (MDSDV) protocol is presented that uses the bandwidth and hop count as the routing metric for selecting appropriate routes in wireless mesh networks. It works very well for wireless networks having large number of nodes.

3 Periodic Adaptive and Intelligent Route Selection

Attacks on HWN have become much more adaptive with passage of time. To deal with these attacks, networks should be able to adjust their security provisions and sophistication levels rapidly and intelligently. Forwarding packets on same route might increase the risk of packet capturing. PAIRS reduce packet capturing by forwarding packets on different routes periodically. Since the route for packets is not known in advance, attacker can not launch either an active or passive attack.

3.1 Tools Used

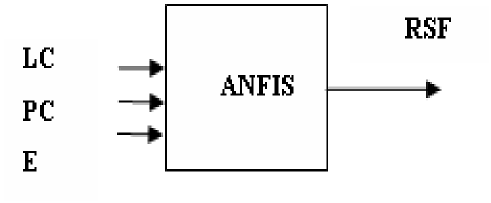


Figure 1. Designed ANFIS for PAIRS

ANFIS with three inputs and one output is designed to rank different routes based on value of route selection factor (RSF). Training data is carefully chosen for tuning of training vector of ANFIS in Fig. 1, so that it can well identify parameters and rules and give reasonable performance. We have used hybrid optimization method and grid partitioning for FIS generation. Set of 27 rules is used in ANFIS to calculate output. PAIRS use ANFIS to select route with highest RSF for packet transmission at any point of time.

3.2 Inputs Parameters for ANFIS

1) LC (Current link capacity of link under consideration)-It gives the current available bandwidth of link under consideration. Lower value of LC signifies that large amount of traffic is currently passing through this link. Further increase in number of packets may lead to increase in congestion and decrease in throughput of network. So, higher the value of LC more is the probability of that route being selected for packet forwarding. In designed ANFIS, range of LC is taken from 0 to 1.

$$LC = (Bt - Bu)/Bt \qquad (1)$$

Where Bt = total bandwidth of network

Bu=used bandwidth of the link under consideration at given point of time

Even if link with low value of LC has two nodes on endpoint with high value of PC and E, choosing that link as part of route may degrade network throughput as it will put additional traffic load on the slow-speed link. It asserts that a link cannot transfer data faster than its remaining bandwidth. So LC has maximum impact on RSF and network throughput.

2) PC (Current processing capacity of node under consideration)-It is related to current CPU usage of node under consideration. PC is the measure of remaining processing capacity of node under consideration. Lower value of PC signifies that much of CPU capacity is currently being used for processing packets. Any further addition of packets to CPU may lead to congestion and this might degrade overall performance of network. So route having higher PC will be selected. In designed ANFIS, range of PC is taken from 0 to 100.

$$PC = (100 - CPUu) \qquad (2)$$

Where CPUu =current CPU usage

3) E (Current power of node under consideration)-It measures current battery backup of node under consideration. Lower value of E signifies that node is running out of power. So it would not be appropriate to use that node for packet forwarding. In designed ANFIS, range of E is taken from 0 to 100.

In ANFIS, range of RSF is taken from -9.558 to 7.331

3.3 PAIRS Algorithm

PAIRS is run in parallel on all nodes in the network that have some data to transmit

Single iteration of PAIRS is given below:

1) Node i that want to transmit packets will calculate RSF using ANFIS for next hop j on all available routes to destination, where j= 1, 2, 3….D.

2) Route with highest RSF is selected for packet transmission by current node i.

3) Current node i resets its packet counter Pi=0.

4) Node i transmit packets.

5) If Pi <=NP go to step 4, else go to step 1.

Detailed description of PAIRS is given in [16].

4 Performance Comparison

The performance of PAIRS is compared with MDSDV [15], in terms of two metrics – packet delivery ratio (PDR) and end to end delay. The effect of increasing number of nodes on packet delivery ratio and end-to-end delay has been anlyzed in next sections. Number of nodes will vary from 10, 20, 30, 50 to 100 in a 1000*1000m network topology.

4.1. Packet delivery ratio (PDR)

Packet delivery ratio is the ratio of packets received to packets sent. PDR can reveal the network throughput and the efficiency of the network resource usage.

4.2. End to end delay

The end-to-end delay can be defined as the total time to deliver all data packets from source to destination.

5 Results and Discussion

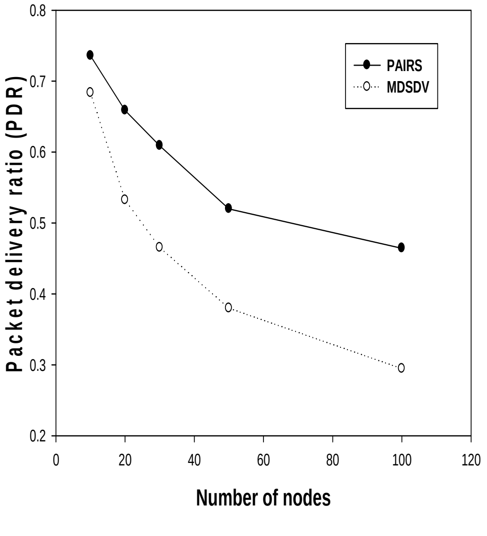


Figure 2. Variation in packet delivery ratio with number of nodes

Fig. 2 shows the effect of varying number of nodes on PDR for PAIRS and MDSDV .It is very much clear that PDR of PAIRS is much higher than that of MDSDV. This is due to the reason that PAIRS effectively balance overall network load and controls congestion. This eventually increases network throughput. As the node number increases, the performance of the PAIRS drops a little but is still higher than MDSDV.

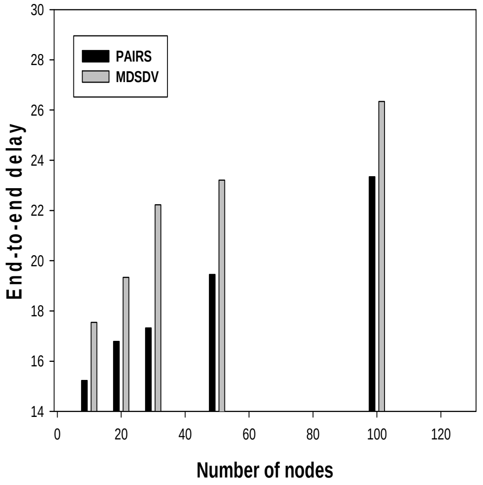


Figure 3. Variation in packet delivery ratio with number of nodes

Fig. 3 shows the effect of varying number of nodes on end to end delay for PAIRS and MDSDV.

Although end to end delay increase with increase in number of nodes for PAIRS, but it is always lesser than MDSDV. This is due to the reason that as number of nodes increase PAIRS has to be run more number of times to find best available route. But at the same time congestion is further reduced due to availability of more routes for diversion of traffic.

6 Conclusions

In this paper a secure route selection algorithm PAIRS has been proposed for HWN. PAIRS is robust against packet capturing and can work well for networks having large number of nodes. It is an intelligent multi-criteria route selection algorithm and selects best route among the available. This effectively balance overall load of the network and prevents DoS attack by offering improved resource management. RSF considers the link capacity of route under consideration, leading to reduction in network congestion. PAIRS can alter its decision to incorporate changes in link and node parameters, making it adaptive at the same time. The performance of PAIRS has been compared with MDSDV and results obtained are far better than MDSDV.

7 References

[1] R. Mikut, J. Jakel, L. Groll. "Interpretability Iissues in Data Based Learning of Fuzzy Systems", In Elesevier Journal of Fuzzy sets and systems, 2004.

[2] J. B. Evans, W. Wang, B. J. Ewy. "Wireless Networking Security: Open Iissues in Trust, Management, Interoperation and Measurement", In International Journal on Security and Networks, Vol. No.1, Issue No. 2, 84-94, 2006.

[3] R. Bhatia, L. E. Li, H. Luo, R. Ramjee, P. Sanjoy. "Icam: Integrated Cellular and Ad-hoc Multicast", In IEEE Transactions on Mobile Computing, Vol. No.5, Issue No. 8, 2006.

[4] W. Enck, P. Traynor, P. McDaniel, T. L. Porta. "Exploiting Open Functionality in SMS-Capable Cellular Networks", In Proceedings of the 11th ACM conference on Computer and communications security (CCS '05) New York, USA, 2005.

[5] H. Yim, T. Kim, J. Jung. "Probabilistic Route Selection Algorithm to Trace DDoS Attack Traffic Source", In Proceedings of IEEE International Conference on Information Science and Applications. Korea, 706-713, 2011.

[6] Shin-Hun Kang , Jae-Hyun Kim. "QoS-Aware Path Selection for Multi-homed Mobile Terminals in Heterogeneous Wireless Networks", In Proceedings of 7th IEEE/ACM conference on Consumer communications and networking conference, 2010.

[7] Malekzadeh et al, "Protected Control Packets to Prevent Denial of Services Attacks in IEEE 802.11 Wireless Networks", In EURASIP Journal on Information Security, Vol. No. 4, doi:10.1186/1687-417X-2011-4, 2011.

[8] Sudip Misra , P. Venkata Krishna , Kiran Isaac Abraham, Navin Sasikumarb, S. Fredun, " An Adaptive Learning Routing Protocol for the Prevention of Distributed Denial of Service Attacks in Wireless Mesh Networks", In Elsevier Journal of Computers and Mathematics with Applications, Vol. No. 60, 294-306, 2010.

[9] Zifan Li , Tao Zhang, Chong Shen "Optimal Path Selection for Multihomed Heterogeneous Wireless Networks", In Proceedings of IEEE Second International Conference on Intelligent Systems, Modelling and Simulation (ISMS). (Kuala Lumpur, 344 – 349, 2011.

[10] Yumin Wu, Kun Yang, Jie Zhang. "An Adaptive Routing Protocol for an Iintegrated Cellular and Ad-hoc Network with Flexible Access", In Proceedings of ACM international conference on Wireless communications and mobile computing, 2006.

[11] Raffaele Bruno, Marco Conti, Antonio Pinizzotto. "A Queuing Modeling Approach for Load-Aware Route Selection in Heterogeneous Mesh Networks", In Proceedings of IEEE international symposium on world of wireless, mobile and multimedia networks and workshops, 1-9, 2009.

[12] Ian D. Chakeres and Elizabeth M. Belding-Royer. "Resource Biased Path Selection in Heterogeneous Mobile Networks", UCSB Technical Report, 2003.

[13] Rohrer, Justin P. Sterbenz, James P.G. Wang, Weichao. "Homogeneous Security in Heterogeneous Networks: Towards a Generic Security Management Protocol", In Proceedings of IEEE Military Communications Conference (MILCOM),1-6, 2007.

[14] Bin Xie, Srinivasan, A.K.S., Agrawal, D.P. "GMSP: A Generalized Multi-hop Security Protocol for Heterogeneous Multi-hop Wireless Network", In Proceedings of IEEE Conference on Wireless Communications and Networking (WCNC), Vol. No.2, Issue No. 3, 634-639, 2006.

[15] Wang Kai, Gao Yang-yang, Guan Jian-feng, Qin Ya-juan. "MDSDV: A Modified DSDV Routing Mechanism for Wireless Mesh Networks", In Elsevier journal of China Universities of Posts and Telecommunications, Vol. No. 18 (Suppl. 2), 34–39, 2011.

[16] Suman, Sukhdip Singh,.Patel R.B., Parvinder Singh. " PAIRS: Algorithm for Secure Heterogeneous Wireless Networks", In Journal of Information Systems and Communication,Vol. No. 3, Issue No. 1, 210-214, ISSN: 0976-8742 & E-ISSN: 0976-8750, 2012.

A Practical Evaluation of Smartphone Application on Mesh Networks

Dalton M. Tavares[1], **Anelisa P. da Silva**[1], **Stella J. Bachega**[2],
Rafael V. Aroca[3], **Jó Ueyama**[4], **Glauco A. P. Caurin**[5],
and Antônio Carlos de Oliveira Jr[1],
[1]Computer Science Department, Federal University of Goiás, Catalão, Goiás, Brazil
[2]Production Engineering Department / Federal University of Goiás, Catalão, Goiás, Brazil
[3]Mechanical Engineering Department / Federal University of São Carlos, São Carlos, São Paulo, Brazil
[4]Computer and Mathematical Science Institute / University of São Paulo, São Carlos, São Paulo, Brazil
[5]Mechanical Engineering Department / University of São Paulo, São Carlos, São Paulo, Brazil

Abstract— *This paper presents a mesh architecture proposal called Mobile mEsh Network to Aid in CountEring drug TRAffiCKing (M.E.N.A.C.E-TRACK). This project was born from the hypothesis we could establish a covert network channel independent of the cell phone companies infrastructures. Therefore, law enforcement agencies could establish connection with field personnel, in a fault tolerant fashion allowing the transmission of multimedia data (instead of only voice). The main contribution for this paper is the strategies involved to configure smartphones on the MANET side of this system. We present the main difficulties and one possible solution to implement ad hoc mode on our testbed so we can enable a MANET organization on M.E.N.A.C.E-TRACK.*

Keywords: OLSR, Android, Mesh networks

1. Introduction

The research project Mobile mEsh Network to Aid in CountEring drug TRAffiCKing (M.E.N.A.C.E-TRACK) proposes the creation of a dynamic mesh network, intended to interconnect field personnel (e.g. in vehicles or on foot) to a base of operations (e.g. a police station) whenever possible. M.E.N.A.C.E-TRACK is organized as a Wireless Mesh Network (WMN), with a particular multi hop ad hoc network consisting of a mesh backbone and mesh clients. The stationary wireless mesh routers (MR) interconnects through single or multi hop wireless links forming the backbone (i.e. the police station and MRs deployed strategically throughout the city). The MRs should have wired connections and act as the Internet gateway (IGW) to exchange the traffic between the Internet and the WMN clients.

The mobile devices can connect to any MRs in reach to access the Internet via the IGW in a multi hop or through any mesh client in reach forming a route to the IGW. In this sense, the behavior of the mobile devices is that of a Mobile Ad hoc Network (MANET), where routes are dynamically established allowing devices to be grouped without any predefined infrastructure. This should account for the field personnel [1]. The scope of this article will be on the MANET devices, specifically, smartphones.

In this scenario, connectivity is a matter so important that governments acknowledge the ability to "be connected" (possibly with Internet access) as a commodity as important as food, shelter or healthcare; not literally as an indispensable element to sustain life, but as a means of offering quality of life to citizens [2]. Therefore, the fundamental research question is: can smartphones perform a more decisive part on the connectivity infrastructure? Considering the number of smartphones in use, we formulate as a hypothesis that it could be used as part of a greater network infrastructure, as an active traffic router, in order to extend the reach of traditional network technologies (e.g. a cellphone tower, an access point etc). The number of devices responsible for connectivity may be an improvement considering the increase in routing diversity and fault tolerance. Therefore, we evaluate the use of smartphones, as potential MANET nodes.

The contributions of this paper relies on the proposal for a method that could be made generic enough for the practical application of MANET routing protocols using smartphones. We also describe the details of our testbed implementation and experience gained during deployment. The rest of this paper covers M.E.N.A.C.E-TRACK application scenario (section 2), related work (section 3), the details of our testbed implementation (section 4) and the conclusion and future work (section 5).

2. Understanding the Scenario

Project M.E.N.A.C.E-TRACK is intended to create a dynamic mesh network, inherently fault tolerant, so data could be transmitted from law enforcement agencies to field agents. Depending on the density of MRs in a city, this transmission could be made with minimal delay, or considering the lack of connected landlines, with variable delay depending on the number of MANET devices used to achieve a mesh node gateway. This is a particularly interesting alternative considering the lack of wireless network coverage quality in Brazil.

According to the Brazilian Network Information Center (NIC.br), none of the Brazilian 3G operators passed quality assurance tests (e.g. TCP/UDP throughput, jitter, latency, packet loss etc) conducted in December 2012 [3]. Preliminary results shows most operators had a packet loss rate per connection 2 percent above the limit when the quality of the Internet connection starts getting compromised. In terms of latency, all major operators in Brazil had values above 200 milliseconds (considered a standard) [4]. Considering the major players in the 3G market today, we have 64.6% of the cities covered, which represents 182,471,019 users and 90.8% of the population [5].

Although the cellphone network could be an alternative to M.E.N.A.C.E-TRACK, this network infrastructure may not survive a devastating earthquake or another natural disaster. Even if the infrastructure stays in place, there can be added overload due to extraordinary circumstances that may render the cellular network useless (e.g. calls from refugees and their families, international aid workers who arrived in the aftermath caused by some unfortunate event etc). Josh “m0nk” Thomas and Jeff “stoker” Robble, both working at Mitre, saw this problem and created a working prototype backup network using only the Wi-Fi chips on Android smartphones [6]. This would be a good staging area for testing MANETs using the approach proposed by Thomas and Robble.

The choice to use mesh networks instead of the traditional cellular or wireless local area networks (WLAN), for the purposes of the proposed system, is based on the following facts [7]: there is no main node, therefore we achieve a certain degree of redundancy innate to the system; it is possible to reach any other node by traversing a number of intermediate nodes, which favors interconnectivity among nodes and a bigger range in some cases; all nodes are equal so there is no centralized control. Therefore, each node participate in networking and as a source or sink of traffic as well; rather than a single hop to a base, multi-hopping/relaying amongst nodes must be a common capability enabling the creation of a new network or the expansion of an existing one. This allows to cope with distance and obstacles by hopping around obstructions; ability to work without infrastructure (e.g. a base station).

3. The Role of Ad hoc Networks on M.E.N.A.C.E-TRACK

The term ad hoc comes from latin and means “for a particular purpose only”. Therefore, an ad hoc network represents a network with its purpose defined for a temporary time frame, such that some network devices can be a part of the network topology only while they are in range or during the communication.

Some applications of mobile computing do not depend on a pre-existing infrastructure and can utilize a MANET. A MANET by definition is a wireless network that does not need a rigid infrastructure and its topology is self-configuring for connected mobile devices [8]. Because of self-configuring and self-organizing characteristics, MANETs can be deployed quickly. There is no infrastructure defined in the network, therefore all of the participating nodes relay packets for other nodes and perform routing if necessary. Because of the limitations in wireless transmission range, communication links could be multi-hop [9]. Although routing protocols are the most important element of a MANET, our discussion will be limited to the establishment of a flexible infrastructure so we can test a broad range of protocols and their features in mobile devices.

3.1 Related Work

For quite some time, universities and research centers have developed and widely used mesh networks as access networks for their users. In this section we present related works that is not new but very similar in nature to our proposal. Some examples that inspired our research are projects RoofNet [10] at the Massachusetts Institute of Technology (MIT), Vmesh [11], [12] at University of Thessaly, Greece, MeshNet [13] at University of California, Santa Barbara (UCSB) etc.

The RoofNet project refers to the deployment of mesh nodes deployed over an area of about four square kilometers in Cambridge, Massachusetts, using volunteer users. These volunteers installed a Roofnet kit at home (PC, an 802.11b card, and a roof-mounted omni-directional antenna) and shared (a fraction) of their DSL lines [14].

VMesh is a low cost and inherent flexible deployment in terms of building a prototype wireless router using an embedded Linux. This was one of the first mesh network projects to use OpenWrt. It also adapted the network configuration for the mesh setup so it would take little to no human intervention. This in turn, was exploited to support the dynamic addition, removal and mobility of network elements. Its network architecture model is also very similar to the one proposed for M.E.N.A.C.E-TRACK and therefore is considered the very inspiration for our proposal.

MeshNet is a 30-node wireless mesh testbed implemented at UCSB. The authors present their experience in designing, deploying and using their mesh network. They also present UCSB MeshNet architecture and discussed the challenges regarding management, nonintrusive and distributed monitoring, and node status visualization. Their implementation were also based on OpenWrt for the mesh nodes [13].

3.2 Routing Protocols for Mesh Networks

Routing can be roughly divided in topology based, location based or energy aware. Topology based has traditionally used the knowledge of instantaneous connectivity of the network with emphasis on the state of network links. Location based uses information related to the physical position of nodes. Energy aware routing, uses information regarding the remaining battery in mobile devices in order to produce paths that comprise nodes with high value of remaining lifetime, as well as to help them adjust their transmission power so that each node keeps the energy required to accomplish the routing task at the minimum [15].

The discussion will be limited to topology based protocols. In this category, the associated routing protocols can be classified as proactive, reactive and hybrid. In proactive protocols all nodes calculate all possible paths to all destinations independently of their effective use such that when a packet needs to be forwarded, the route is already known, eliminating routing delays. The main drawback is the periodic broadcasts sent taking some time to converge (i.e. to create the routing table at each node). In reactive protocols, the network is evaluated if needed. Routes are created only if there is the need to carry data traffic. This protocol exempts the creation of a routing table, scaling well for large populations. Hybrid routing mixes the features of proactive and reactive protocols and is used when there is a set of circumstances where neither protocol perform well [15]. As our testbed is very simple, we chose to use a proactive routing protocol as it works efficiently for a small scale mesh network with high mobility [16].

3.2.1 Optimized Link State Routing Protocol (OLSR)

According to [17], OLSR is a routing protocol for proactive ad hoc networks, developed by Institut National de Recherche en Informatique et en Automatique (INRIA) and standardized by the Internet Engineering Task Force (IETF) in RFC 3626 [18] as an experimental protocol. Its goal is to calculate and maintain routes for every node in a wireless network using a mesh topology. OLSR is able to do that by executing in each node a process that keeps track of network paths for every other node. Therefore, each node fills a routing table that indicates how he can reach every other node and so the algorithm converges.

Each node regularly exchanges information with each other, updating every routing table detecting the insertion and removal of mesh nodes. Usually, in an ad hoc networks, as a node receives the routing update information, he sends a broadcast message, retransmitting information to every neighbor (a.k.a. flooding). This flood of routing information is performed many times, meaning a node can receive the same packet time and time again unnecessarily, generating an undesired overhead in the network.

A powerful and efficient policy to control these many broadcast messages that are flooded across the network is a mechanism called multipoint relay (MPR). This technique is the main difference between OLSR and other proactive protocols. MPR is an optimization that regards the election of some nodes that are able to broadcast update messages. Therefore, a controlled flooding is achieved avoiding the replication of update messages in the network.

OLSR presents some clear advantages when used in the M.E.N.A.C.E-TRACK MANET environment. These concern its nature as intended to be used in high density networks, greatly due to the MPR approach. The bigger and more cluttered the network is, better the route optimization provided. OLSR was developed to work in a completely distributed fashion, avoiding any dependency with a central entity. It also does not need to transmit reliable control messages using TCP. All the communication is done using UDP port 698 for the transmission of periodical unreliable messages. The loss of some messages is of no consequence to its operation. OLSR messages do not have to be delivered in sequence, since each message contains a sequence number. Therefore, the destination can control the sequence of the messages delivered, and in case of loss, request the retransmission of the missing part. It also supports IPv4 and IPv6 [18].

OLSR does not change the TCP/IP protocol suite in any way. It only interacts with layer 2 management tables. OLSR networks support IP addressing to identify each node. The use of multiple interfaces is also supported, although a preferred IP must be chosen for routing. Each node in an ad hoc OLSR network has a direct and bidirectional (i.e. symmetric) relationship. The uncertainties about the propagation of the radio signal may cause some communications to be restricted to unidirectional links. Nevertheless, each communication must be verified in both directions so that one link can be considered valid. To accomplish that, each node periodically sends a HELLO message that contains the information of neighbors (link sensing, neighborhood detection and MPR selection signalling), and are transmitted in broadcast mode [18].

A HELLO message contains a list of neighbor addresses which have a valid bidirectional connection and a list of the neighbor addresses that are listened by this node, but whose link is still not valid as bidirectional. If a node has its own address in a HELLO message, the link is considered bidirectional for the sender node. The HELLO messages transmitted by a node are received by each of its neighbors with a distance of one hop, but they are not retransmitted by them.

The OLSR protocol is considered an optimization of the link state protocol adapted to MANETs because it reduces the size of control messages. Instead of stating all the links, he states only a subset of the neighbors' relationships. As a consequence, OLSR minimizes the flooding, controlling the traffic using only selected MPRs to broadcast HELLO messages from the second hop onwards. Only the MPRs related to a given node retransmit its broadcast messages. Therefore, MPRs minimize the overhead of HELLO messages which otherwise would be coming from all the active nodes in a mesh network, avoiding the broadcast of redundant information. To select MPRs, each node in the mesh network selects a set of symmetric nodes a hop away. The premise is a node must reach every second order node using the fewer MPRs possible, allowing a source node to reach any other node at a distance of two hops.

Neighboring nodes to a given S, that are not MPRs, receive the broadcast message but do not relay it. Each node chooses a neighbor to be its MPR considering this is a symmetric (i.e. bidirectional) hop. This selection is performed so the coverage of the radio link of all symmetric nodes are at 2 hops distance. S is also known as a selector node given it is choosing its MPRs. Each node elected to be MPR to S (MPR(S)) keep information about the set of neighboring nodes belonging to MPR(S). The node's set of MPR nodes chosen by S is known as Multipoint Relay Selector Set (MPRSS). A node discovers the MPRSS from periodic information received from its neighbors. A broadcast message destined to be sent in the network, from any MPR(S) is assumed to be relayed back to S, in case S has still not received the message. The set can change over time (i.e. when a selector node pick another MPR). This is indicated by the HELLO message sent from the selector node. The premise here is that the node will only retransmit an OLSR packet if it is chosen as MPR by the last node that retransmitted the message and if the packet TTL is major than zero.

Each OLSR node keeps the information about the network topology. This information is acquired from TC messages and used to calculate the routing tables. A node keeps a routing table that allows finding a path to other network nodes. This routing table is created from the local link information base. The local link information base stores the information about the paths to neighboring nodes. If any of these paths is modified, the routing table is recalculated to update the routing information about any node destiny in the network. The routing entries are defined in [18] as presented in Listing 1.

Listing 1: Routing table format ([18] pg 46).

```
1.  R_dest_addr    R_next_addr    R_dist
    R_iface_addr
2.  R_dest_addr    R_next_addr    R_dist
    R_iface_addr
3.      ,,             ,,            ,,          ,,
...
```

Each entry in the table consists in R_dest_addr, R_next_addr, R_dist, e R_iface_addr, where R_dest_addr distance is estimated in R_dist hops from the local node, its symmetric neighboring node is R_next_addr, which is the next hop in the route to R_dest_addr and its local interface has the address R_iface_addr. These entries are recorded in the routing table for each network destiny for which a route is known. For every destiny, when a route is broken or just partially known, this entry is not registered in the table.

MANETs are usually isolated, however, there are situations where there is the need to access other networks. The HNA is the solution presented to this situation. It works as a host in the mesh network identifying itself as a gateway to other network and it can present its services by means of Host and Network Association (HNA) messages. When a node receives HNA messages from other node, it adds the transmitter as a gateway to other network. The address to this other network is obtained reading the fields `Network Address` and `Netmask`. In general, if the transmitter is an Internet gateway the field `Network Address` and `Netmask` will both have the value 0.0.0.0.

4. Case Study: M.E.N.A.C.E-TRACK – MANET side

The research project M.E.N.A.C.E-TRACK proposes 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 in case a node or group of nodes leaves or returns to the main base. Some important features for M.E.N.A.C.E-TRACK when considering field personal are high bandwidth (if available), end to end communication with the MRs, all mobile network nodes are also traffic routers forwarding packets until they reach the destiny using some kind of topology control for the deliberate adjustment of certain system parameters (e.g.antenna direction, transmission power, routing protocols etc) to form a particular and more adequate network topology, end to end IP support, data transfer, audio and video streaming, geographic positioning with or without the use of GPS (depending on the accuracy needed) and support to mobility and scalability [19], [20].

The steps to build the M.E.N.A.C.E-TRACK infrastructure consist in defining the devices that can be used as nodes, defining the operating system for such devices and defining the best routing protocol(s) to provide routing adjustments considering mobile nodes with varying ranges. This paper will focus mainly in the first and second steps. We chose a group of mobile devices for testing (notebooks and smartphones) and the most suitable operating system for each one. OLSR will be used as the routing protocol 1 in every device.

4.1 Testbed Preparation

We used as our MANET testbed two Macbook Air 11" and two smartphones Galaxy S3 GT-I9300. The notebooks used OS X 10.7.5 and the smartphones, a modified version of Android OS (version 4.2.2). All of the were configured to behave like mesh clients. One problem that immediately arose when configuring the devices as mesh clients was that not all of then could work in ad hoc mode. The notebooks did not present any problem to be configured in ad hoc mode (just a matter of using the OS X GUI and configure a new

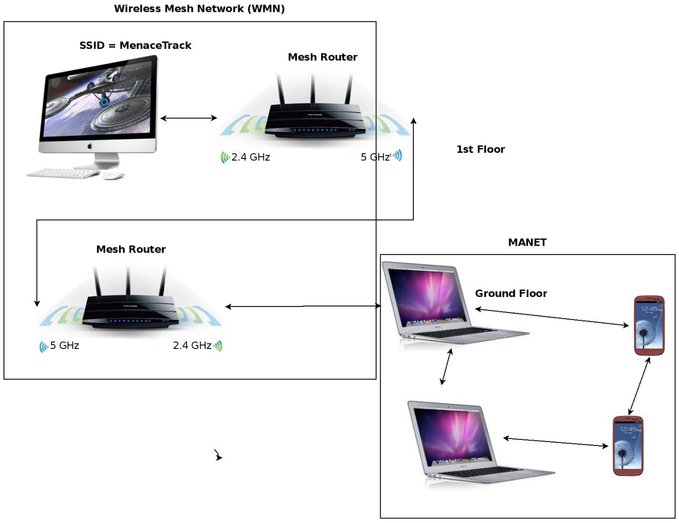


Fig. 1: Proposed testbed for the case study (MANET side).

network). For the smartphones (Galaxy S3 GT-I9300) it was not so easy.

First of all, Android OS does not support ad hoc mode for its mainstream version (i.e. factory release). The need for Ad hoc or Independent Basic Service Set (IBSS) support is not something new and the users and developers community is asking for a solution for more than five years [21]. Users usually understand by "ad hoc" mode the ability to share 3G via W-Fi (i.e. to create a hotspot) [22]. As presented in section 3, the definition of ad hoc is much more broad and complex than this understanding of ad hoc. This capability won't be shipped on new smartphones anytime soon, but it is a really interesting trend and for some an Android project to which they might contribute [6].

In order to create true ad hoc mode in the smartphones, one will have to completely reinstall the operating system. This procedure includes rooting the device and choosing a version of Android that supports ad hoc mode. As of now, this modified versions of Android are ver scarce (e.g. Galaxy Nexus, Nexus 7, Nexus S, Samsung Epic 4G). Although, Thinktube [23] did not have an image for the Galaxy S3 GT-I9300, the documentation available in the site and the personal answers provided by Mr. Bruno Randolf helped a lot understanding the complexities involved in enabling ad hoc mode in Android smartphones.

Basically, Mr. Randolf and the Thinktube fellows [23] created patches to bring the missing Ad-Hoc (IBSS) mode to Android for the aforementioned devices, in a way that is fully integrated into the Android system API and user-interface. Therefore, it is possible to create and connect to Ad-Hoc networks from the standard user interface (Settings - WiFi) and Applications have an API to configure their "own" ad hoc network. Although this is a commendable effort, it opens some question regarding some official ad hoc support for Android OS provided by Google and even, a timeframe for this support to be available. Otherwise, we are in a situation were one would have to adapt the OS for particular devices in order to gain the possibility to use this feature.

The problem is not the hardware support, but simply a matter of software limitations introduced in the Android OS in order to block the use of the ad hoc mode. In more detail, in order to enable the ad hoc support, one would have to modify the WLAN driver "`bcmdhd`", extend the Android framework, the public Android API, and add the missing additional hooks to the "Settings" application. It is really a shame the other great mainstream version of Android OS (codename cyanogenmod [24]) also does not support ad hoc mode.

Fortunately while researching alternatives for ad hoc mode on the project smartphones we stumbled on the Smart Phone Ad-Hoc Networks (SPAN) project [25]. This project reconfigures the onboard Wi-Fi chip of a smartphone to act as a Wi-Fi router with other nearby similarly configured smartphones, creating an ad-hoc mesh network. These smartphones can then communicate with one another without an operational carrier network. A modified version of Android OS was created to root specific models of Android smartphones in order to expose and harness the ad-hoc routing features of the onboard Wi-Fi chip [6]. Bottom line, the modified version of Android puts the network interface in ad hoc mode with the SSID configured as AndroidAdhoc. A minor discomfort of this modified Android OS is its inability to connect to infrastructure networks (usual APs). Therefore, we cannot download directly any applications.

So, the approach we used to put the wireless interface in ad hoc mode using this modified version of Android was to install the MANET Manager app [26] in OTHER smartphone (with an official Android OS version), install Apk Extractor [27] and extract the package file (apk) for MANET Manager, use a USB cable and copy the MANET Manager app to a computer and then, to the target smartphones. Later, we installed the MANET Manager in the target smartphone and start MANET Manager and turn the app on. At this point the MANET Manager app put the wireless interface in ad hoc mode. Now, all that's left is to create a method to use the smartphone as a mesh client node. In personal conversations with Mr. Jeff "stoker" Robble and Mr. Bruno Randolf it became clear that the best way to put this devices to use as mesh clients would be to explore the structure of Android OS in-depth. The problem is, the authors are not that much well versed in Android OS modification and programming. Therefore, we came with an alternative approach that would allow us to use the smartphones as intended without the need to modify directly the Android OS.

We chose to install a Linux on top of the Android OS. To do so we downloaded an Ubuntu ARM compliant version of Linux (there are other Linux flavors) [28]. After downloading the image file, we unpack and put it on the smartphones (via USB cable). It is recommended to rename the file to `ubuntu.img` and put it in `/sdcard/ubuntu`. After you need a shell script to load the Linux OS [29]. This file have to be put on `/sdcard/ubuntu`. Using a terminal emulator on the smartphone in super user mode will grant `root` access. After, one must issue `sh /sdcard/ubuntu/ubuntu.sh`. This will trigger the Ubuntu OS, mounting the ubuntu.img on /data/local/mnt and chrooting it to /.

We installed Linux on Android so we could have all the flexibility of Linux in Android. Therefore, we need to install some useful tools like `iwconfig`, `ping` and `olsrd`. As we saw earlier, we cannot do that from the target smartphones because they do not have support to infrastructured networks. The best way to accomplish this would be to create an ad hoc gateway or simply install all the packages needed on another standard (rooted) smartphone and copying the directory `/var/cache/apt`. As we already had another smartphone with cyanogenmod installed (from earlier case studies) we installed all the needed packages in it and them, generated a `tar.gz` file and copied it to the target smartphones. There we just issued the appropriate `apt-get` commands and installed all the needed tools (e.g. `apt-get install olsrd`). Although this method is not the best due to the overhead caused by a second operating system running on top of Android OS, this approach was chosen to simplify the case study for mesh networks.

4.2 olrsrd configuration

There is one OLSR implementation that is becoming the standard and most widely used known as OLSRd (old Unik-OLSR). One key advantage observed in OLSRd implementation for Linux is its support for IPv6 (this feature won't be used for this case study). OLSRd is an implementation based on the INRA C code, but has been almost completely rewritten (it's almost GPL). OLSRd also is under continuous development [30].

OLSRd fully complies to the RFC 3626, including support for plugins and an optional GUI. The implementation also has a informative up-to-date web-page with links to mailing lists and papers [31]. The experimental network topology used was as follows: Smartphone A ↔ Smartphone B ↔ Notebook A ↔ Notebook B. One very particular feature of this testbed is the 20 cm reach of the smartphones. This is pretty bad when comparing to a 10 to 15 m of the notebooks but it allows for testing the mesh reconfiguration. For instance, if we move Smartphone B to the reach of Notebook B, the routing to reach Smartphone B would have to be done through Smartphone B.

This first configuration of the file `/sw/etc/olrsrd.conf` (we used `fink` to install it on OS X) was kept pretty much default. The only changes we made were:

- to set `DebugLevel = 2`, kept `IpVersion = 4`;
- enabled `LoadPlugin` "`olsrd_httpinfo.so.0.1`" and put it on port 8080 and available to the localhost and the network range we were using (192.168.0.0/24). Therefore, we can access any of the network nodes from each other;
- for this test we didn't configure any device as an Internet gateway (therefore, no need to set Hna);
- we set each interface appropriately (e.g. Interface "`en1`").

Considering this parameters for each and every mesh node, and the suggested topology, we just have to configure every node (except the smartphones) to connect to the ad hoc network with SSID set to AndroidAdhoc (set by MANET Manager). All the IPs at this point are configured manually but we are working on a DHCP setup. Therefore, with all the IPs in the range 192.168.0.0/24 (Smartphone A = 192.168.0.100, Smartphone B = 192.168.0.101, Notebook A = 192.168.0.10 and Notebook B = 192.168.0.20) we triggered `olsrd` in all of the nodes.

With the debug level set to 2, we can follow everything that is happening on every node. For each node, the routing table is created and we can ascertain that by issuing `netstat -r`. We can also view the nodes that are accessible by one hop count and two hop counts. At first, Notebook A has nodes 192.168.100, 192.168.101 and 192.168.0.20 as nodes reached by one hop count. Due to the terrible reach of the smartphones, when we get Smartphone B and take it next to Notebook B, the cost to reach Smartphone B from Notebook A turns immediately to INFINITE. Therefore, it is removed from the nodes reached by one hop count. It takes sometime for it to be "reintegrated" to the mesh. After about 3 minutes, the routing tables are slowly updated and we can reach Smartphone B again. This testbed was not completely stable regarding some failures of the smartphones and the slow reintegration when we move them out of reach and into the reach of another node.

5. Final Remarks

This paper presented the bare bones of project M.E.N.A.C.E-TRACK and its intended application context. We also have shown the complexities regarding the MANET side of this project. Although the initial idea of using smartphones as a covert channel for a MAN size mesh network, the lack for ad hoc mode support is worrisome. The demand for this feature exists for at least five years and the main market players haven't issued an answer to date. What we have is an stoic effort from the open source community with individuals going to the extent of modifying the kernel of some smartphone OSs so the community can tinker with the ad hoc mode. Although this is not the best case scenario, as far as research goes, we can still test mesh routing protocols on the proposed testbed. In this paper, we achieved a configuration for the smartphones of the project that will allow us flexibility to configure any Linux supported mesh routing protocol and see its effects on a controlled test environment. The ridiculously short reach of the smartphones is quite interesting considering tests involving the change in topology of an established mesh network. Therefore, the answer for the proposed research question is NO, considering the current state of smartphones, they could not play a key part in M.E.N.A.C.E-TRACK.For our future works, we intend to explore OLSR in depth and test various configuration parameters so we can determine its effectiveness in the aforementioned testbed. We also need to integrate the MANET side of M.E.N.A.C.E-TRACK with the WMN side (i.e. with the OpenWrt APs).

Acknowledgment

The authors would like to thank the Research Assistance Foundation of the State of Goiás (FAPEG – Fundação de Amparo a Pesquisa do Estado de Goiás) for the full sponsorship of this research – edict number 006/2012.

References

[1] J. Wang, B. Xie, and D. P. Agrawal, *Guide to Wireless Mesh Networks*. London UK: Springer, 2009, ch. Chapter 1 - Journey from Mobile Ad Hoc Networks to Wireles Mesh Networks, p. 28.

[2] R. Woods, "Connectivity as commodity," *Latin America*, pp. 36–37, 2010.

[3] F. Tamusiunas. Aferição da qualidade pelo usuário e defesa de seus direitos (evaluation of quality by the user and defense of their own rights). [Online]. Available: http://az545403.vo.msecnd.net/uploads/2013/08/Fabricio-Tamusiunas-.pdf

[4] (2013, January) Nic.br divulga diagnóstico da internet brasileira durante cp 2013 (nic.br discloses diagnosis of brazilian internet for cp 2013). [Online]. Available: http://www.ebc.com.br/tecnologia/2013/01/nicbr-divulga-diagnostico-da-internet-brasileira-durante-cp-2013

[5] Teleco, "3g: 3rd cellular generation in brazil (3g coverage)," April 2014. [Online]. Available: http://www.teleco.com.br/en/en_3g_cobertura.asp

[6] S. M. Patterson, "Previous article next article android phones are connecting without carrier networks," February 2013. [Online]. Available: http://www.networkworld.com/community/blog/android-phones-are-connecting-without-carrier-networks

[7] S. Methley, *Essentials of Wireless Mesh Networking*, ser. Cambridge Wireless Essentials Series. Cambridge University Press, 2009.

[8] X. Zhao, W. N. Hung, Y. Yang, and X. Song, "Optimizing communication in mobile ad hoc network clustering," *Computers in Industry*, vol. 64, no. 7, pp. 849–853, September 2013.

[9] M. Naserian, *Game Theoretic Approach in Routing Protocols for Wireless Mobile Ad Hoc Networks*, ser. Canadian theses, U. of Windsor, Ed. Canada: University of Windsor, 2008, vol. 1.

[10] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris, "Link-level measurements from an 802.11 b mesh network," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 4, pp. 121–132, 2004.

[11] U. of Thessaly, "Vmesh – wireless network testbed at university of thessaly, volos, greece," Volos, Greece. [Online]. Available: http://vmesh.inf.uth.gr/

[12] N. Tsarmpopoulos, I. Kalavros, and S. Lalis, "A low-cost and simple-to-deploy peer-to-peer wireless network based on open source linux routers," in *Testbeds and Research Infrastructures for the Development of Networks and Communities, 2005. Tridentcom 2005. First International Conference on*. IEEE, 2005, pp. 92–97.

[13] H. Lundgren, K. Ramachandran, E. Belding-Royer, K. Almeroth, M. Benny, A. Hewatt, A. Touma, and A. Jardosh, "Experiences from the design, deployment, and usage of the ucsb meshnet testbed," *Wireless Communications, IEEE*, vol. 13, no. 2, pp. 18–29, April 2006.

[14] J. Bicket, D. Aguayo, S. Biswas, and R. Morris, "Architecture and evaluation of an unplanned 802.11b mesh network," in *Proceeding of the MobiCom'05*, Cologne, Germany, August 2005.

[15] G. Aggelou, *Wireless Mesh Networking*. McGraw-Hill Communications, 2009.

[16] P. Garnepudi, T. Damarla, J. Gaddipati, and D.Veeraiah, "Proactive, reactive and hybrid multicast routing protocols for wireless mesh networks," in *Proceedings of the 2013 IEEE International Conference on Computational Intelligence and Computing Research*. IEEE, 2013.

[17] M. Bahr, J. Wang, and X. Jia, *Wireless Mesh Networking - Architectures, Protocols and Standards*, ser. Wireless Networks and Mobile Communications. Boca Raton, FL: Auerbach Publications - Taylor & Francis group, 2007, ch. Chapter 4 - ROUTING IN WIRELESS MESH NETWORKS, p. 125.

[18] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)," RFC 3626 (Experimental), Internet Engineering Task Force, Oct. 2003. [Online]. Available: http://www.ietf.org/rfc/rfc3626.txt

[19] M. Natkaniec, "Ad hoc mobile wireless networks: Principles, protocols, and applications (sarkar, s. k. et al.; 2008) [book review]," *Communications Magazine, IEEE*, vol. 47, no. 5, pp. 12–14, May 2009.

[20] H. X. T. Tan and W. K. S. Seah, "Dynamic topology control to reduce interference in manets," Institute for Infocomm, Tech. Rep. NET-2004-AHN-GENERAL-0027, Nov 2004.

[21] Google Groups - android-platform, "Android adhoc ibss wifi support." [Online]. Available: https://groups.google.com/forum/#!msg/android-platform/tLLspmSySbY/fm0zqCFBvQkJ

[22] Google Groups - android, "Issue 82: Support wi-fi ad hoc networking." [Online]. Available: https://code.google.com/p/android/issues/detail?id=82

[23] B. Randolf, "Ad-hoc (ibss) mode support for android 4.2.2/4.3/4.4." [Online]. Available: http://www.thinktube.com/android-tech/46-android-wifi-ibss

[24] cyanogenmod, "Introducing the cyanogenmod installer." [Online]. Available: http://www.cyanogenmod.org/

[25] J. "stoker" Robble, J. "m0nk" Thomas, and D. Doria, "The span project." [Online]. Available: https://github.com/ProjectSPAN

[26] J. "stoker" Robble, "Manet manager." [Online]. Available: https://play.google.com/store/apps/details?id=org.span&hl=en

[27] zxcasd12@gmail.com, "Apk extractor." [Online]. Available: https://play.google.com/store/apps/details?id=net.sylark.apkextractor&hl =en

[28] zacthespack, "Linux-on-android – working to getting a range of linux distros running on android." [Online]. Available: http://sourceforge.net/projects/linuxonandroid/files/

[29] tiborr, "Samsung galaxy s3 with ubuntu." [Online]. Available: http://forum.xda-developers.com/galaxy-s3/general/ubuntu-backtrack-i9300-samsung-galaxy-s3-t1700228

[30] The Linux Tutorial, "Linux optimized link state routing protocol (olsr) ipv6 howto." [Online]. Available: http://www.linux-tutorial.info/modules.php?name=Howto&pagename=OLSR-IPv6-HOWTO/olsrlinux.html

[31] A. Tønnesen, T. Lopatic, H. Gredler, B. Petrovitsch, A. Kaplan, S.-O. Tücke, *et al.*, "olsrd – an ad hoc wireless mesh routing daemon." [Online]. Available: http://www.olsr.org/

Access Point Reconfiguration Using OpenWrt

Dalton M. Tavares[1], **Maicon J. Lima**[1], **Rafael V. Aroca**[2],
Glauco A. P. Caurin[3], **Antônio Carlos de Oliveira Jr**[1],
Tércio A. Santos Filho[1], **Stella J. Bachega**[4],
Marcos A. Batista[1], **and Sérgio F. da Silva**[1]
[1]Computer Science Department, Federal University of Goiás, Catalão, Goiás, Brazil
[2]Mechanical Engineering Department / Federal University of São Carlos, São Carlos, São Paulo, Brazil
[3]Mechanical Engineering Department / University of São Paulo, São Carlos, São Paulo, Brazil
[4]Production Engineering Department / Federal University of Goiás, Catalão, Goiás, Brazil

Abstract—*The research project Mobile mEsh Network to Aid in CountEring drug TRAffiCKing (M.E.N.A.C.E-TRACK) proposes 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. To configure a mesh node using, for instance, an access point, usually a modified firmware is needed. In this paper we present the first steps to build the M.E.N.A.C.E-TRACK infrastructure concerning the configuration of the access point, the chosen firmware and some configuration scenarios on an infra-structured network in order to demonstrate its flexibility.*

Keywords: Mesh Network, Ad hoc, OpenWrt

1. Introduction

The research project Mobile mEsh Network to Aid in CountEring drug TRAffiCKing (M.E.N.A.C.E-TRACK) proposes the creation of a dynamic mesh network, intended to interconnect field personnel (e.g. in vehicles or on foot) to a base of operations (e.g. a police station) whenever possible. This type of network accepts the dynamic disconnection and reconnection of a node or group of nodes leaving or returning to the main base network range. The use of day to day low cost and off the shelf devices, like access points, notebooks or smartphones and open source software is one of the main attractors to our approach.

To configure a mesh node using, for instance, a wireless router, usually a modified firmware is needed. This firmware allows the creation of a dynamic route between the base station (the one that have the Internet connection and access to the main systems) and the client nodes in a mesh environment. The dynamic route can be stablished directly between the base station and mesh nodes, or using each mesh node as a "bridge" (in-between nodes) to amplify the range of the original access point. Therefore, each mesh node receives a data connection from a given node and conveys data to the next node, extending the range of communication for each passing node [1].

The steps to build the M.E.N.A.C.E-TRACK infrastructure involve the definition of the devices that can be used as nodes, the choice of the operating system for such devices and the selection of the best routing protocol(s) to provide routing adjustments considering mobile nodes with varying ranges. The chosen devices are off the shelf with the ability to use wireless communication, i.e. IEEE 802.11 b/g/n, and the TCP/IP protocol suite. Specifically, we will cover the adaptation of access points (APs) so they can be used as mesh nodes in the future. This procedure addresses the selection of a firmware, preferably open source, which is flexible enough to be modified according to the choice of the wireless environment to be implemented (e.g. infra-structured or ad hoc network).

2. Infra-structured versus Ad hoc networks

An IEEE 802.11 W-LAN can be implemented either with infrastructure or without infrastructure (i.e. ad hoc). In an infrastructure based network, there is a centralized controller for each cell. This cell represents the fundamental building block of the 802.11 architecture, known as the basic service set (BSS). A BSS typically contains one or more wireless stations and a central base station, known as access point (AP). The wireless stations, which may be either fixed or mobile, and the central base station communicate among themselves using the IEEE 802.11 wireless MAC protocol. Multiple APs may be connected together (e.g. using a wired Ethernet or another wireless channel) to form a distribution system (DS). The DS appears to upper-level protocols (e.g. IP) as a single 802.11 network, in much the same way that a bridged wired 802.3 Ethernet network appears as a single network to the upper-layer protocols. We can use the same analogy considering that an AP, which is normally connected to a wireline backbone (either wired or wireless) is also considered a DS, thus providing Internet access to mobile devices. All traffic goes through the AP, even when this is sent to a destination that belongs to the same cell [2], [3].

In an ad hoc network there is no central control with connections to the Internet. Here, the network is a temporary arrangement as and when required. The benefit is that no infrastructure is needed and the users themselves may extend the area of coverage. Hence, mobile devices that have found themselves in proximity to each other, that have a need to communicate and find no pre-existing network infrastructure in the location (e.g. a pre-existing 802.11 BSS with an AP) may communicate using each other as the current medium [2], [4].

The focus of this paper will reside in extending the range of an infra-structured network using multiple APs. The main objective to be accomplish with this course of action is to gain the necessary expertise concerning the firmware operation and configuration, which will be invaluable in future experiments involving mesh networks.

2.1 OpenWrt Features

OpenWrt is described as a GNU/Linux distribution for embedded devices, which provides a fully writable filesystem with package management. In that sense, by definition, it is not strictly a firmware but a complete and modifiable operating system. This frees the developer from the static application profile provided by a vendor and allows the customization of the device through the use of packages that suit any particular need [5].

Compared to other distributions OpenWrt may also not be regarded as a true end-user or user friendly firmware. Nevertheless, it can be used as such sometimes, depending on the feature set provided in addition to the main package [5]. Therefore, OpenWrt was chosen as our base architecture to implement M.E.N.A.C.E-TRACK. Section 3 presents a set of experiments devised to study OpenWrt flexibility regarding a simple research question: how can I extend the range of an infra-structured IEEE 802.11 wireless network?

3. Experimental Testbed

We chose as our experimental testbed a TP-LINK TL-WDR 4300 router. According to [6] this router provides a simultaneous dual band (concurrent 2.4 GHz and 5 GHz) and is advertised as 750 Mbps in dual-stream mode on the 2.4 GHz band, and triple-stream on the 5 GHz band. It's hardware can be summarized as an Atheros AR9344 CPU operating at 560MHz, 8MB internal flash, 128 MB RAM, 4 Gigabit Ethernet ports, 1 Gigabit Ethernet WAN port, 2 USB 2.0 ports, Serial and JTag interfaces and support up to 12 VLANs.

To simplify the presentation of the experimental results, we organized this section to discuss the device preparation (section 3.1) and the case studies involving the two APs so they can expand the indoor range of the Wi-Fi connection shared by a desktop computer (section 3.2). The experimental testbed is presented in Figure 1.

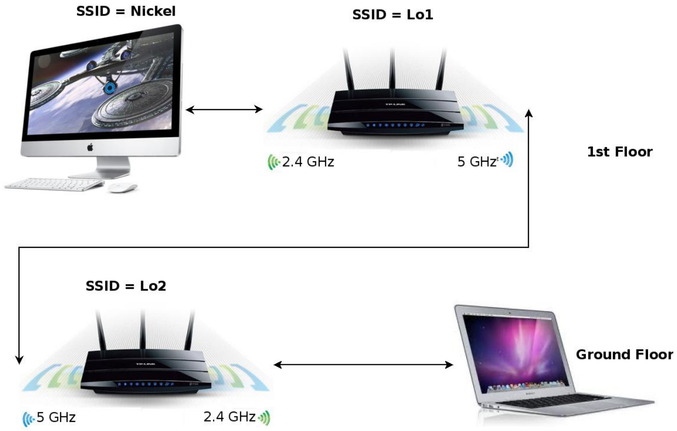


Fig. 1: Proposed testbed for the case study.

3.1 Preparing the AP Devices

The first steps taken to install OpenWrt on the TP-LINK TL-WDR 4300 router are described at [7]. They involve consulting the Buyers' Guide [8], and verifying how much a given router is compatible with OpenWrt using the table of hardware [9].

The first installation of the OpenWrt firmware is done using the original web interface of the TP-LINK firmware. It is performed just like a flash update of some new version of the original TP-LINK firmware. According to [6], it is recommended to turn off the Wi-Fi interfaces manually (there is a switch behind the device to disable them). Therefore, at the first boot, we will not have any wireless interface enabled by default. The first login test is done using an Ethernet cable plugged into any Ethernet port (not the WAN port). The computer interface must be set to any address in the range 192.168.1.0/24, exception made to 192.168.1.1, which is the default address of the OpenWrt interface [10].

After the first connection it is recommended to setup the root password (the only account). After the password is set, the telnet daemon is disabled and all future accesses are done using the ssh service [10]. Considering our testbed is essentially experimental and is intended to test multiple mesh routing protocols, implying in a frequent change in its configuration files, we still did not consider the procedures described at [11] for network and system hardening.

3.1.1 USB Storage

Considering the size of the internal flash memory (8 MB), it is recommended to increase this size as soon as possible. After the base installation, we have up to 4 MB available, which runs out relatively fast depending on the number of packages the administrator chooses to install. Fortunately, OpenWrt allows the extension of this available size by means of an overlay file system. This means we can extend the root file system (stored at the internal flash memory) to an external storage (stored in a USB stick) [12].

This configuration is called pivot overlay on version 12.09 of OpenWrt. To ensure pivot overlay will work, it is recommended some packages are installed for the correct USB support and mounting of the external device (i.e. `kmod-usb-core`, `kmod-usb-ohci`, `kmod-usb-storage`, `kmod-usb-uhci`, `kmod-usb2`, `libusb-1.0`, `usbutils`, `kmod-fs-ext4`, `e2fsprogs`) [13].

Support to pivot overlay is granted by the package `block-mount`. This package allows the mounting of all block devices by just calling the commands `block mount` and `block umount` [14]. To create the pivot overlay on the external USB device, one can either use an empty new rootfs or copy the contents of the current overlay (JFFS2) to the new rootfs. Assuming the filesystem for the new external rootfs is mounted, for example on `/mnt/sda1`, one could issue `tar -C /overlay -cvf - . | tar -C /mnt/sda1 -xf -` [12].

Considering the load of the overlay file system at boot time, it is necessary to create the file `/etc/config/fstab`. This can be done by simply issuing `block detect > /etc/config/fstab` [15]. After enabling the pivot overlay at boot time, we must ensure the external file system is being mounted correctly. Special attention to the target and device options. If after a system reboot a command `mount /mnt` issues no error, than everything is correct. The last remaining step is to change `option target /mnt` to `option target /overlay` and in the next boot, the file system size will be the size of the USB stick plugged in the USB port (minus the space already in use).

3.2 Case Studies

Our experimental case study considers the range extension of a connection shared by a desktop computer (SSID = Nickel) to a notebook. When in an indoor environment in the first floor of a building, we found it would not go further than approximately 10 m. The next step was to establish a simple connection to the two APs as client/server and the connection of a notebook on their respective SSIDs (i.e. Nickel ↔ Lo1 ↔ Lo2 – the notebook can access either of the SSIDs). As each antennae has its own frequency (2.4 GHz and 5 GHz) we had to match them as illustrated on Figure 1. In OpenWrt, this configuration is done at `/etc/config/wireless` (Listing 1).

Listing 1: Excerpt from `/etc/config/wireless`.

```
...
config wifi-device  radio0
        option hwmode    11ng ←
...
#radio0 is connected to the Desktop
config wifi-iface
        option device   radio0
        option mode     sta
        option ssid     'Nickel'
        option encryption psk2
        option key       password
        option network  wwan
...
config wifi-device  radio1
        option hwmode   11na ←
...
config wifi-iface
        option device   radio1
        option network  lan
        option mode     ap
        option ssid     Lo1
        option encryption psk2
        option key       password
```

Listing 1 shows that `radio0` operates in the 2.4 GHz band (`hwmode 11ng`) while `radio1` operates in the 5 GHz band (`hwmode 11na`). `radio0` is in mode `sta`, meaning it is configured to be a client of `Nickel`. It also shows that `Nickel` uses encryption (WPA2) and is related to a network called `wwan`. The configuration for `radio1` is analogous, but it is operating in AP mode (`option mode ap`), with `SSID Lo1` and is related to network `lan`. The definitions for networks `wwan` and `lan` are showed on Listing 2.

Listing 2: Excerpt from `/etc/config/network`.

```
...
config interface 'lan'
        option ifname 'eth0.1'
        option type 'bridge'
        option proto 'static'
        option ipaddr '192.168.0.1'
        option netmask '255.255.255.0'
...
config interface 'wwan'
        option ifname 'wlan0'
        option proto 'dhcp'

config interface 'stabridge'
        option 'proto'      'relay'
        option 'network'    'lan wwan'

config 'zone'
        option 'name'        'lan'
        option 'network'     'lan wwan'   #
            Important
        option 'input'       'ACCEPT'
        option 'forward'     'ACCEPT'    #
            Important
        option 'output'      'ACCEPT'
```

Listing 2 shows the settings for network `lan` (Ethernet ports) and `wwan` (wireless – `radio0`). Network `lan` is configured in bridge mode and its static IP address is 192.168.0.1, which ensures access to the AP via Ethernet cable (or wireless via `radio1`). Any computer connected via Ethernet cable receives an IP address via DHCP, in the range 192.168.0.100/192.168.0.150 (Listing 3). The `wwan` DHCP config ignores the `lan` pool and gets addresses outside this pool.

A bridge is defined between `lan` and `wwan`. Therefore, any computer connected to the wireless network (on `radio1`) or via Ethernet cable (via `lan`) will receive an

IP address automatically. Instead of a true “bridge”, in this configuration, the traffic forwarded is affected by firewall rules, such that the `wwan` network and the `lan` network should be configured according to the same “zone” created by a firewall with the “forward” policy adjusted to “accept”, so that all the traffic flows between both interfaces [19].

The insertion of firewall rules inside `/etc/config/network` is not recommended, but it was implemented as such for the sake of simplicity in this first test. According to the official documentation [20], this configuration should not work. It is explicitly mentioned that “...STA and AP at the same time is not yet supported...”. Fortunately, this was true for trunk version up to r22989. We are using trunk version r40555.

Listing 3: Excerpt from `/etc/config/dhcp`.

```
...
config dhcp ’lan’
        option interface ’lan’
        option start ’100’
        option limit ’150’
        option leasetime ’12h’
...
config dhcp ’wwan’
        option interface ’wwan’
        option ignore ’1’
...
```

This configuration demonstrates how we can connect each AP in series to Nickel. The configuration of the second access point (Lo2) is analogous. Attention to the frequency of the interfaces. In Lo2, we must connect `radio1` (`hwmode 11na - 5 GHz` in `mode sta`) to `radio 1` in Lo1 (`hwmode 11na - 5 GHz` in `mode ap`) and `radio0` (`hwmode 11ng` - 2.4 GHz in `mode ap`) must be available to the clients (e.g. notebooks or other APs).

The objective is to set each AP in different floors in a building (i.e. first floor and ground floor) in order to extend the range of Nickel. To enable Internet access from either AP, it is necessary to set a packet forwarding service on both APs so that each could route traffic appropriately. We will also discuss a Wireless Distribution System (WDS) implementation in order to extend the original SSID beyond its original range. The specifics of each approach will be further discussed in sections 3.2.1 and 3.2.2.

3.2.1 IP Forwarding

According to [21], netfilter is used for packet filtering, NAT and mangling. This firewall is configured by means of a proper syntax called Unified Configuration Interface (UCI) [22]. This “language” is intended to centralize the configuration of OpenWrt. The problem with UCI is the fact one needs to learn a new syntax for something that is already well known (i.e. `iptables` syntax). Although UCI is intended to simplify the configuration of OpenWrt, depending on the complexity of the intended scenario, the use of UCI can lead to some confusion. For instance, in the beginning of this section, we saw how to configure bridging according to OpenWrt documentation [19]. The configuration uses the firewall to establish the packet transport from the `lan` to `wwan` as discussed before. Therefore, this can be assessed as an application of the UCI firewall. This section will perform some minor modifications on the discussion regarding the connection of both APs and display a configuration profile easier to follow using the `iptables` scripting allowed on OpenWrt.

In order to obtain Internet access from Lo1 and Lo2, the easiest way is to perform source network address translation (SNAT). SNAT translates an outgoing packet, which may come from Lo1 or Lo2 clients, so that each intermediary OpenWrt system looks like the source. Considering the intent to use Lo1 and Lo2 as extensions of Nickel, a SNAT rule must be created on both. In Lo1 each outgoing packet will have its source IP changed to the Lo1 outgoing IP address (interface in `mode sta`). From Lo2, the outgoing packets coming from its clients will suffer SNAT so that they seem to come directly from Lo2. Then, in Lo1, these packets will suffer another SNAT so the clients connected to Lo2 can access the Internet. Therefore, a client must be able to access the Internet connecting directly through Nickel, through Lo1 (going further in the same floor) or through Lo2 (covering the ground floor). The modified configuration for `/etc/config/wireless` and `/etc/config/network` is presented on Listings 4 and 5.

The main differences from the version presented before are the removal of the firewall rules originally inserted to establish the bridge rules and the creation of two distinct interfaces for each radio (`wlan0` and `wlan1`) (on `/etc/config/network` – see Listing 2). We renamed interface `wwan` to make it simple to correlate the “virtual” interface names (`interface` parameter) to the “real” interface names (`ifname` parameter). Therefore, `wwan` will be renamed to `wlan0` and we will create a new interface called `wlan1` which will be in `mode 'ap'`. We also modified accordingly file `/etc/config/dhcp`. There are no modifications to file `/etc/config/wireless`.

Listing 4: Excerpt from modified `/etc/config/network`.

```
...
config interface ’wlan1’
        option ifname ’wlan1’
        option proto ’dhcp’

config interface ’wlan0’
        option ifname ’wlan0’
        option proto ’dhcp’
...
```

Listing 5: Excerpt from modified `/etc/config/dhcp`.

```
...
config dhcp ’wlan0’
        option interface ’wlan0’
        option ignore ’1’

config dhcp ’wlan1’
        option interface ’wlan1’
        option ignore ’1’
...
```

This scenario seems simple enough, however we have some operational problems. First, considering the UCI syntax, it is not clear how to specify a rule that considers any source IP to be translated to a given destination IP. Second, considering Lo1 and Lo2, we have to detect the destination IP address every time the packet is translated. The outgoing interface is being configured via DHCP so, the address is very likely to change in each boot (or every 12 h as described in Listing 3).

The best way to address this issues was to use the UCI firewall interface to `iptables` commands. Therefore, we achieved a simplified configuration model that is fit for the purposes of our case study. We created a user script that is processed by the UCI firewall, after the firewall rules are loaded. This script is stored in `/etc/firewall.user` (Listing 6).

Listing 6: Excerpt from `/etc/firewall.user`.

```
...
WANIF="wlan0"
WANIP="`/sbin/ifconfig $WANIF | grep
    ’inet addr’ | awk ’{print $2}’ |
    sed -e ’s/.*://’`"
iptables -t nat -I POSTROUTING 1 -o $WANIF
    -j SNAT --to $WANIP
```

The main part of the user script of Listing 6 is the deduction of the IP address of the outgoing interface. We create a filter using `grep` and `awk` in oder to extract from the `ifconfig` command output exactly the IP address of the outgoing interface (defined by the user in the `WANIF` parameter). This is used as the input for the SNAT rule (in the `iptables` rule). Therefore, for any IP address coming from the `WANIF` interface we have the translation to the new outgoing `WANIP` address.

One last quirk of OpenWrt is that the processing of the `firewall.user` is not deterministic in our system. It sometimes worked (i.e. the `firewall.user` file is processed) and it sometimes didn’t. Therefore, as a workaround, we used the init script `/etc/rc.local` (Listing 7). The commands appearing inside this file are executed once the system init is finished.

Listing 7: Excerpt from `/etc/rc.local`.

```
# Workaround to load the SNAT firewall rule
# correctly
sleep 10
/etc/init.d/firewall stop # clean up the system
                          # firewall rules
sleep 10
/bin/sh /etc/firewall.user # insert SNAT firewall
                           # rule
```

3.2.2 Wireless Distribution System

According to [23], WDS is a misunderstood concept. WDS is usually referred to as a “wireless DS” or a “DS” that operates over a WLAN. A WDS (as defined by [24]) is neither. This confusion perhaps results from an extremely poor choice in naming the WDS capability. The “WDS” capability actually has nothing to do with either of those terms.

Still according to [23], WDS is a mechanism for constructing 802.11 frames using a 4-address format. The content of the data frame address fields are dependent upon the values of To DS and From DS bits and is defined in Table 1. If the content of a field is shown as not applicable (N/A), the field is omitted. Note that Addr. 1 always holds the receiver address (RA) of the intended receiver and Address 2 always holds the address of the station that is transmitting the frame (TA). Addr. 3 and 4 refers to the usual destination address (DA) and source address (SA).

Table 1: WDS 4-address format [23].

To DS	From DS	Addr. 1	Addr. 2	Addr. 3	Addr. 4
1	1	RA	TA	DA	SA

OpenWrt implements WDS between a client AP and a master AP using the 4-address format, which enables transparent bridging on the client side. In this scenario, a bridged host (e.g. computer A) sends a packet to a target host (e.g. computer B). The frame is relayed via the client AP (i.e. Lo2) and the sender MAC (i.e. computer A) is preserved. The master AP (i.e. Lo1) receives the frame and redirects it to the target (i.e. computer B) using the original sender source MAC (computer A). The target (computer B) receives the frame and generates a response, using the given source MAC (computer A) as destination. The master AP relays the frame to the client AP with the right destination MAC as target (computer A). The client AP receives the frame and redirects it to the final destination using the computer A MAC as target. Computer A receives the response frame and the connection is established [25].

The aforementioned scenario was inspected using a network sniffer (Wireshark) on both computer A and Computer B. We used `ping` to send an ICMP Echo Request from computer A and inspected the received packet on computer B. It was possible to verify that the MAC address was really from computer A instead of the master AP where the traffic is relayed. We also had a Wireshark running on computer A and we also inspected the return message (ICMP Echo Reply) in computer B. The source MAC address from the packet was the one from computer A and not from the relay station (client AP).

To configure WDS, on the master AP (Lo1), we need to add the line `option wds ‘1’` on the `config wifi-iface` section for `radio 1` (the one configured in `mode 'ap'`), at `/etc/config/wireless` [26] (see Listing 1). That's all there is to it. If a client AP connects to this master AP, the WDS interface is created as `wlan1.sta1`. This can be verified using the `ifconfig` command.

The client AP (Lo2) configuration is a little bit more complicated. First, all the firewall configurations described on section 3.2.1 must be removed. WDS approach is a level 2 bridge, not a level 3 translation via netfilter. For all purposes, all the firewall rules can be disabled on the client AP. Second, at `/etc/config/network`, `interface 'lan'` and `interface 'wlan0'` are bridges. Care must be taken considering the configuration of the static IP addresses in both cases. Both bridges must not be in the same IP address range (see Listing 8). We must also observe that the bridge configuration for `interface 'lan'` is not done inside `/etc/config/network`. This will be done afterwards manually using the `brctl` command (Listing 10) [27]. This is done because UCI, in our test case, for some reason, did not allow the configuration of two bridges simultaneously.

Listing 8: Excerpt from `/etc/config/network`.

```
...
config interface ’lan’
        option ifname ’eth0.1’
        option proto ’static’
# Quirk: invalid IP address to avoid conflict with
# br-wlan0 address
        option ipaddr ’192.168.1.199’ ←
        option netmask ’255.255.255.0’
        option ip6assign ’60’

config interface ’wlan0’
        option ifname ’wlan0’
        option type ’bridge’ ←
        option proto ’static’
        option ipaddr ’192.168.0.200’ ←
        option netmask ’255.255.255.0’
...
```

The `br-wlan0` bridge is a wireless to wireless bridge (`radio 1` ↔ `radio 1` at 5 GHz). Its definition becomes clear in the specification presented at `/etc/config/wireless`, for the `radio 1` (Listing 9). One particular quirk that must be observed is the use of both Wi-Fi interfaces (`radio 0` and `radio 1`). `radio 0` is configured as the local AP for Lo2, but its SSID is identical to the SSID used by Lo1 (including the password). `radio 1` is configured as a client (`mode 'sta'`) of the Lo1. This configuration is necessary to assure a seamless integration between Lo1 and Lo2. That way, for example, a client that is moving across the boundaries of Lo1 and Lo2 cells will not realize the transition from one AP to the other. One particular feature in this case study is that the APs Lo1 and Lo2 incidentally operate in different frequencies (Lo1 uses 5GHz for its clients and Lo2 uses 2.4. GHz). Therefore, the client must also have the ability to operate in both frequencies.

Listing 9: Excerpt from `/etc/config/wireless`.

```
...
config wifi-iface
        option device ’radio0’
        option ssid ’Lo1’  ←
        option encryption ’psk2’
        option key ’password’ ←
        option network ’wlan0’
        option mode ’ap’ ←
...
config wifi-iface
        option device ’radio1’
        option mode ’sta’ ←
        option ssid ’Lo1’ ←
        option encryption ’psk2’
        option key ’password’
        option network ’wlan0’ ←
# This line is important to establish the WDS
        option wds ’1’  ←
...
```

The next step is to add manually a second network interface to the established bridge, in this case, the `br-wlan0`. As this procedure must be done on every boot of `Lo2`, we recommend putting it in `/etc/rc.local`. Another problem is that even though all the clients of `Lo2` correctly access the Internet, `Lo2` itself cannot do it. This is important mainly to keep the ability to install new packages from `Lo2`. The problem is its routing table needs a default gateway, therefore we use `Lo1` address and create a static route [28]. We put all this in `/etc/rc.local`, so the complete procedure is executed for every boot (Listing 10).

Listing 10: Excerpt from `/etc/rc.local`.

```
...
# ### Quirks to load static route and local DNS
# (192.168.0.1)
rm /etc/resolv.conf
ln -sf /etc/config/resolv.conf.auto /etc/resolv.
   conf
# ### Create the static route for the default
# gateway
sleep 5
route add default gw 192.168.0.1 br-wlan0 ←
# ### Add Ethernet interfaces to the wifi bridge
# br-wlan0
sleep 5 # Important to give time for the WDS
        # connection establishment
# ### Add eth0.1 to the br-wlan0 bridge using
   brctl
brctl addif br-wlan0 eth0.1 ←
exit 0
...
```

4. Final Remarks

The proposed research question for this paper regards which methods could be used to extend the range of an infra-structured wireless network. The concealed objective was to find an open firmware, flexible enough so it could be used in the context of project Mobile mEsh Network to Aid in CountEring drug TRAffiCKing (M.E.N.A.C.E-TRACK). In that sense, this paper succeeded considering the scenarios tested on OpenWrt provided a rich environment for interaction and customization of the firmware. Therefore, we tested the configuration of APs used to extend the original reach of a wireless network using a level 3 solution (i.e. a firewall) and a level 2 solution (bridging via WDS).

The main advantage of using packet forwarding on the APs is the solution is hardware independent. We can use packet forwarding in virtually any device (considering it has at least up to the network layer of TCP/IP protocol stack). The main drawback is transparency. The user must choose explicitly the SSID of the network he/she is using. This limits the kind of application one can implement considering the need for a seamless integration of APs in order to provide a unified SSID for a bigger range (e.g. in the transit of a mobile robot which exchanges information with an external computer in the same Wi-Fi network).

Considering transparency of integration among devices, WDS is the best choice between the test cases. It provides the possibility for using multiple AP devices without the need to choose between SSIDs. There is only one SSID and the integration is seamless. The drawbacks, on the other hand, can be more extreme when we compare this solution to packet forwarding. WDS is not a certified IEEE standard and, therefore, every vendor (e.g. Ralink, Atheros, Broadcom etc) can have its own implementation, resulting in incompatibility among devices [29].

One design problem we managed to solve with our configuration concerns the loss of bandwidth in WDS Repeater mode. According to [29], WDS Repeater mode will sacrifice half of the bandwidth available from the primary router for clients wirelessly connected to the repeater. This is a result of the repeater taking turns talking to not just one partner but two, and having to relay the traffic between them. As the AP used for the WDS use case has two independent antennae, this will not occur (i.e. it does not have to take turns between communicating pairs).

Considering the coverage area in both cases, we achieved roughly 40 m when compared to the original 10 m of the desktop computer sharing its Internet connection. When using the packet forwarding implementation, we had to exchange SSIDs on the 1st floor (Lo1) and ground floor (Lo2) explicitly. When using WDS, we could roam both floors without connection loss using only Lo1.

As a future work, we will devise the case study for implementing a mesh network using TP-LINK TL-WDR 4300. Therefore, OpenWrt will be the basis to configure part of our mobile ad-hoc network (MANET), which will be implemented in the context of the M.E.N.A.C.E-TRACK system.

Acknowledgment

The authors would like to thank the Research Assistance Foundation of the State of Goiás (FAPEG – Fundação de Amparo a Pesquisa do Estado de Goiás) for the full sponsorship of this research – edict number 006/2012.

References

[1] D. Johnson, K. Matthee, D. Sokoya, L. Mboweni, A. Makan, and H. Kotze, *Building a Rural Wireless Mesh Network A do-it-yourself guide to planning and building a Freifunk based mesh network, version 0.8*, Meraka Institute, South Africa, 2007.

[2] J. F. Kurose and K. W. Ross, *Computer Networking - A Top-Down Approach*, 6th ed. Pearson Education, 2012.

[3] G. Aggelou, *Wireless Mesh Networking*. McGraw-Hill Communications, 2009.

[4] S. Methley, *Essentials of Wireless Mesh Networking*, ser. Cambridge Wireless Essentials Series. Cambridge University Press, 2009.

[5] OpenWrt, "About openwrt." [Online]. Available: http://wiki.openwrt.org/about/start

[6] ——, "Tp-link tl-wdr4300." [Online]. Available: http://wiki.openwrt.org/toh/tp-link/tl-wdr4300

[7] ——, "Beginners' guide to openwrt." [Online]. Available: http://wiki.openwrt.org/doc/howto/user.beginner

[8] ——, "Buyers' guide." [Online]. Available: http://wiki.openwrt.org/toh/buyerguide

[9] ——, "Table of hardware." [Online]. Available: http://wiki.openwrt.org/toh/start

[10] ——, "Openwrt – first login." [Online]. Available: http://wiki.openwrt.org/doc/howto/firstlogin

[11] ——, "Secure your router's access." [Online]. Available: http://wiki.openwrt.org/doc/howto/secure.access

[12] ——, "Rootfs on external storage (extroot)." [Online]. Available: http://wiki.openwrt.org/doc/howto/extroot

[13] ——, "Usb basic support." [Online]. Available: http://wiki.openwrt.org/doc/howto/usb.essentials

[14] ——, "Mounting block devices." [Online]. Available: http://wiki.openwrt.org/doc/techref/block_mount

[15] ——, "Fstab configuration." [Online]. Available: http://wiki.openwrt.org/doc/uci/fstab

[16] subsignal.org, "Luci," 2014. [Online]. Available: http://luci.subsignal.org/trac

[17] OpenWrt, "Luci essentials." [Online]. Available: http://wiki.openwrt.org/doc/howto/luci.essentials

[18] ——, "Openwrt sysupgrade." [Online]. Available: http://wiki.openwrt.org/doc/howto/generic.sysupgrade

[19] ——, "Network configuration." [Online]. Available: http://wiki.openwrt.org/doc/uci/network

[20] ——, "Wireless configuration." [Online]. Available: http://wiki.openwrt.org/doc/uci/wireless

[21] ——, "Firewall configuration." [Online]. Available: http://wiki.openwrt.org/doc/uci/firewall

[22] ——, "The uci system." [Online]. Available: http://wiki.openwrt.org/doc/uci

[23] D. Engwer, "Ieee p802.11 wireless lans - "wds" clarifications," Nortel, July 2005. [Online]. Available: http://www.ieee802.org/1/files/public/802_architecture_group/802-11/4-address-format.doc

[24] IEEE, "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," ANSI/IEEE," Standard, 2003.

[25] OpenWrt, "Client mode wireless (solution using wds)." [Online]. Available: http://wiki.openwrt.org/doc/howto/clientmode#solution.using.wds

[26] ——, "Atheros and mac80211 wds to implement a wireless network bridge." [Online]. Available: http://wiki.openwrt.org/doc/recipes/atheroswds

[27] "Building bridges with linux." [Online]. Available: http://bwachter.lart.info/linux/bridges.html

[28] N. Craft, "Linux setup default gateway with route command," August 2006. [Online]. Available: http://www.cyberciti.biz/faq/linux-setup-default-gateway-with-route-command/

[29] dd-wrt.com, "Wds linked router network," July 2012. [Online]. Available: http://www.dd-wrt.com/wiki/index.php/WDS_Linked_router_network

A Review of Handover Schemes in Overlaid Macro-femto Cellular Networks

Suman Deswal, Anita Singhrova
Department of Computer Science & Engineering, Deenbandhu Chhotu Ram University of Science & Technology, Murthal, India

Abstract: In today's scenario, large number of disparate wireless technologies account for heterogeneity of the access environment. For seam less mobility in such networks the handover plays an important role. In cellular a network, which is an integral part of heterogeneous environment, coverage and subsequently QoS (Quality of Service) are the main issues. Femtocells have emerged as the most promising solution by improving coverage and the Quality of Service to users on one hand and by offloading the macrocells and thus benefitting the service providers on the other hand. Femtocells are the small, low power base stations deployed inside the homes or buildings to provide better coverage in those regions and to increase the capacity of the networks. This paper presents works on the handover management between macrocells and overlaid femtocells. In this direction, the different handover techniques are reviewed for improved throughput, energy efficiency, reduction in unnecessary handovers, network load balancing, interference management and network security. The study demonstrates that more number of objectives achieved leads to better QoS.

Keywords: Femtocells, handover, Quality of Service(QoS), throughput, energy efficiency, network load balancing.

1. Introduction

The wireless revolution today is having a profound impact on the way people work and live. More people all over the world are having a mobile phone than having a PC. The mobile wireless handheld devices have overtaken the wired computers as dominant internet access throughout the world. With the introduction of new and varied mobile devices like smartphones, tablets every now and then in the market, their efficient use requires a serious thought as these devices now a days are not used for voice only but data services are equally important. With the wireless revolution new wireless technologies are also taking shape. A fast research and development has lead to the emergence of Bluetooth, IEEE 802.11 wireless LAN and satellite networks [1]. This growing demand to access communication services that too with mobility has lead to an accelerated technological development towards integration of various types of access technologies. Such a mix of radio technologies and cell types working together seamlessly is called a heterogeneous network. A Heterogeneous Network (HetNet) is based on the coordinated radio network with integrated Wi-Fi, advanced traffic management and high-performance backhaul. As the user moves, he chooses the best access technology which provides the best data speed, high quality and which supports mobility also.

Mobility Management is one of the important issues of Heterogeneous Networks. It is the set of tasks required to supervise the mobile user terminal in a wireless network to check that it is always connected to the network even when moving [2]. Mobility management has several aspects like maintaining Quality of Service, Handover management, Location management and power management etc. A wireless network supports mobility and provides QoS through handover management. It is the process of transferring an ongoing call from one cell to another. Handover process faces several challenges like maintaining the QoS across different systems, deciding the correct handover time, the correct handover decision, packet losses, latency, signaling traffic overhead, security and increased system loads [3].

Out of many existing heterogeneous network, one is femtocell-macrocell integrated network. A femto cellular network is the technology to meet the demands of ever increasing wireless capacity [4] [5]. Existing networks do not have the flexibility to adapt to changes in demand. Moreover much of the traffic is generated indoors and the existing networks suffer from the problem of poor indoor coverage. Various ways of overcoming this challenge are installation of more sites and consequently more base stations, deploying signal boosters or installing more relay stations, add more spectrums to the network to improve the spectral efficiency or reuse the existing spectrum. Undoubtedly, this leads to a reduction in the above cited problems but the overhead in acquiring the permission and cost of installation is also significantly high. Small cells such as microcell, picocell and femtocell can be deployed to enhance the coverage in specific locations and to improve voice and data capacity. The microcells and picocells cannot be deployed inside the house by the user as their base stations are big enough and specific standards are required for installation which can be fulfilled by operator only. Therefore, over a past few years, Femtocells have evolved as a new small cell technology to improve the indoor cellular coverage. A femtocell is a low power, low cost, and user-deployed cellular base station (BSs) with a small coverage range e.g. 30–40 m. A large number of femtocells are deployed in the coverage area of macrocells. The mobile industry is now flooded with femtocells as they provide a good solution to the various problems suffered by macrocellular networks. As most of the calls are made indoors and the signal from the macrocell base station deteriorates very quickly as it reaches indoors and due to the interference in the cell, the true 3G service is available to end users only if there is a good quality signal indoors and the number of users per cell is limited [6]. The femtocell technology helps to offload the traffic from macrocell. In other words, Femtocell is a low cost solution and easy to install which at the same time provides significant coverage indoors and offload macrocells [7].

This paper studies the parameters for handover in different handover algorithms in femtocells and presents their comparison.

1.1 Handovers in Femtocells:

Figure 1 shows an overlaid femtocells deployment over the macrocell i.e. a smallest unit of cellular network. Handover in such a network can be performed in the following three ways [8]:

Hand-In : Hand In takes place when a user moves from a macrocell to femtocell. It is very complex as the user chooses the best femtocell out of several options after considering the neighbouring cells.

Hand-Out : Hand-Out takes place when a user moves from a femtocell to macrocell. It is simple as compared to Hand In as the target macrocell is always one.

Femto to femto handover : It takes place as a user moves out of the boundary of one femtocell and enters into the boundary of other.

Hand-In and Hand out falls under the category of vertical handover as it involves two different access techniques and Femto to Femto handover is horizontal handover as two similar access techniques are used.

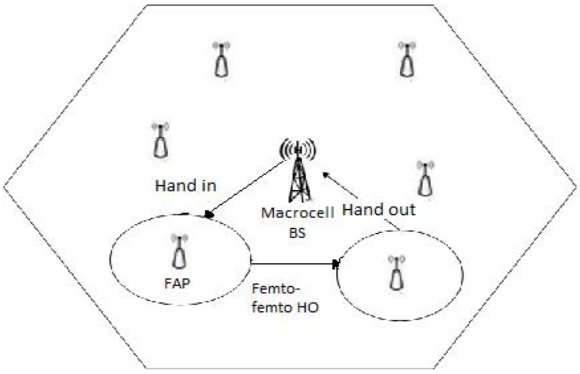


Figure 1: Handover Scenario in overlaid macro femtocellular networks

The number of handovers increases as the femtocells are more densely deployed in a macrocell. Moreover, as the range of a femtocell is very limited, therefore the mobile device tend to leave coverage area of femtocell very soon which leads to handover to either a nearby femtocell or the macrocell. Handover management is one of the most challenging issues in femto cellular network as in any other heterogeneous network. A wide range of schemes have been presented to efficiently handle the handovers [9-14]. The authors have considered a wide range of factors like Received Signal Strength (RSS), velocity of the user, peak bandwidth reception, energy efficiency etc. for performing handover.

2. Literature Survey

A few approaches have been studied for handovers in macro-femto cellular networks. The mechanisms are based on different parameters. However, almost all of the approaches have considered Received Signal Strength (RSS) as one of the parameters for handover. Various other parameters such as speed of mobile device, mobility prediction, available bandwidth etc. are combined to provide improved Quality of Service (QoS) to the users.

In paper [9], the decision for handover from macrocell to femtocell is based on the future mobility pattern prediction scheme to prevent macro femtocell handovers of temporary femtocell visitors who stay in the femtocell coverage area for a short period of time thereby reducing the unnecessary handovers. Each mobile device periodically transmits its movement history to server. The server collects the histories and based on the mobility rules, it generates the mobility pattern. If the received signal strength of the femtocell is greater than the threshold then before performing the handover, the mobile device predicts its next consecutive movements and if the next consecutive movements are within the coverage area of femtocell for enough length of time, only then handover is performed. That is why this handover algorithm is named as smart.

In [10], the author presents a new handover strategy between femtocell and macrocell for LTE (Long Term Evolution) based network. It analyzes the scenarios of hand-in and hand-out after the handover decision. For macrocell to femtocell handover, it first checks whether the user belongs to Closed Subscriber Group (CSG) or not. If mobile device is a CSG and received signal strength and velocity are lower than a threshold value and the application not being a real time application then handover is performed depending on the availability of bandwidth. It reduces the unnecessary handovers and eliminates the cross layer interference.

In paper [11], the authors proposed a handover strategy based on neighbour cell list. A neighbour cell list is created together by macro base station and the neighbour femtocell access points (FAP). The mobile device performs a pre authentication with all the femtocells included in the list.

The handover decision is taken based on the pre-authentication and the received signal strength from the target femtocell access points. This handover algorithm overcomes the hidden FAP problem and also reduces power loss by decreasing the size of neighbour cell list because of the pre authentication.

The authors in [12] presented a handover strategy for hybrid access mode based on Received Signal Strength Indicator (RSSI), velocity of the UE(user equipment), Signal to Interference level (SINR), capacity bandwidth that one Femtocell can accept, the user type and the duration UE maintains the signal level above the threshold level. If the velocity of a mobile device falls below a threshold value, then six base stations with the highest available RSSI are determined. The base station which can support the bandwidth and having the highest SINR is chosen for handover. If the mobile device is a registered user then handover is performed immediately. Un registered users must stay in the femtocell area for the threshold time interval for the handover to occur. In this way it minimizes the unnecessary handovers.

The authors in [13] proposed a Handover Scenario and Procedure in LTE-based Femtocell Networks. It is based on the mobility prediction to reduce unnecessary and very frequent handovers. If the velocity of a mobile device falls below 10km/h but greater than 5km/h then a mobility prediction is performed to predict the heading position of the mobile device. A pro active handover is then performed for the real time traffic and for non real time traffic reactive handover is performed which reduces the unnecessary handovers by postponing the handover as long as possible until the UE reach the target FAP as the result of mobility prediction. The target FAP is chosen based on signal quality (RSSI/CINR) and QoS.

A handover decision algorithm which helps to conserve the energy of the network and provides for load balancing both at the same time is proposed in [14]. The algorithm is based on two main factors viz. the velocity of mobile station and the service type of the mobile station. The user moving at a speed higher than the threshold and undergoing real time transmission will not undergo any macro/femto handover. Therefore unnecessary handover is eliminated. To overcome the wastage of power in the network, the femto base stations are assumed to remain in low power ‘idle mode’ when no user is present in its coverage.

3. Requirements for Handover in Femtocells

This section explains the various requirements to be kept in mind before performing handover. These are as follows:

(i) Improved throughput: Femtocells are introduced in the existing macrocellular network to improve the user throughput and in turn the overall system throughput by efficiently using the available bandwidth. A femtocell reduces the load on macrocell. But at the same time a mobile device moving in from a macrocell should experience minimum call dropping and call blocking. A good bandwidth allocation ensures lower call blocking and call dropping probabilities and hence an increase in throughput. A handover technique from macrocell to femtocell should always help to increase the user and network throughput. For this reason, the bandwidth should be efficiently distributed among maximum number of users.

(ii) Energy efficiency: Energy efficiency has always been an area of concern as mobile phones lose its battery power rapidly. Both network discovery and activation of interface between networks causes battery drainage. Performing handover further enhances battery consumption. So, unnecessary handovers should be avoided for saving power of the device.

(iii) Reduction in unnecessary handovers: As a femtocell covers only a small area, therefore the mobile device may go out of the coverage of femtocell very soon, if moving at a high speed. In that case large number of handovers only leads to wastage of network resources and thus should be avoided.

(iv) Network load balancing: One of the main limitations which the cellular network is suffering from is very high network load which leads to congestion in the network. The idea of introducing femtocells is to divert data and voice traffic to internet and offload macrocells. A good handover technique would also help in offloading traffic from macrocells.

(v) Interference Management: The overlaid integrated femtocell/macrocell networks offer an efficient way to increase access capacity by improving coverage and quality of service. One of the major technical challenges that femtocell networks are facing is their interference behaviour when they are placed within macrocells. A handover technique should always lead to reduced interference levels among networks.

(vi) Increased Network Security: A femtocell access point is located in customer’s location and accesses the operator’s core network through an IP link. As a result, security threats to operator’s core network increases. The network security deals with the policies to prevent and monitor unauthorized access, misuse and modification of network-accessible resources. Network security features should be included in handover technique to attain the highest levels of integrity, authentication, and confidentiality [15].

4. Comparison of the Existing Handover Techniques

In this section existing handover techniques are compared against the objectives for a vertical handover. Table 1 lists the achievement of objectives by each technique.

Table 1: Comparison of handover techniques

S No	Handover technique	Parameters for handover decision	Objectives					
			O1 Improved throughput	O2 Energy efficiency	O3 Reduction in unnecessary handovers	O4 Network load balancing	O5 Interference Management	O6 Network security
1	Smart Handover Decision Algorithm using location prediction for hierarchical macro/femto-cell networks [9]	Mobility prediction, RSS, velocity, time for which device stays in femtocell area			Yes(by identifying the time a user spends in femtocell area)			
2	A new handover strategy between femtocell and macrocell for LTE based network [10]	RSS, available bandwidth, velocity, interference level, real/non-real time transmission, user type			Yes(improved handover decision due to more number of parameters for decision)		Yes(by considering interference level for handover decision)	Yes(by allowing the network access to CSG)
3	Handover management in high-dense femtocellular networks [11]	Neighbour cell list, RSS, SINR, available bandwidth	Yes(by reducing call block/drop)	Yes(by creating reduced neighbour cell list)	Yes(by taking a high value for SINR)	Yes(by increasing macrocell channel release rate)		Yes(preauthentication pertformed)
4	Performance Analysis of Handover Strategy in Femtocell Network [12]	RSS, velocity, interference level, capacity bandwidth, user type			Yes(by delaying the handover for unregistered users)		Yes(by considering interference level)	Yes(registered users are allowed to access network
5	Handover Scenario and Procedure in LTE-based Femtocell Networks [13]	RSS, Mobility prediction, velocity, real/non-real time transmission			Yes(by identifying the time a user spends in femtocell area)			
6	Load balancing with reduced unnecessary handoff in energy efficient macro/femto-cell based BWA networks [14]	RSS, velocity, real/non-real time transmission	Yes(better call quality)	Yes(by introduction of active-idle mode of femto base station)	Yes(by denying handover to slow users or undergoing real-time transmission)	Yes(by setting the macro RSS threshold value above the balanced threshold level)		

Using the requirement parameters listed in section 3, it is found that none of the techniques achieves all the objectives, although maximum number of mechanisms help in the reduction of unnecessary handovers. While handover techniques at S.No 3 and 6 improve throughput and energy efficiency of the network and also provide for load balancing but they fail to look at another important aspect of interference management. The techniques at S.No. 2, 3 and 4 tend to improve security of the network by pre authentication of the user to access the femtocell network. The techniques at S.No. 1 and 5, although able to achieve only a single objective of reducing unnecessary handovers yet they are using a very efficient scheme of mobility prediction to calculate the time a user device will spend in the femtocell area.

The most important aim of overlaying the macrocellular network with femtocells is to provide enhanced end point service accessibility to users and offloading macrocellular network. All the objectives described above leads to improved QoS in one way or another. The QoS can be taken as a function of all the objectives mentioned in section 3 i.e.

QoS = f (O1,O2,O3,O4,O5,O6)

In above equation, if all the objectives are assigned equal weights, then techniques at S.No. 4 and 6 provide the most significant improvement in quality to users. However, different objectives can be assigned different weights based on their priorities and impact on quality.

5. Conclusion:

The femtocells have emerged out as a promising solution to improve the network coverage indoors and to offload the traffic from already overloaded macrocells. Handover is an important area of research in overlaid macro femtocellular networks. In this paper, a few works on the handovers in femtocell overlaid macrocell have been described. The objectives for handover are proposed which include throughput, Energy efficiency, Reduction in unnecessary handovers, Network load balancing, Interference Management, Network security. The comparison of existing handover techniques has been done against the stated objectives. The comparison indicates the need of an efficient handover technique that tries to cover all the objectives. To achieve all the objectives is certainly difficult but more number of objectives achieved leads to improved Quality of Service to user.

References:

[1]. Debabrata Sarddar, Shovan Maity, Arnab Raha, Ramesh Jana, Utpal Biswas, M.K. Naskar, ”A RSS Based Adaptive Hand-Off Management Scheme In Heterogeneous Networks” IJCSI International Journal of Computer Science Issues, vol. 7, no. 6, ISSN (Online): 1694-0814, Nov 2010.

[2]. G. Giannattasio, A Guide to the Wireless Engineering Body of Knowledge (WEBOK). Wiley: 2009.

[3]. A. Singhrova, N. Prakash, “ Vertical handoff decision algorithm for improved quality of service in heterogeneous wireless networks,” IET Communications, vol. 6, no. 2, pp. 211-223, 2012.

[4]. 1. S Yeh, S Talwar,”WiMAX femtocells: a perspective on network architecture, capacity and coverage” IEEE Commun. Mag., vol.46, no. 10, pp. 58–65, 2008.

[5]. RY Kim, JS Kwak, K Etemad,” WiMAX femtocell: requirements, challenges and solutions” IEEE Commun. Mag., vol. 47, no. 9, pp. 84–91, 2009.

[6]. Amevi Acakpovi, Henry Sewordor, Koudjo M. Koumadi,” Performance Analysis Of Femtocell in an Indoor Cellular Network” IRACST – International Journal of Computer Networks and Wireless Communications (IJCNWC), vol. 3, no. 3, ISSN: 2250-3501, June 2013.

[7]. M Z Chowdhury, Y M Jang, Z J Haas,” Cost-effective frequency planning for capacity enhancement of femtocellular networks”, International Journal of Wirel. Personal. Commun., vol. 60, no. 1, pp. 83–104, 2011.

[8]. Z. Jie and D. Guillaume, Femtocells: Technologies and Deployment. UK: Wiley, 2010.

[9]. Byungjin Jeong, Seungjae Shin, Ingook Jang, Nak Woon Sung, Hyunsoo Yoon,” A Smart Handover Decision Algorithm Using Location Prediction for Hierarchical Macro/Femto-Cell Networks” IEEE conference on Vehicular Technology Conference VTC Fall 2011, IEEE, pp. 1-5, ISSN 1090-3038, 978-1-4244-8327-3/11, 2011.

[10]. Shih-Jung Wu,” A New Handover Strategy between Femtocell and Macrocell for LTE-based Network”, Fourth IEEE International Conference on Ubi-Media Computing, pp. 203-208, 978-0-7695-4493-9/11, 2011.

[11]. Mostafa Zaman Chowdhury, Yeong Min Jang,” Handover management in high-dense femtocellular networks” EURASIP Journal on Wireless Communications and Networking 2013, doi:10.1186/1687-1499-2013-6, http://jwcn.eurasipjournals.com/content/2013/1/6.

[12]. Azita Laily Yusof, Siti Sabariah Salihin, Norsuzila Ya’acob, and Mohd Tarmizi Ali,” Performance Analysis of Handover Strategy in Femtocell Network” Journal of Communications vol. 8, no. 11, Nov 2013.

[13]. Ardian Ulvan, Robert Bestak, Melvi Ulvan,” Handover Scenario and Procedure in LTE-based Femtocell Networks” UBICOMM 2010 : The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, pp. 213-218, ISBN: 978-1-61208-100-7, 2010.

[14]. Prasun Chowdhury, Anindita Kundu, Iti Saha Misra, Salil K Sanyal,” Load balancing with reduced unnecessary Handoff in energy efficient Macro/femto-cell based BWA networks”, International Journal of Wireless & Mobile Networks (IJWMN) vol. 4, no. 3, June 2012.

[15]. 3GPP TS 33.320 v0.2.0: “3GPP Security Aspect of Home NodeB and Home eNodeB; Release 9”.

Performance Evaluation of NeBuST-wide Burst Sensor Data Transmission Protocol

Hiroaki HIGAKI
Department of Robotics and Mechatronics
Tokyo Denki University
Senjyu-Asahi 5, Adachi, Tokyo 120–8551 Japan
Email: hig@higlab.net

***Abstract*—In an event-based sensor network, a sequence of burst sensor data messages are required to be transmitted in shorter transmission delay without losses. However, it is difficult due to limited communication buffers in intermediate sensor nodes in a wireless multihop transmission route to the sink node. This paper proposes NeBuST-wide routing and transmission protocols in which sensor data messages are transmitted along dynamically determined transmission routes in message-by-message manner. In case of a full communication buffer in the next-hop wireless sensor node, an intermediate node forwards data messages to another neighbor node to make detour transmissions for achieving shorter transmission delay and for avoidance of losses of the sensor data messages. In addition, this paper evaluates the performance of NeBuST-wide in comparison with the original NeBuST to make clear the effect of wider distribution of burst sensor data message transmissions and in comparison with a static multi-route transmission protocol.**

I. Introduction

A wireless sensor network consists of multiple wireless sensor nodes and a sink node and sensor data messages are required to be transmitted within the required time. Until now, various methods for sensor data message transmissions have been proposed for sensor networks with regular and periodical transmission requests. However, in an event-driven sensor network, distribution of sensor data transmission requests is wide and the memory capacities in wireless sensor nodes are not always enough. Hence, buffer overflow might be caused and some sensor data messages might be lost. For reliable and stable transmissions of event-driven sensor data messages, they are forwarded only when there are enough amount of free communication buffer in a next-hop sensor node. In addition, in spite of this forwarding condition, end-to-end transmission delay of a sequence of sensor data messages are required to be shorter. This paper proposes NeBuST-wide which is an extension of NeBuST which achieves reduction of transmission delay caused by a sequence of filled communication buffers. NeBuST-wide uses dynamically determined detour routes more widely distributed than the original NeBuST.

II. Wireless Sensor Networks

A wireless sensor network consists of multiple wireless sensor nodes with wireless communication devices and a sink node. Sensor data messages containing sensor data achieved by the sensor nodes are transmitted to the sink node. A wireless multihop transmission route $R = ||S_0 \dots S_n\rangle\rangle$ from a source sensor node S^s $(= S_0)$ to a destination sink node S^d $(= S_n)$ is a sequence of intermediate sensor nodes S_i $(0 < i < n)$. Each intermediate sensor node S_i forwards sensor data messages from S_{i-1} to S_{i+1}.

There are two kinds of sensor data messages as follows:

- Regular sensor data messages which are transmitted periodically from various sensor nodes.
- Event-driven sensor data messages which are transmitted from dedicated sensor nodes in an on-demand manner.

This paper discusses transmission methods for the latter. For transmissions of event-driven sensor data messages, constant high traffic communication is not always required to be supported. In case of an occurrence of an event, sensor data messages are burstly initiated and required to be transmitted without losses in a shorter delay. It is assumed that the capacity of the wireless sensor network is enough to transmit all the regular and event-driven sensor data messages. However, since many event-driven sensor data messages are initiated burstly, communication buffers in intermediate sensor nodes are temporarily filled. It may cause losses of the event-driven sensor data messages and their end-to-end transmission delay may get longer which are inadequate for sensor network applications.

III. Problems

Resources in a wireless sensor node are limited for smaller-sized implementation and its lower price. Not only its battery capacity but also its memory capacity are limited. Thus, in a wireless multihop transmissions of a sequence of event-driven sensor data messages along a wireless multihop transmission route $R = ||S_0 \dots S_n\rangle\rangle$, communication buffers in intermediate nodes S_i may be filled. Such filled communication buffers are caused around the sink node and intermediate sensor nodes at which multiple transmission routes join together when multiple sensor nodes transmit event-driven sensor data messages simultaneously as shown in Figure 1.

In wireless multihop communication, S_i contends with S_{i-2}, S_{i-1}, S_{i+1} and S_{i+2} which reduces transmission opportunities in S_i. In widely available contention-based wireless LAN protocols [5], [6], contention avoidance is realized by introduction of randomly determined backoff periods. This provides long-term feasibility due to probably distributed backoff periods; however, short-term infeasibility is inevitable. This

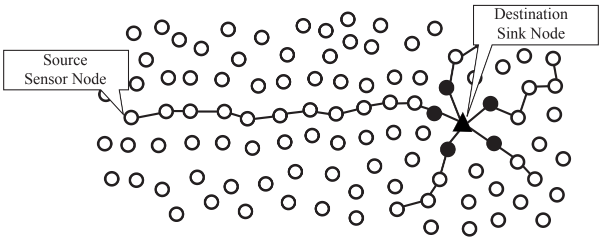


Fig. 1. Filled Buffers in Intermediate Sensor Nodes around Sink Node.

causes maldistribution of transmission opportunities which causes filled communication buffers in the intermediate sensor nodes.

While a communication buffer in S_i is filled, it is impossible for S_i to receive a sensor data message forwarded by S_{i-1}. In this case, S_{i-1} does not receive an acknowledgement message from S_i and retransmits the sensor data message. After the predetermined numbers of retransmissions, S_{i-1} discards the sensor data message. Each intermediate sensor node S_i stores sensor data messages in its communication buffer for their retransmissions while it does not receive an acknowledgement message from its next-hop sensor node S_{i+1}. Due to less amount of communication buffers and maldistribution of transmission opportunities, the communication buffer of S_i is filled temporarily and S_{i-1} cannot forward sensor data messages to S_i. For sensor data message from S_{i-1} to S_i, S_i is required to transmit a buffered sensor data message to S_{i+1} to make space in its communication buffer; however, it may contend with its 1-hop and 2-hop previous-hop intermediate sensor nodes S_{i-1} and S_{i-2} which are exposed and hidden nodes. Thus, communication buffers in $S_{i-1}, S_{i-2}, \dots$ are also filled; that is, a sequence of communication buffers are filled as in Figure 2. Due to the same reason discussed above, it requires much more longer time to clear the sequence of filled communication buffers.

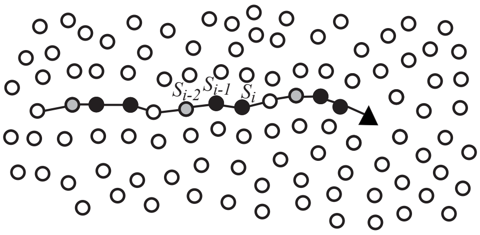


Fig. 2. Sequence of Filled Communication Buffers.

IV. Related Works

Wireless LAN protocols are not designed for wireless multihop communications. Its collision avoidance methods CSMA/CA and RTS/CTS reduce the performance of multihop transmissions of a sequence of sensor data messages. MARCH [7] is a modified protocol for RTS/CTS control message exchanges to adapt to multihop transmissions. RH2SWL [2] avoids contentions between hidden nodes by a transmission power based routing protocol. These protocols concentrate only on transmission delay of a sequence of sensor data messages but do not consider the buffer capacities in intermediate sensor nodes. Various scheduling algorithms [1] for sensor data message transmissions have been proposed which reduce end-to-end transmission delay and indirectly reduce the required communication buffer capacities in intermediate sensor nodes. However, the transmission schedule is not flexible and an advertising method of the schedule and a close synchronization method among sensor nodes are required. Hence, it is difficult to apply them to transmissions of event-driven sensor data messages.

The authors have proposed NeBuST (Neighbor Buffering for Congested Sensor Data Transmissions) by which sensor data messages waiting for their spaces in the next-hop sensor nodes are forwarded to neighbor nodes nearer to the sink node for shorter end-to-end transmission delay. In order to store sensor data messages into communication buffers of sensor nodes nearer to the sink node, an intermediate sensor node S_{i-1} forwards them to different neighbor sensor node S'_i from its next-hop one S_i as shown in Figure 3. S'_i is a neighbor sensor node of both S_{i-1} and S_{i+1}. In case that S_{i-1} tries to forward a sensor data message to S_i whose communication buffer is filled, S_i transmits back a nack (negative acknowledgement) control message to notify its filled communication buffer. On receipt of the nack control message, S_{i-1} tries to forward it to S'_i. If it is successful, transmission delay of the sensor data message is expected to be reduced since S_{i+1} is a neighbor sensor node of S'_i. Otherwise, i.e., the communication buffer of S'_i is also filled and S_{i-1} receives a nack control message, S_{i-1} stores the sensor data message to its own communication buffer.

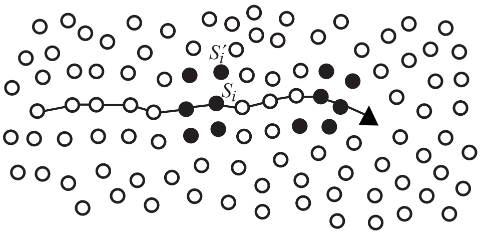


Fig. 3. Neighbor Buffering in Original NeBuST.

V. Proposal

A. NeBuST-wide

For transmissions of burst event-driven sensor data messages, the original NeBuST avoids additional transmission delay caused by a sequence of filled communication buffers by using communication buffers in 1-hop neighbor sensor nodes as backups. Sensor data messages are transmitted nearer to the destination sink nodes even while communication buffers in intermediate sensor nodes are filled. In a wireless multihop transmission route $R = ||S_0, \dots, S_n\rangle\rangle$, sensor data messages buffered in a backup sensor node S'_i is forwarded to S_{i+1} which is a next-hop node of S_i in R.

As discussed in the previous section, communication buffers in any intermediate sensor nodes may be filled independently of distance to the sink node. Thus, a sequence of filled communication buffers may be configured even far from the sink node. In this case, the original NeBuST also contribute to reduce the transmission delay by using communication buffers in 1-hop neighbor sensor nodes as shown in Figure 4. However, for avoidance of a sequence of filled communication buffers far from the sink node, since few concurrent transmissions of burst event-driven sensor data transmissions through neighbor sensor nodes are expected, wider distribution of sensor data message transmissions may reduce more transmission delay. Thus, this paper proposes an extension NeBuST-wide. Here, a backup sensor node S'_i of S_i does not always transmit to S_{i+1} but also to its backup node S'_{i+1}. In addition, S'_i also transmits sensor data messages to a backup node S''_{i+1} of S'_{i+1} which may be away from R. Therefore, as shown in Figure 5, if a communication buffer in S_i is filled, transmissions of sensor data messages are widely distributed and sensor data messages are transmitted and buffered in sensor nodes nearer to the sink node. The widely distributed transmissions becomes independent of R and shorter end-to-end transmission delay is achieved.

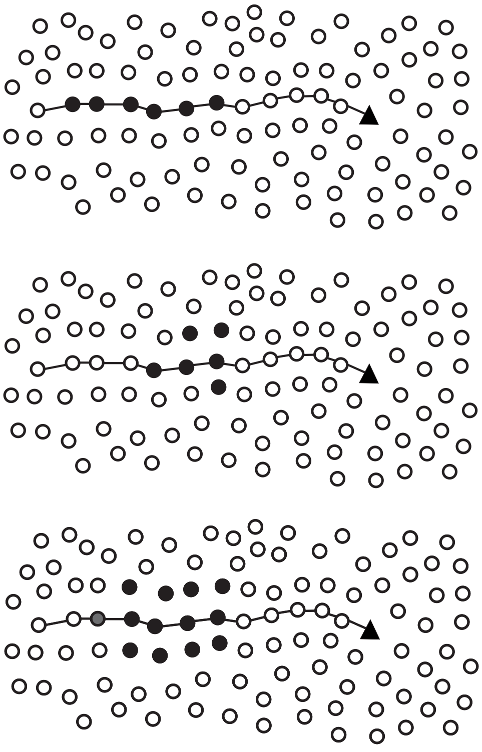


Fig. 4. Filled Communication Buffers in Original NeBuST.

B. Routing Protocol

In the extended NeBuST-wide proposed in the previous subsection, not only 1-hop neighbor sensor nodes S'_i of intermediate sensor nodes S_i in a wireless multihop transmission route $R = ||S_0, \ldots, S_n\rangle\rangle$ but also other sensor nodes are engaged in transmissions of sensor data messages from S_0. There are many possible intermediate nodes widely distributed. Hence, differently from the reactive routing protocol in the original NeBuST, NeBuST-wide adopts a proactive routing protocol in which each sensor node maintains its next-hop sensor nodes for lower routing overhead. One of the possible methods for all the sensor nodes to achieve and update their next-hop sensor nodes is periodical flooding of a routing control message from the destination sink node.

Here, based on a well-known ad-hoc routing protocol

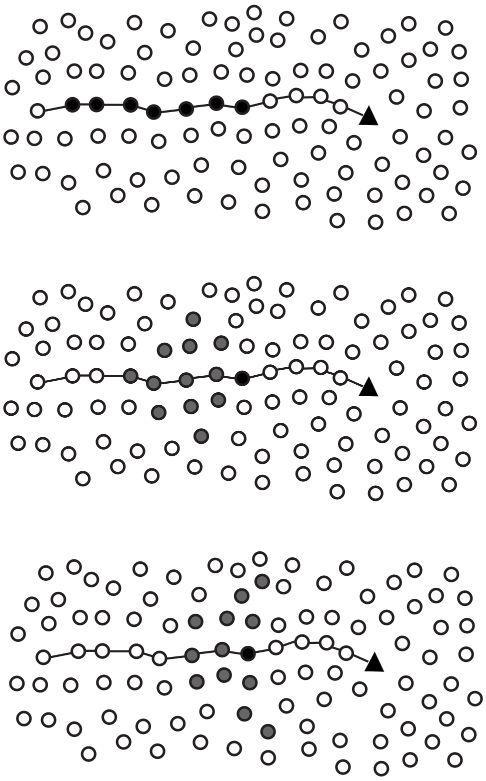


Fig. 5. Filled Communication Buffers in NeBuST-wide.

TORA [4], each sensor node achieves and keeps its hop counts to the sink node. The sink node broadcasts a routing control message to which the initial hop count 0 is piggybacked. On receipt of the first arrived routing control message, a sensor node increments the piggybacked hop count and broadcasts it. On receipt of the last arrived routing control message, a sensor node determines its hop count as the minimum hop count piggybacked to the received routing control message. Then, it advertises a pair of its identification and hop count to its neighbor sensor nodes by broadcasting a routing control message containing it. By receipts of the routing control messages, a sensor node determines its next-hop nodes advertising the minimum hop counts. In addition, it also achieves its possible previous-hop sensor nodes whose hop counts are larger than its own hop count. A sensor node forwards sensor data messages to one of its next-hop sensor nodes whose communication buffers are not filled.

C. Protocols

This section describes NeBuST-wide routing and data message transmission protocols proposed in this section.

[Routing Protocol]

1) The sink node broadcasts a routing control message $Rreq(s)$ with a sequence number s periodically[1]. $Rreq(s)$ also carries a hop-count $Rreq(s).hop$ from the sink node which is initially 0.
2) On receipt of the first copy of $Rreq(s)$, a wireless sensor node S_i registers $Rreq(s).hop + 1$ as its temporary hop

[1] The interval depends on the degree of topology change of the wireless sensor network caused by mobility and failures of wireless sensor nodes.

count $thop_i$. Then, S_i broadcasts a copy of $Rreq(s)$ where $Rreq(s).hop := thop_i$.

3) On receipt of another copy of $Rreq(s)$, only if $Rreq(s).hop + 1 < thop_i$, S_i registers $Rreq(s).hop + 1$ as $thop_i$.

4) If S_i has received copies of $Rreq(s)$ from all its neighbor wireless sensor nodes, S_i registers $thop_i$ as its final (minimum) hop-count hop_i to the sink node. S_i broadcasts a routing control message $Rrep(s)$ to which hop_i is piggybacked as $Rrep(s).hop$.

5) On receipt of the $Rreq(s)$ from S_i, a neighbor wireless sensor node S_j registers a pair $\langle S_i, hop_i\rangle := \langle S_i, Rreq(s).hop\rangle$ into its routing table. □

A routing table of S_i is a set of pairs of its neighbor wireless sensor nodes S_j and their hop counts hop_j to the sink node. If $hop_j < hop_i$, S_j is a candidate of its next-hop node. S_j with the minimum hop count hop_j is a default next-hop node of S_i and S_i forwards sensor data messages to S_j if a communication buffer of S_j is not full, i.e., S_j does not return a $nack$ control message for a reply to a data message.

[Data Transmission Protocol]

1) A wireless sensor node S_i which initiates a multihop transmission of a data message or receives a data message from one of its neighbor nodes transmits the data message to its default next-hop node S_j with the minimum hop count hop_j to the sink node.

2) If S_i receives an ack control message from S_j, S_i deletes the data message and the transmission is complete.

3) If the receipt of the ack control message is timeout, S_i retransmits the data message to S_j.

4) If S_i receives a $nack$ control message from S_j or the number of retransmissions reaches the predetermined upper limit, S_i transmits the data message to its another candidate of its next-hop node S_j whose hop count is less than its hop count, i.e., $hop_j < hop_i$, according to Step 1).

5) If no ack control message is received from its all next-hop candidate nodes, S_i stores the data message to its communication buffer. After a certain longer interval, S_i retries the transmission from Step 1). □

VI. Evaluation

This section evaluate the performance of NeBuST-wide. First, we evaluate it by comparison with the original NeBuST. During transmissions of a sequence of sensor data messages, sensor nodes nearer to the sink node than a threshold distance suspend forwarding sensor data messages in order to cause filled communication buffers temporarily. Then, the sensor nodes restart forwarding and end-to-end transmission delay of the sensor data messages are measured. As shown in Figure 6, sensor nodes with 100m transmission range and communication buffers for 5 sensor data messages are located in a lattice with 60m spaces. A source sensor node is 30 hops away from a destination sink node along a straight transmission route and transmits 30–50 sensor data messages per second for 0.5–2.0 second duration. Sensor nodes within 300m from the sink node suspend communication by transmitting back nack control messages in duration 1.0 second after starting transmissions of sensor data messages. Here, the required time to transmit all the burst sensor data messages to the sink node is measured in the naive multihop transmission with nack control message returns, in the original NeBuST and in NeBuST-wide.

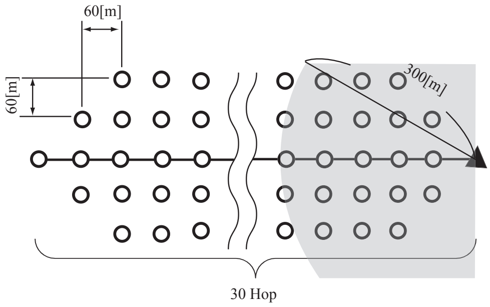


Fig. 6. Simulation Setting.

Figure 7 shows the results with 1.0 second suspension period. In comparison with the naive multihop transmissions with nack control message returns, NeBuST reduces 4.67% transmission delay. Thus, neighbor buffering in NeBuST contributes to reduction of end-to-end transmission delay. Furthermore, NeBuST-wide achieves 15.9% shorter transmission delay. Thus, wider distribution of sensor data message transmissions by extended neighbor buffering effects on the reduction of end-to-end transmission delay.

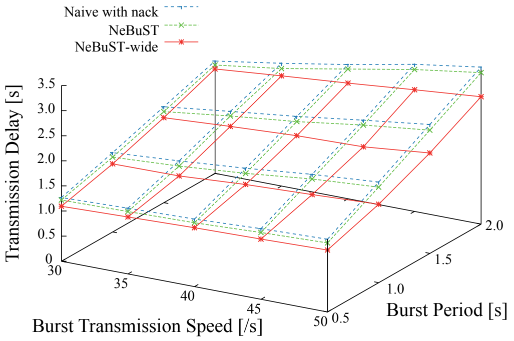


Fig. 7. Transmission Delay with 1.0s Suspension.

Figure 8 shows instances of transmission delay of all the sensor data messages in cases of 2.0 second burst period with 50 sensor data messages per second and 1 second suspension period. In NeBuST and NeBuST-wide, the front sensor data messages are buffered in an intermediate node along a transmission route and requires longer transmission delay; however the other sensor data messages are transmitted through detour routes even during the suspension period and the reduction of transmission delay is realized. Hence, though the tail sensor data messages require longer transmission delay in the naive transmissions with nack control message returns due to the effect of the sequence of filled communication buffers, NeBuST and NeBuST-wide are less effected and transmission delay of the tail sensor data messages are much shorter than the others especially in NeBuST-wide.

Next, we compare it with another multi-route routing protocol AODVM [3] for node-disjoint ad-hoc routing. In a 2,000m × 2,000m square field, the sink node is at the center of the field and 3,000 wireless sensor nodes with a 100m wireless transmission range and a communication

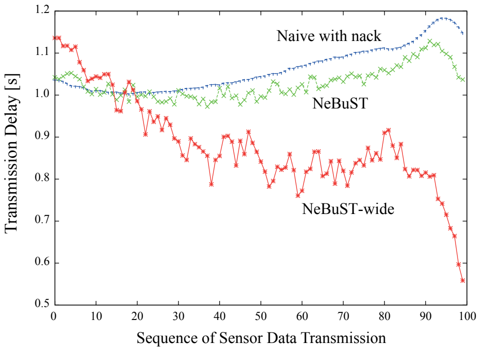


Fig. 8. Transmission Delay of 100 Sensor Data.

buffer for 5 data messages are randomly distributed. AODVM detects multiple transmission routes in advance; however, NeBuST-wide dynamically changes transmission routes in a message-by-message manner. Figure 9 shows detected two node-independent transmission routes in AODVM and Figure 10 shows a default transmission route in NeBuST-wide. The randomly selected source wireless sensor node initiates 30–50 burst data messages in each second and the burst period is 1.5–3.0 second. For comparison of the two protocols, during 1.0–2.0 second in simulation time, all wireless sensor nodes return *nack* control messages in order to buffer some data messages in transmission.

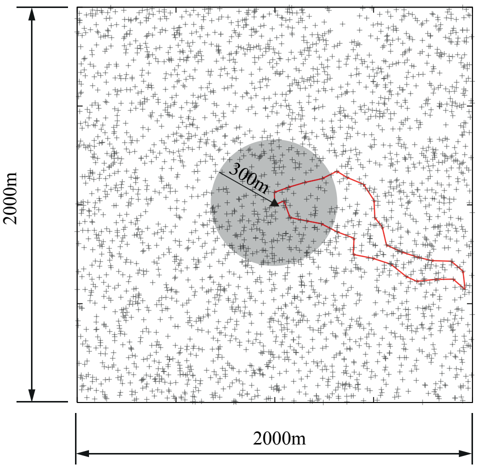


Fig. 9. Transmission Routes in AODVM.

VII. Conclusion

This paper proposes NeBuST-wide as an extension of NeBuST which realizes shorter delay transmissions of event-driven sensor data messages. In spite of less memory capacities in wireless sensor nodes, it avoids losses of sensor data messages caused by filled communication buffers and improves end-to-end transmission delay by wide distribution of sensor data message transmissions away from the original transmission route. For NeBuST-wide, routing and sensor data message transmission protocols are designed and the performance is evaluated by comparison with the naive transmissions and the original NeBuST. The proposed NeBuST-wide achieves shorter transmission delay due to avoidance of effects of sequences of filled communication buffers.

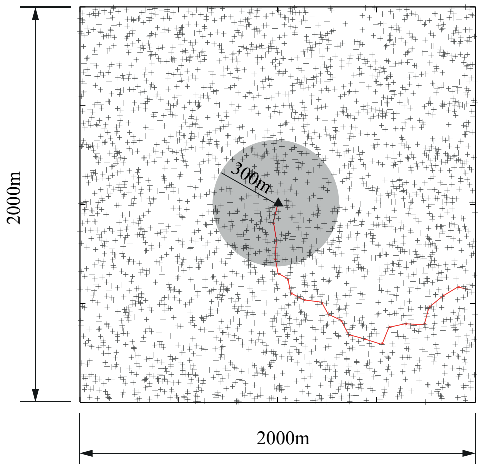


Fig. 10. Default Transmission Route in NeBuST-wide.

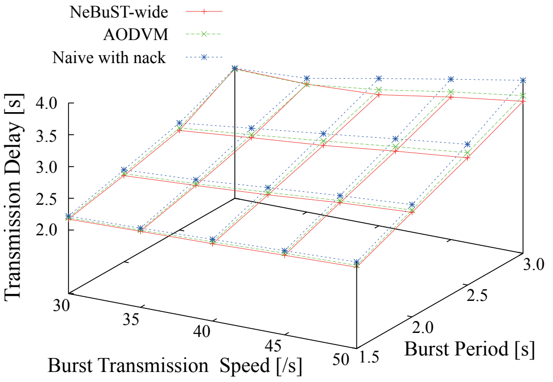


Fig. 11. shows the average transmission delay of all the burst sensor data messages.

References

[1] Akamine, R. and Watanabe, T., “On a Scheduling Method for Reducing MAC Collision in Sensor Networks” IEICE Technical Report, Vol. 107, No. 293, pp. 63–67 (2007).

[2] Matsumura, S. and Higaki, H., “Extension of RH2SWL for

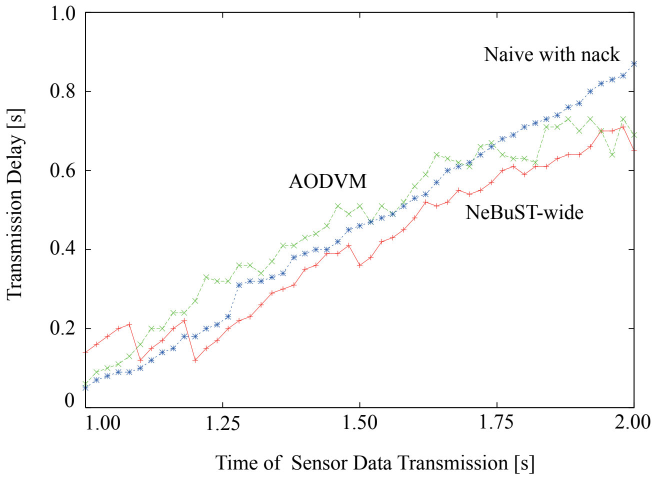


Fig. 12. Average Transmission Delay of Data Messages.

Collision-Free Data Message Transmissions by Subsidiary Channel in Wide-Area Wireless Multihop Networks," Proceedings of the 11th IEEE International Wireless Communications and Networking Conference, CD-ROM (2010).

[3] Motegi, S. and Horiuchi, H., "AODV-Based Multipath Routing Protocol for Mobile Ad Hoc Networks," IEICE Transactions on Communications, Vol. E87-B, No. 9, pp. 2477–2483 (2004).

[4] Park, V. and Corson, S., "Temporally-Ordered Routing Algorithm (TORA) Version 1 Functional Specification," Internet Draft, draft-ietf-manet-tora-spec-04.txt (2001).

[5] Perkins, C.E., "Ad-Hoc Networking," Addison-Wesley (2001).

[6] Perkins, C.E. and Royer, E.M., "Ad hoc On-Demand Distance Vector Routing," RFC 3561 (2003).

[7] Toh, C.K., Vassiliou, V., Cuichal, G. and Shih, C.H., "MARCH: A Medium Access Control Protocol for Multiple Wireless Ad Hoc Networks," Proceedings of the IEEE Military Communications Conference, pp. 512–516 (2000).

New Mobile Digital Signature Solution Based on Micro SD

Khaled Moawad (khaledauc99@hotmail.com), Mostafa Abdel-Aazim (melbakary@aast.edu),
Bahaa Hassan(bahaa.hasan@asc-egypt.org)
The Arab Academy for Science, Technology and Maritime Transport

***Abstract*- Mobile payment is receiving such a significant care because it enables an easy payment method. It can replace the traditional payment means. However, m-payment over open devices and networks reveal security challenges related to the phone & open wireless communication nature. Mobile phone nature as small device, low cost, low power and portable, it cause that mobile cannot do complicated computing encryption. Moreover the evolving of phone viruses is also a huge threat to the mobile payment. At the same time, open wireless communication can be easily tapped by illegal user. Illegal user can intercept the information and change its content. This paper is going to present a mobile payment solution with distributed key based on current mobile multi-interface. It uses an additional equipment (micro SD) to achieve keys distribution and improve the mobile encryption capability. Moreover it uses J2ME security architecture to do data encryption, digital signing, and identity authentication to secure the wireless communication. The proposed system is to raise the security level of mobile payment to a standard that is legally accepted without affecting the user experience or raise the cost of the implementation.**

Key words:
Payment Security; Mobile Payment; Multi-interface Authentication; Mobile Encryption; Micro SD encryption.

I. INTRODUCTION

By time more and more people depend on internet to do their business by using e-commerce. Now, internet is being used to access online services such as e-commerce, e-voting, e-banking, e-government [1, 2]. Security is an important issue in internet based services. Most of online services uses passwords, personal identification numbers and keys to access their account. These types of authentication cannot verify the real identity of the user who is doing the transaction. Using any online service today always need username and password to authenticate. This mechanism is not secure enough as password can be easily hacked by the man in the middle and later used to get the user information.

Financial organization that offers online services should use stronger secure method to authenticate their customer. The current mechanism that based on password is not enough to protect user information against identity theft. As a result any hacker can gain access on confidential information like credit card number.

Single interface authentication increases the risk of revealing customer information by hacker or middle man. A need is rising for a multi interface authentication technique for secure web transactions and to increase trust in web user identity.

II. PROBLEM DEFINITION

Many organization uses username and password (one interface) as authentication credential for remote login. Using the right password will grant the user an access over company database, email account and other information. But password are naturally is not secure enough. Password s can be cracked or easily guessed. Hackers can use many ways to get the password like [3]:

A. Phishing attacks which trick the user into providing access information.
B. Key-loggers and "spyware" which clearly capture access information.
C. User Session Hijacking – Attacker gets control over the active user session and monitors all user activities.

Moreover, it‘s very difficult to know who access your account or even if someone is accessing your account illegally [11].

The most concrete way to give a better security is to add another level of security. Since the password is something that user know, we can add something that user also can have.

	PIN/ password List	One Time password OTP Token	Mobile SMS- OTP	PKI smart card + reader
Security	☹	😐	😐	🙂
Easy to use	🙂	☹	🙂	☹
Mobility	😐	☹	🙂	☹
Low Usage Cost	😐	😐	🙂	😐
Low Maintenance	😐	😐	🙂	🙂
Legal Qualified signing	☹	☹	☹	🙂

🙂 : Compliant ☹ : Not Compliant 😐 : Partially Compliant

Fig.1 different M-Payment solutions for internet based system

In Fig.1 we have examined many tokens or hardware that user can have as a second authentication interface. As a result, none of them is legally accepted expect PKI smart card, however it is not easy to use and very difficult user to carry it around. According to available options, if we need an accepted legally secured transaction, we will complicate the system implementation which is not accepted if we are targeting large number of users. Moreover the cost of the system will be very high.

III. EXISTING PAYMENT SYSTEM & PROTOCOLS

There are many types for M-payment solutions for internet based system. M-payment systems can be categorized into three major categories [4].

A. Account-based system: each customer has an account that managed by third party. The customer can perform pre-paid or post-paid financial. There are three sub-categories for account-based M-payment systems [5]:
   - Payment system based on mobile phone which enable customer to do commerce transaction.
   - Smart card payment system.
   - Credit-card M-payment system which enable customer to transaction through their mobile devices.

B. Mobile POS (Point OF Sale) payment system which customers can purchase by using a vending machines or online with their mobile device. It can be divided into :
   - An automated POS like any fixed terminal like vending machine and parking meters.
   - Attended POS payment, which customer can make payments with assistance from third party like taxi driver, counter clerk, etc [6].

C. E-wallets or E-cash: customers can stress cash from their debit/credit card into tier E-wallet. E-cash can be used to pay micro payments or small payments [7].

There are two existing protocols are being used in all present's systems:

A. Secure Electronic Transaction (SET), is a protocol used in existing credit card payment system to secure the web based financial transactions as well as authentication of transaction. SET is considered as a standard international protocol [8].

In a basic scenario, customer will access the seller website using normal computer to select goods or services and pay it through credit/debit card. SET protocol will support the customer to do online payment to any seller who can accept online banking facilities.

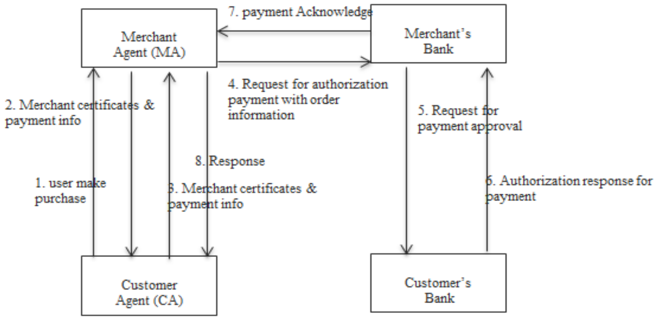

7. payment Acknowledge
Merchant Agent (MA)
Merchant's Bank
4. Request for authorization payment with order information
2. Merchant certificates & payment info
5. Request for payment approval
8. Response
1. user make purchase
3. Merchant certificates & payment info
6. Authorization response for payment
Customer Agent (CA)
Customer's Bank

Fig.2 Transaction flow in Secure Electronic Transaction (SET) protocol

A brief description of SET protocol steps as follow

Step 1: Customer logged into the seller web site to select goods or services. Web site will show the total cost of the goods/ services with taxes and shipping cost.
Step 2: The system asks the customer to choose payment method.
Step 3: Customer will fill payment info to pay through credit card.
Step 4: Merchant agent will raise a request to merchant bank for payment and customer authorization.
Step 5: Merchant bank will contact customer bank for payment authorization.
Step 6: Customer bank will reply back with authorization response. .
Step 7: Merchant bank will reply back to the agent with the customer bank response
Step 8: Customer will receive a web notification based on transaction processing result.

Most of the banks provide online service to enable customers to make payments for goods/services. But they need to be registered before in the bank. Each customer is provided with bank account number and account type. The bank will generate a transaction reference and be provided to the customer for tracking purposes.

Disadvantage of SET

- SET protocol is designed to maintain the normal transaction flow (Customer Agent- Merchant Agent-Merchant Agent Bank). There is a need for customer identity verification.
- There is no notification with the account balance to the customer or asking for verification.
- SET is designed only for card based(credit/debit) transactions. Account transaction is not supported or any other singing/authorization activity.

B. E-Wallet or Digital cash is a way of money transfer. This method is electronic money transfer using computer via internet or on mobile via GPRS (General Packet Radio Service) connection. This method is easy and flexible to pay small payment. However, it has many disadvantages which include fraud, device failure, tracking of specific person. Most common implementation of the E-wallet is the Personal wallet [12].

Personal wallet is a software or hardware installed in the user machine. There is no need for any server. The user saves the credit information locally.

Advantages:

- Its implementation cost is very low.
- It does not need any server, so customer can do offline payment.

- It can be used through smart card or secret pin.
- All customers private information is under user control, as it is save to the local device.

Disadvantages:

- It needs a huge amount of space on user machine, as all logs stored on user machine.
- As the information stored on user machine. It presents a single point of failure. It also opens a lot of possibilities for hacking.

IV. MOBILE DIGITAL SIGNATURE SOLUTION BASED ON MICRO SD.

We need a strong authentication mechanism when dealing with electronic money transaction. Single interface authentication is not considered the right choice to use. Therefore, we need multi-interface authentication. In the proposed system we will add a second authentication layer using SMS. We believe that the customers will carry their phone and will be able to receive and send SMS. After getting the SMS user can confirm his choice. The secure Micro SD will take the responsibility to sign the SMS. Micro SD will provide a physical space which is not accessible by phone application.

Micro SD design and device communication:

Customer can use the micro SD as normal storage, however the secure element that will do the signing, it is not accessible unless through a special procedure.

In Fig3, we are presenting the micro SD architecture overview. As shown, the micro SD from high level has no difference than any normal micro SD. As a result it can be used in any phone that has SD slot. It does not need any special requirements to be attached to a mobile phone.

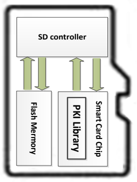


Fig.3 Micro SD Architecture overview

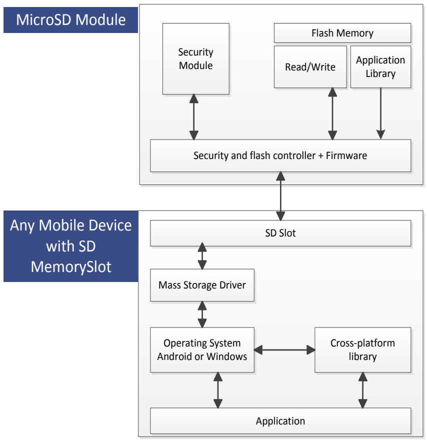


Fig.4 Micro SD Communication flow

As shown in fig4 we are presenting the communication flow of the secure micro SD. As shown, the micro SD has a great flexibility, as micro SD storage can be accessed simultaneously with the security functions.

When a customer receives a verification SMS (with specific Short Code) an application will be activated on the phone asking the user to enter Pin code then a screen will open with the SMS content. Customer will have the option to choose to sign or not. Then the application will send the SMS back with the customer response. All these operations will happen with support from micro SD to encrypt/ decrypt using PKI.

By using micro SD, we have achieved the following:

1- Decryption of Received SMS will happen in secure physical place with no chance that any other application on the phone can access it.

2- In case of losing the phone no one can access the application as he will not have the Pin code. This is much more securing the VISA Card.

3- Micro SD will not work in any other phone. Micro SD will be linked to the phone through IMEI (International Mobile Station Equipment Identity) phone number. However, it is still can work as a normal storage.

4- Micro SD will not work in case of SIM changing. Micro SD will be linked to the SIM by storing the SIM IMSI number (International mobile Subscriber Identity).

More information about how the micro SD is linked to other components will be present in the system sequences diagram.

System Architecture:

In the coming section, we will explain the system architecture overview in simple steps using a basic scenario.

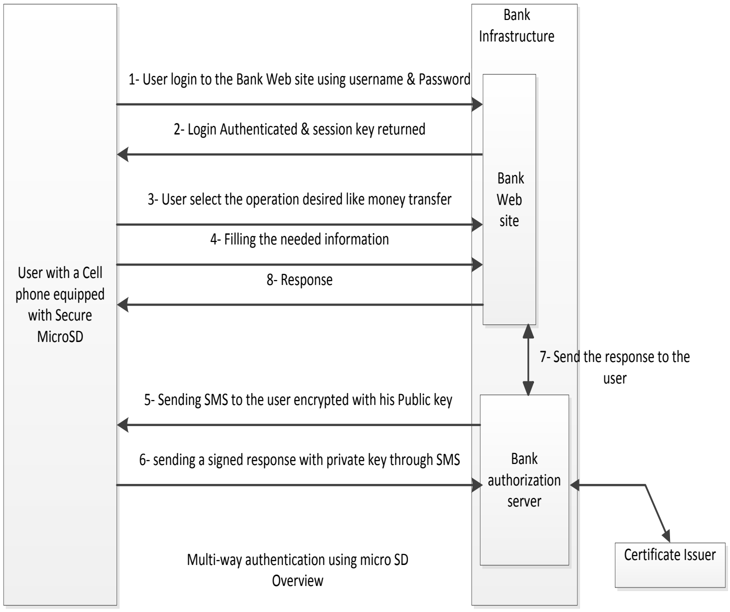

User with a Cell phone equipped with Secure MicroSD
Bank Infrastructure
Bank Web site
Bank authorization server
Certificate Issuer
1- User login to the Bank Web site using username & Password
2- Login Authenticated & session key returned
3- User select the operation desired like money transfer
4- Filling the needed information
8- Response
7- Send the response to the user
5- Sending SMS to the user encrypted with his Public key
6- sending a signed response with private key through SMS
Multi-way authentication using micro SD Overview

Fig.5 Basic authentication scenario using secure micro-SD

already configured username/ password.

Step 2: Bank will allow user to login in, as a result of success comparison between entered username & password with the saved one.

Step 3: User choose the operation that he needs to perform.

Step 4: User filling the needed information.
Step 5: Bank will use the user phone number to send an encrypted SMS to the user asking for his sign. The SMS will be encrypted by user public key. As soon as user's phone receives the SMS, an application will be activated as the SMS has a special short code. The application will decrypt the SMS using user's private key. Private Key information stored in the phone's secure micro SD.
Step 6: User will have a screen on his mobile with the following information:

a. SMS content.
b. One button to sign the SMS.
c. One button to reject it.

Based on user choice a signed SMS will be send back to the bank.
Step 7: Bank will decrypt the message and comparing between the message and its hashing value.
Step 8: Bank will send the user a response back through Web interface or SMS.

System Class Diagram:

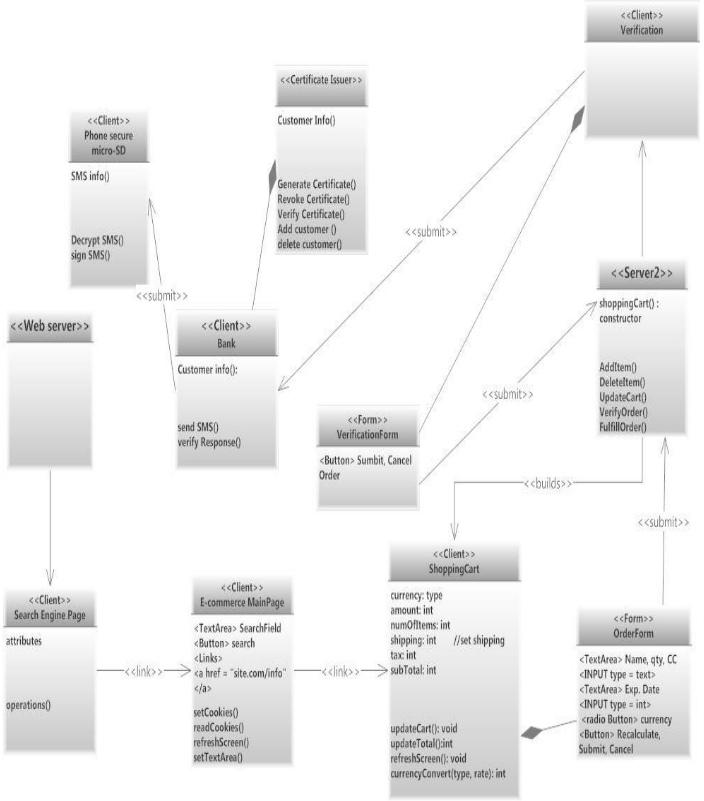


Fig.6 Multi-interface web transaction using secure Micro SD

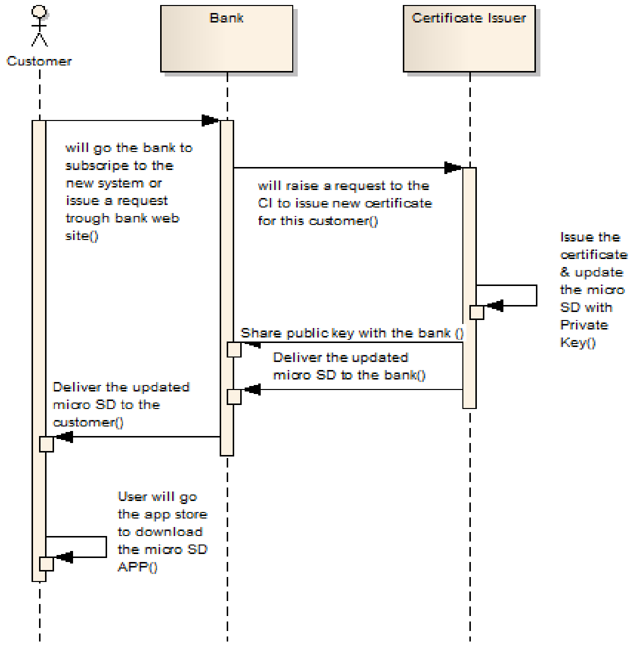
As shown in Fig 6, we have added two new relations to the normal web transaction class diagrams to support multi interface authentication[10].

1- Certificate issuer is newly presented to the class diagrams as he will be responsible to issue the certificate & revoke it.
2- Mobile secure micro-SD: it will be responsible to decrypt the message that will be received from the band and sign the response back to the bank.

By adding these two components to the class diagrams we have reached the multi-interface authentication with minimum impact on user& cost.

System sequences diagram:

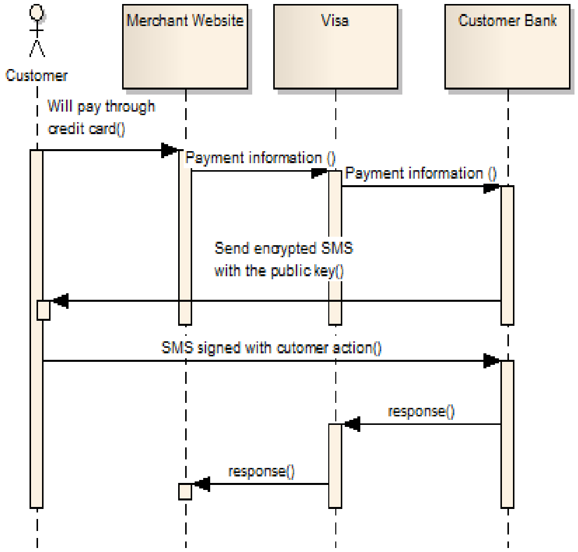
- Customer uses the system for first time: in order to build the system and operate correctly, we need to add a relation between the customer's bank and a new component that will presented to the transaction life cycle which is certificate issuer (CI). Each customer will have his own certificate (1024 or 2048 bits). Public key will be shared with the bank and the private key will be stored in the micro SD [16]. Then, micro SD will be delivered to the customer. During the customer [13] registration in the bank security links will be created .
  a. Micro SD will be configured to have customer IMSI (International mobile subscriber identity) number.
  b. Micro SD will be configured to have customer IMEI (International Mobile Station Equipment Identity) number.



Customer
Bank
Certificate Issuer
will go the bank to subscripe to the new system or issue a request trough bank web site()
will raise a request to the CI to issue new certificate for this customer()
Issue the certificate & update the micro SD with Private Key()
Share public key with the bank ()
Deliver the updated micro SD to the bank()
Deliver the updated micro SD to the customer()
User will go the app store to download the micro SD APP()

Fig.7 customer uses the system for first time (SET)

- E-payment sınging: the proposed system can be used to certify an E-payment. Customer can enter any website and select the desired goods or services. Then, customer will choose the credit card as a payment method. Payment information will be submitted to VISA which will contact the customer bank. Customer's bank will send an encrypted SMS to the customer mobile phone. Then Customer will reply back in signed SMS. In this way, we are sure that the customer is he /she claim to be.



Key management: customers will receive their micro SD already loaded with the customer private key. It's recommended to change the customer private key every 3 months. It is possible to change the customer private key in a secure way remotely. We can use OTA (over the Air) technology to send the new private key to the customer in a secure way. OTA is a telecom standard to send secure/encrypted SMS to mobile device to execute commands. Each SIM has its own key that will be used to encrypt the SMS before sending. This key cannot be repeated between customers.

V. HOW THE PROBLEM GET SOLVED

The issue with multi-interface authentication is solved by adding a new component to the transaction life cycle like [14 , 15]:

1- Certificate issuer: it will be responsible for issuing a new certificate per customer.

2- Micro SD: it's a secure component will be used from inside the mobile phone to encrypt /decrypt and signing SMS

Step 1: Customer logged into the seller web site to select goods or services. Web site will show the total cost of the goods/ services with taxes and shipping cost.

Step 2: The system asks the customer to choose payment method.

Step 3: Customer will fill payment info to pay through credit card.

Step 4: Merchant agent will raise a request to merchant bank for payment and customer authorization.

Step 5: Merchant bank will contact customer bank for payment authorization.

Step 6: Customer bank will send an encrypted SMS to get customer signing

Step 7: Customer will reply will sign SMS [16].

Step 8: Customer bank will reply back to the merchant bank with the customer response based on the signed SMS verification.

Step 9: Merchant bank will reply back to the merchant agent based on customer's bank response.

Step10: Customer will get a response from the Merchant agent which is the same response he gave already when signing the SMS.

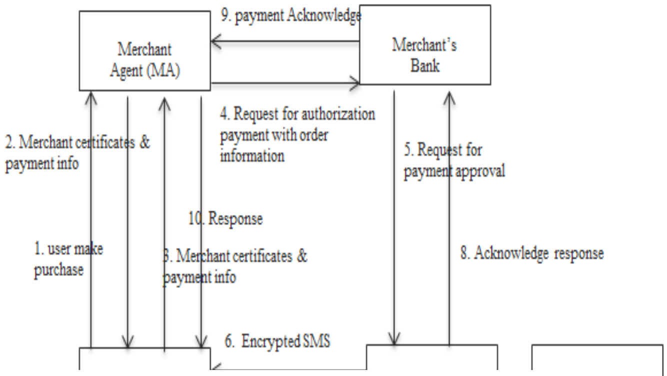

Merchant Agent (MA)
Merchant's Bank
9. payment Acknowledge
4. Request for authorization payment with order information
2. Merchant certificates & payment info
5. Request for payment approval
10. Response
1. user make purchase
3. Merchant certificates & payment info
8. Acknowledge response
6. Encrypted SMS

Fig.10 transaction flow in multi interface authentication transaction

VI. SYSTEM ANALYSIS AGAINST INTERNET THREATS

The proposed system can handle many internet threats like phishing, loss of cell phone etc. we will discuss the resistance of the proposed system under many internet threats [9]:

Security against Phishing: phishing fraud is very popular technique to capture important data like passwords and credit card. Phishers can use the hacked information to gain access to customer data or transfer money. Phishing is generally performed using email or instant message or false web site. The proposed system is against phishing attacks. Even if the phishers can get the username / password, he cannot do any operation. Multi-way authentication is protecting the system, so phisher will need to sign the verification SMS that will be sent from the bank.

Virus attack on phone: infecting phone with viruses will not affect the micro SD. Micro SD is doing all encryption /decryption and signing operation in a separated space which is not accessible to any application installed on the phone. It is impossible for any application to get the private key. Infected phone can still be used to sign SMS without known issue.

Cell phone loss: no one can access the security module on the micro SD except its owner, because you will need to enter the pin code to activate the application. Moreover, we can use the OTA technology to send an encrypted SMS to the phone to change the private key to mismatch the public key, so the security module will be disabled and no one can sign any requests.

SMS is delayed or destroyed: it is highly recommend having a well-equipped network, if you are going to rely on it to verify the user identity. This issue addressed by placing a timer on the bank and on the phone to put a time frame regarding when you should get a SMS delivery receipt.

If someone changed the phone: if someone tried to move the micro SD from one phone to another, he can use it as storage media not as a security module. The micro SD and phone will be linked together through the IMEI number.

VII. CONCLUSION& FUTURE WORK

The proposed system has achieved many benefits compared to the available options on the market.

	PIN/ password List	One Time password OTP Token	Mobile SMS-OTP	PKI smart card + reader		Mobile secure Micro SD
Security	🙁	😐	😐	🙂		🙂
Easy to use	🙂	🙁	🙂	🙁		🙂
Mobility	😐	🙁	🙂	🙁		🙂
Low Usage Cost	😐	😐	🙂	😐		🙂
Low Maintenance	😐	😐	🙂	🙂		🙂
Legal Qualified signing	🙁	🙁	🙁	🙂		🙂

🙁 Not compliant 😐 : Partially compliant 🙂 : Compliant

We have introduced legally secured payment transaction system based on multi-interface authentication instead of normal interface which is based on username /password.

Phone will be your token, as a result of wide usage. Using an already existing device will decrease the overall system cost. User familiarity with mobile phone interface will not reduce the user experience. Adding a secure micro-SD to the phone can help to do cryptographic operation in isolated space, which will grant the user privacy. However the software can be easily downloaded from the phone app store

We did affect the ease of use or user journey, as all encryption/decryption and signing is hidden for the user. Customer identity is confirmed and protect in a way that cannot be deniable. We have reached a level of encryption which is legally accepted.

For future work, same system needs to be implemented with NFC Micro SD or Bluetooth that will eliminate the need for SMS which will make the system easier to use.

References:

[1] Global mobile statistics 2014 Part A: Mobile http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats/a#subscribers

[2] Mobile usage of e-commerce http://www.screenpages.com/about/articles/mobile-research

[3] Abdurrahman, U.A. ; Dept. of Comput. Eng., Mevlana (Rumi) Univ., Konya, Turkey ; Kaiiali, M. ; Muhammad, J. A new mobile-based multi-factor authentication scheme using pre-shared number, GPS location and time stamp, "Electronics, Computer and Computation (ICECCO)", 2013

[4] Castro,P.C.IBM Research Division, Thomas J. Watson Research Center, YorktownHeights, NY, USA Ligman, J.W. ; Pistoia, M. ; Ponzo, J. ; Thomas, G.S. Runtime adaptive multi-factor authentication for mobile devices "IBM Journal of Research and Development ". Nov.-Dec. 2013

[5] Internet Social Networking Risks http://www.fbi.gov/about-us/investigate/counterintelligence/internet-social-networking-risks

[6] European Commission, "Green Paper: Towards an integrated European market for card, internet and mobile payments." 2012.

[7] XuanZhang Sch. of Software, Yunnan Univ., Kunming, "ChinaModified SET protocol for mobile payment" Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference.

[8] Salajegheh, Mastooreh Computer Science Department University of Virginia, USA, "Unleashing the Wild Card for mobile payment" Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on

[9] Roland, Michael NFC Research Lab Hagenberg University of Applied Sciences Upper Austria,"Applying relay attacks to Google Wallet", Near Field Communication (NFC), 2013 5th International Workshop.

[10] Wang Jie Sch. of Comput. & Commun. Eng., Zhengzhou Univ. of Light Ind., Zhengzhou, China," Research and Design of Mobile Payment Security Middleware" Industrial Control and Electronics Engineering (ICICEE), 2012 International Conferenc

[11] C. TOMA, "M-Payments Issues and Concepts." Informatica Economică, vol. 16, no. 3, 2012,pp. 117-123.

[12] J. Venkatesh and D.S. Kumar, "Evaluation of Mobile Payment System and It's Service Providers.", International Journal of Multidisciplinary Research, vol. 2, no. 4, 2012,pp. 118-123.

[13] V.K. Raina, U.S. Pandey, and M. Makkad, "A User Friendly Transaction Model of Mobile Payment with reference to Mobile Banking in India". International Journal of Information Technology, vol. 18, no. 2, 2012,

[14] Sekhar, V.C. Dhirubhai Ambani Inst. of Inf. & Commun. Technol., Gandhinagar, India,"Secure Lightweight mobile payment protocol using symmetric key techniques"Computer Communication and Informatics (ICCCI), 2012 International Conference.

[15] Wan Zhongbao ; Inf. Security Center, East China Jiao Tong Univ., Nanchang, China ; Wang Qing."Secure mobile payment based on Super SET protocol",Advanced Computer Control (ICACC), 2010 2nd International Conference.

[16] Kitahara, M. ; Dept. of Inf., Kyushu Univ., Fukuoka, Japan ; Nishide, T. ; Sakurai, K."A Method for Embedding Secret Key Information in RSA Public Key and Its Application"Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference.