SESSION

INTERNET OF THINGS

Chair(s)

TBA

IoT: A Frost Penetration Sensor Network

D. Sawka[1], M. Frost[1], D.P.U. Tran[1], S. Virk[1], J. Blatz[2] and R.D. McLeod[1]
[1] Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Canada
[2]Department of Civil Engineering, University of Manitoba, Winnipeg, Canada
Robert.mcleod@umanitoba.ca

Abstract - *Frost Penetration Monitoring is a very practical IoT project and demonstrates a means to pre-emptively predict potential service interruptions due to water supply line freezing below grade. Such interruptions occurred at over 2600 homes in Winnipeg, Manitoba in winter 2014 as well as Eastern Canada in 2015 and 2016, and can cost municipalities millions of dollars per season. Temperature sensors were developed to be installed at grade, and down to a depth of 8 ft. at 2 ft. intervals to monitor temperatures and hence infer frost penetration. Data is collected remotely and transmitted over WiFi to a cloud data service. Data integration includes weather data from Environment Canada, estimated frost penetration from the City of Winnipeg, as well as the sensor data collected directly. This types of infrastructure monitoring is an ideal IoT endeavor, integrating sensors, cloud data collection, and analysis. In addition to real time value, the data collected is essential to develop heat transfer models (analytical and numerical) that would allow better prediction of frost penetration during annual cycles. These techniques would be useful in other northern climate cities as well as in higher latitude regions to better understand changes in permafrost due to climate change*.

Keywords: Remote temperature sensing, cloud services, infrastructure monitoring.

1 Introduction

In the Department of Electrical and Computer Engineering at the University of Manitoba we have been developing applications that center on sensing and control within an IoT framework. These applications typically are senior year Capstone design projects or Graduate projects. This paper outlines a Capstone design of a distributed frost depth monitoring system as an early warning system to mitigate against freezing water pipes. Frost penetration has been of interest in both academia as well as in practice for a considerable period of time [1][2]. Soil composition, moisture content, and environmental conditions have been used to build models to estimate frost penetration. In addition, surface type is also an important consideration in particular for highways and airfields [3]. Complicating matters in terms of building predictive models also include non-uniform soil (multilayer soil) [4]. More recently, numerical modeling methods have become standard often augmented or supported by field measurement. In many cases, the principle interest is in predicting frost heaves, whereas the principle interest here is in simply knowing if the frost has penetrated sufficiently to cause the freezing of a water pipe.

An increasingly attractive alternative to sophisticated models and methods is to measure ground temperatures directly or indirectly. This paper describes methods and techniques well suited to this type of infrastructure monitoring.

2 Sensors for Frost Penetration

This project was created in response to the widespread freezing of municipal water supply lines in Winnipeg two winters ago [5]. The winter of 2014 left thousands of residences with frozen pipes and without water [6]. Little was known about properties that were at risk and residences in which frozen pipes occurred were only discovered when they were reported to the city. Having proper and thorough information as to the actual temperature of the ground all over the city could greatly aid in the prediction of frozen pipes each year. This would allow preventative measures to be put in place for specific areas of Winnipeg and would also help to pinpoint at-risk locations for freezing. Although it was water pipe freezing which prompted this project's creation, temperature profile monitoring can be used for a variety of applications. Moreover, the backend and microcontroller system can be deployed for basically any sensing application.

2.1 System Overview

As shown in Fig. 1, a complete sensor controller unit for an individual location consists of a solar panel, the temperature sensor array or pole, and the sensor controller box. Four such units are installed at three locations in Winnipeg. All three locations have setups identical to the sensor on the right as shown in Fig. 1 (middle of the yard). Fig. 1 depicts a location that has a second sensor unit installed with the temperature sensor array within one meter from the roadway. This was done to compare the temperature profile next to the road versus the profile in the middle of the yard. The choice to install this second unit near the road was a consequence of the conjecture that the uncovered road would result in the frost penetrating deeper than in the middle of the yard where the snow can act as an insulator. The results and comparisons are presented subsequently.

The temperature monitoring system was divided into four sections:

- Power Management – The power generation, charging, and regulation for the entire unit.
- Sensors – The attributes of the sensors and the construction of the sensor arrays or poles.
- Communications – The wireless communication and cloud storage for easy access to the data.
- Microcontroller System – The design and construction of the microcontroller with interfaces to the previous three sections as well as the software that it runs.

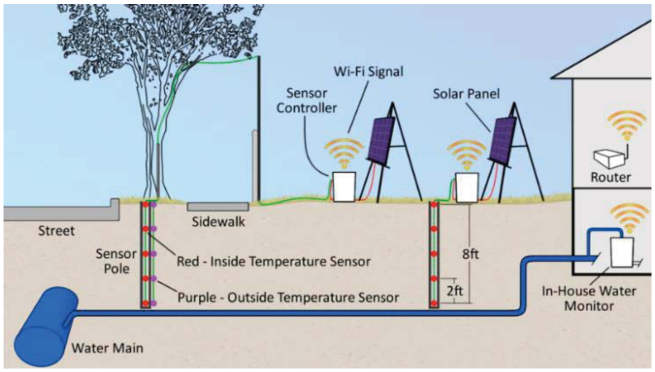


Figure 1: System Diagram

To maintain power for each temperature monitoring unit, a solar panel is connected the Sensor Controller. Inside this box, power from the solar panel is used to charge a battery which is responsible for supplying the power to the entire system. This system also controls the power to the Microcontroller System. The Microcontroller System is powered up every six hours and cuts power when the Microcontroller System signals that it has completed its task.

To monitor the temperature of the ground, temperature sensor array extends over eight feet into the ground. The top of each array lies at ground level and the bottom at the eight foot mark. The sensor arrays are made of PVC conduit or poles and have five temperature sensors placed inside them. The first temperature sensor is placed at ground level, then in two foot increments along the pole, stopping at the eight foot mark. In order to protect the sensors from possible corrosion and water damage, each pole has been completely waterproofed. They are just over eight feet in height to accommodate for waterproofing at the top and bottom. The poles were installed in late November before the ground froze. Fig. 1 also shows one temperature sensor pole that has temperature sensors on the outside of the pole as well as the inside. This pole has ten sensors, one on the inside of the pole and one on the outside for each depth increment. This was done as there were concerns from outside parties that the isolative properties of the temperature pole would affect the results of the measurements. The comparison of the inside and outside temperature sensor results is discussed subsequently.

The data that is recorded from the temperature sensors are uploaded to the cloud service ThingSpeak [7]. This is done via Wi-Fi to the Wi-Fi network of the house to which the unit is installed.

To monitor the temperature from the sensors, a custom assembled controller was built that records the temperature from the sensors and uploads this data using the Wi-Fi module or shield. This system has been designed to use as little power as possible in order to maintain battery power throughout the entire winter, even when the sun is at its weakest. When deployed in the configuration of five temperature sensors read every six hours, the unit can record over 24 days of sensor readings before the oldest data is overwritten. This controller has been specifically designed to utilize the sensor's unique ability of communicating using only one shared data wire. The implication of this functionality is that any number of temperature sensors up to 102 can be connected to the controller with no modification to the software.

Fig. 2 provides a map of the location of the sensor arrays. Location 1 is as configured in Fig. 1 with locations 2 and 3 each having a single sensor array within the respective property. Ideally, as these types of IoT applications are further developed, the sensors would be located in areas historically known to be more prone to water line pipe freezing.

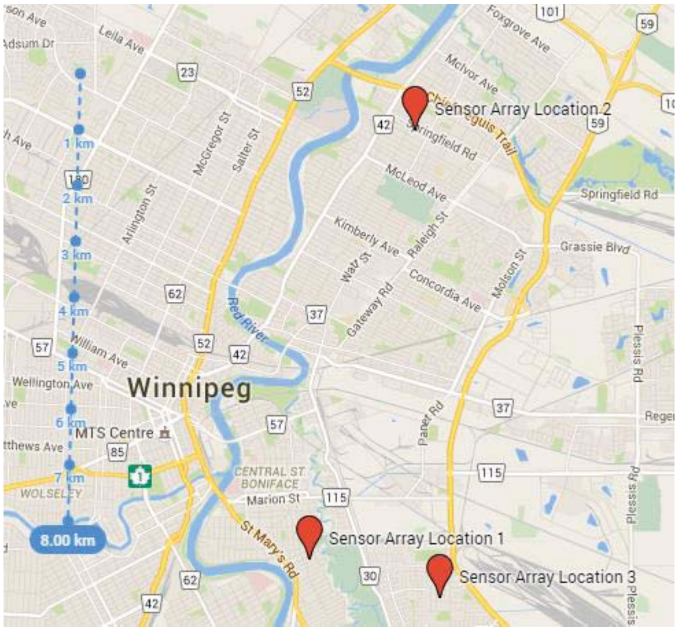


Figure 2: Sensor Array Network

2.2 Subsystem Design

The following sections outline specifications and components used in the system build.

2.2.1 The Temperature Sensors

The temperature sensors play an important role in the design of a temperature profile monitoring system. The overall goal is to keep track of how fast the frost penetrates into the ground. In order to do this, a temperature sensor array with sensors every 2 feet from ground level down to 8 feet underground was constructed. The temperature sensor specifications are:

- One-wire data interfacing to limit the number of wires used to connect the sensors with the control system.
- High temperature measuring accuracy.
- Allowed distance from sensor to controller of at least 9 feet.
- Low power consumption.

The DS18B20 sensor was selected because it satisfies all of the requirements. Operating temperature range is -55°C to +125°C [8]. Temperature measuring accuracy is ±0.5°C from -10°C to +85°C. On top of that, it is compatible with many controllers such as Arduino and Raspberry-Pi. The supply voltage range could be between 3.0 V to 5.5 V. The controller provides a 3.3 V to enable the temperature sensor. This minimizes the power consumption, and gives a consistent and common supply voltage among the Wi-Fi shield, controller, and temperature sensors.

An excellent feature of the DS18B20 sensor is its one-wire digital interfacing data line. This allows multiple sensors to be connected using only one pin at the controller. One line of connections can link up to 127 sensors to an instrument to reduce the amount of wires. For our system design, as the sensor is at least 9 feet away from the controller. A long wire could introduce errors in an analog system. Thus, with the digital interface, the system does not suffer these types of errors.

2.2.2 The Sensor Array

Temperature sensors buried in the ground are vulnerable to environmental effects such as corrosion and physical impacts. The soil could be acidic or alkaline. Soil with different pH levels could be harmful to wire connections and the sensors if exposed for a long period of time. Therefore, the sensors were placed inside a PVC pipe to protect them from such exposure.

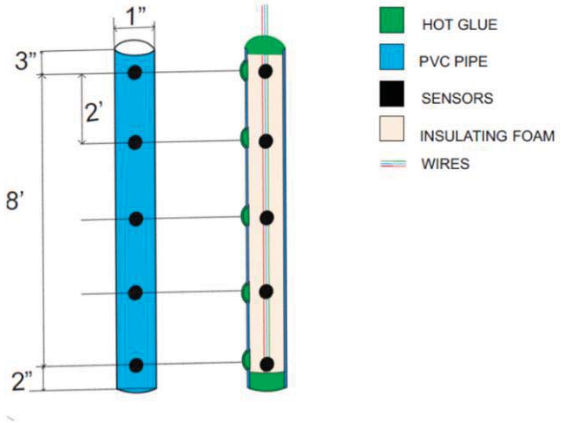


Figure 3: Temperature Sensor Array

Inside a hollow PVC pipe, air could flow up and down. Air circulating leads to unwanted temperature differences between the inside of the pipe and the outside environment. The goal of the project is to measure the temperature of the ground at different depths as precisely as possible. To solve this concern, insulating foam was used to fill the pipe. Insulating foam not only keeps the air from circulating, minimizing the erroneous temperature variation, but also keeps the sensors secured at the right depth.

Fig. 3 illustrates the sensor array pole. A PVC pipe of the length of 8 feet 5 inch is used to hold the sensors. The pipe is filled with insulating foam. After the insulating foam is set, hot glue is used to seal the cavities to keep water from entering the pipe.

2.2.3 The Communication Subsystem

The specifications for the communication subsystem are as follows:

- Transfer data from the microcontroller to a cloud based storage service.
- Perform successful data transfer when connected to a wireless router that supports IEEE's 802.11 b, g or n standards.
- Perform successful data transfer when connected to a wireless router at a distance of approximately 50 meters.
- Reliable data storage.
- Display data for all the sensors in a graphical form.
- The ability to analyze data easily.

The Arduino WiFi shield 101 is used to transfer data from the microcontroller to the cloud storage service used for this project. The Arduino WiFi Shield 101 is a shield designed to connect any Arduino based microcontroller to the internet. It is compliant with the IEEE 802.11 b/g/n standard [9]. The two main reasons for choosing this shield are that this shield had the greatest range of all the WiFi shields available at the time of selecting components when using the IEEE 802.11g standard and it is compliant with the IEEE 802.11n standard to boost range even further with a compatible router. It has a rated range of 70 meters indoors and 250 meters outdoors making it sufficient for this project. The wireless routers providing internet connections to the shields for all the frost measurement locations are a distance well within this range. The third reason to choose this shield was that it has the ability to operate at both 3.3V and 5V. Since the microcontroller can operate at 3.3V, it was ideal to choose a shield that could operate at the same voltage as well. This simplifies the design of the power system and aids in efficient energy use.

The WiFi shield can transfer data to the internet either using the traditional HTTP POST and GET methods or by using any third party library; provided that the library has been downloaded and included in the Arduino software. For this project, the ThingSpeak software library created by the ThingSpeak cloud service to simplify data transfer between its channels and the microcontroller was initially used to transfer data to the cloud service. Eventually, due to an issue encountered during the integration phase of the project, it was decided to use the HTTP GET method to perform the data transfer. Because of this issue the format in which the data is transferred to ThingSpeak is the HTTP GET method, with specific fields as specified by ThingSpeak.

ThingSpeak is the cloud service that is used for this project. It is a free web service that offers storage, display and analysis of data. Although the ThingSpeak service could be used for any type of data, it has been specially designed for sensor data, making it appropriate for our usage. The main reasons for using this service was that ThingSpeak eliminated the need for creating a separate website for displaying the sensor data, displayed the data very nicely in a graphical form, allowed data analysis within its own domain and that it was a free service.

2.2.4 Microcontroller System

The Microcontroller System is a custom assembled unit based on an Atmel ATmega328P-PU microcontroller. It controls the reading of the temperature sensors, the storing of the temperature data, and the transmission of the data to ThingSpeak. The Microcontroller System has been designed to operate using very low power as this was one of the utmost concerns for the finished prototype. The Microcontroller System is powered up by the Power System every six hours. After power up, the Microcontroller System reads the digital temperature sensors and saves the data in its onboard storage. Then the microcontroller connects to ThingSpeak using the Wi-Fi card and the Wi-Fi of the house that it is deployed to. The microcontroller then uploads the temperature sensor data. The system can upload multiple sets of data in the event it was previously unable to connect to the Wi-Fi network or ThingSpeak. The controller will only try to send up to five sets of data in one burst in order to reduce the worst case power usage. After it has uploaded all of the sensor data that is required, it sends a power-down signal to the Power System, which then shuts off the power to the Microcontroller System for another six hours. The cycle then repeats indefinitely.

3 Installation Procedure

The frost penetration measurement system was installed at three locations across the City of Winnipeg (Fig. 2). To install the system, an eight feet deep hole was dug at these locations using a manual auger. The sensor pole that contained the insulated temperature sensors inside it was then placed in the hole. As the hole had a slightly larger diameter than the sensor pole, the gap between the sensor pole and the soil was filled with sand.

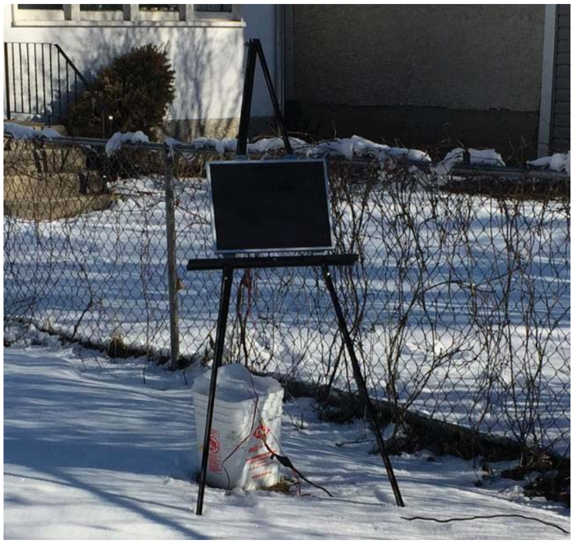


Figure 4: Solar Power Install and Microprocessor System

The battery along with the timer and the voltage converter circuit, the Atmel microcontroller and the Arduino WiFi Shield 101 were placed inside a large plastic container; the system container. The system container was closed with a lid and a small hole was created on one side of the container to connect the wires from the sensor pole to the microcontroller and the wires from the solar panel to the battery. All the wire connections were first sealed and insulated with hot glue and then secured with an electric tape.

The Wi-Fi signal strength was checked at various spots near the sensor pole using a smartphone app and the system container was placed at the spot with the best Wi-Fi signal strength.

A solar panel was placed at an angle on an A-frame artist's easel at a spot near the sensor pole that received direct sunlight for most of the day. The solar panel was placed at an angle as the angled solar panel captures the sun rays reflected from the snow on the ground along with the direct sun rays that it receives, thus increasing its efficiency. The solar panel installed in one yard can be seen in Fig. 4. The system container is just below the easel.

4 Results

The city of Winnipeg collects data from various sources such as water and sewer excavations to monitor the current penetration depth of the frost line during the winter months. The results are updated on a weekly basis and displayed publicly on their website among other related statistics [10]. This is one of the city's current methods for predicting when city pipes will begin freezing. Table 1 summarizes the data as of February 20th and includes a linear interpolation estimation of what our monitoring measures the frost line depth to be each day.

Table 1: Frost Depth as estimated by the city and inferred by measurement

Date	City maximum frost depth	Estimate frost depth in yard	Estimate frost depth near street
January 13	0	0.5 Feet	1.2 Feet
January 15	2 Feet	0.6 Feet	1.4 Feet
January 22	2 Feet	1 Feet	1.7 Feet
January 29	2 Feet	1.2 Feet	1.9 Feet
February 5	2 Feet	1.2 Feet	2 Feet
February 12	3.9 Feet	1.3 Feet	2.1 Feet

The city's estimate is more aggressive than our data suggests, especially compared to the readings taken from the middle of a property. This suggests the city's data is derived mainly from the ground under or in close proximity to city streets. The estimates the city currently acquires are also highly imprecise, with large spans of time where the estimate doesn't change mixed with multi foot changes in the estimation. The readings are also hand acquired, introducing an element of human error into the frost penetration estimations. This imprecision and the lack of more localized frost line depth mapping highlights the importance in upgrading the measurement method.

4.1 Temperature Profiling at One Location

The city regularly clears the snow from the city streets during the winter months, exposing the bare pavement to the frigid winter air. It is expected that the reduced snow cover due to this snow clearing and the higher heat conductivity property of pavement compared to soil will promote faster frost penetration near the roads. To investigate the potential effects of this, two systems were set at different locations on the same property; the "Roadway" monitoring system is 1 meter from the edge of a paved road, while the 'Yard' monitoring system is 15 meters deeper into the property with a large grass-only surrounding. In Fig.5 illustrates the temperature profiles associated with the yard while Fig. 6 illustrates the temperature profile at the Roadway. It is clear from a closer analysis of the data that frost penetration is slightly greater at the Roadway as expected.

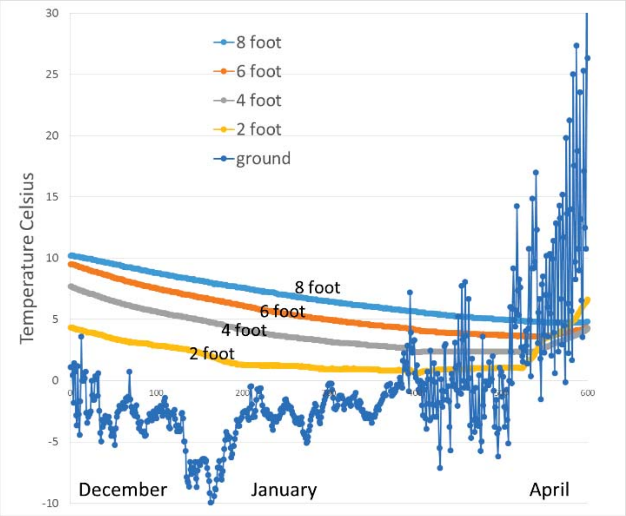


Figure 5 Temperature Profile in Yard

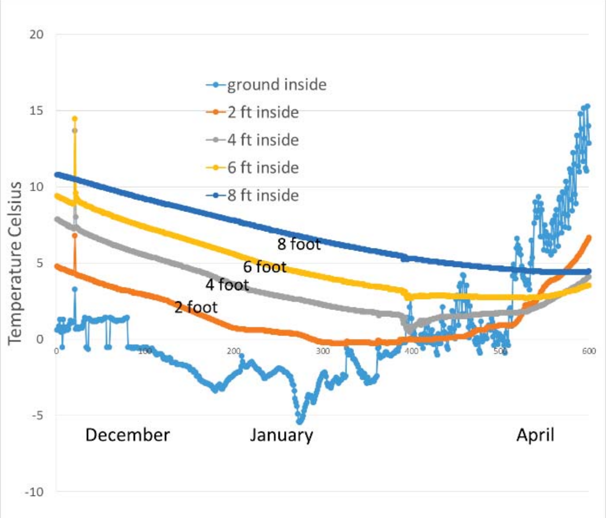


Figure 6: Temperature Profiles at Roadway

The readings at ground level are hard to draw any conclusions from because of their variance. This data suggests the ground level sensor of the Street system is about an inch below the soil level, unlike the ground level sensor of the Yard system which isn't covered by any soil. A direct comparison between the two ground level sets of readings is somewhat unimportant as the interest is really in the sensors below grade. Turning attention to the other eight sets of readings strongly supports the conjecture that the frost penetrates at a more rapid pace nearer to the road and its bare pavement. This is most easily seen from the two foot and eight foot underground sets of readings. When initial readings were recorded from the systems on December 12th the Street and Yard readings were identical, but as the winter has progressed the readings have consistently and gradually diverged. February 19th readings show the two foot underground temperature of Yard is 1.69°C higher than Street's, while the eight foot underground reading of Yard is 0.62°C higher than that of Roadway's. Considering the total changes in temperature at each depth this winter, this is a significant result.

4.2 Temperature Inside Vs. Outside PVC

As mentioned, the temperature profiling system has temperature sensors at five depths: ground level, 2 feet, 4 feet, 6 feet, and 8 feet below ground level. The sensors are secured inside a PVC pipe to keep them level with the desired depth. The pipe is filled with insulating foam then sealed with hot glue to keep water from entering. One of the temperature sensor arrays has a double sensor systems due to concern about temperature differences between the protected sensors inside the pipe versus the outside environment. The pole has two sensors at each level, one is inside the pipe and one is securely taped outside the pipe. Data was collected over a 120 day period, and several graphs were produced to illustrate the impact of the temperature probe positioning or set-up on the sensor readings.

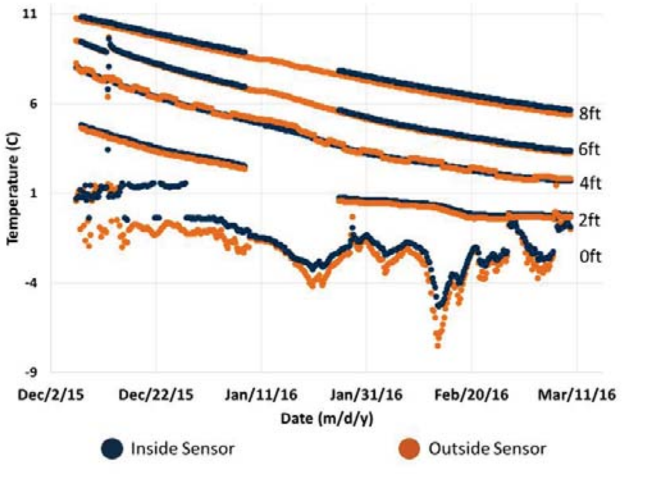


Fig. 7 Temperatures inside and outside of the PVC

Fig. 7 shows the temperature of inside versus outside sensors at a depth of 8 feet. The 10 sensor pole has a gap of lost data for 18 days, between Jan 18th to 26th 2016. Unexpected changes in cloud service properties caused Wi-Fi communication problems. The bug was corrected but it was not possible to recover the data during this period.

Notwithstanding, the collected temperature of the outside sensors show a consistent trend of equal or slightly lower temperature compared to the inside ones. The worst case scenario was at eight feet depth, shown in Fig. 6. However, these differences, although correlated are all within the sensors uncertainty of ± 0.5°C. The result is that one may reasonably infer the below grade temperature using this set-up. Alternatively, it may be very practical to simply install temperature probes using displacement method similar to a hydraulic push apparatus used in taking core samples.

5 Extensions

In addition to monitoring frost penetration for purposes of modeling and as an early warning system, the same techniques may be deployed and serve as a more active control or mitigation system. An automated flushing system is an application using the frost penetration data and/or water temperature measurements at point of water supply ingress. When deemed necessary to prevent pipe freezing, the system periodically flushes water from the house water supply to the drainage system, bypassing the rest of the house plumbing. The flushing system consists of a solenoid, a microcontroller, a WiFi shield and a temperature sensor. The solenoid and the temperature sensor are deployed in the water supply line inside a house as illustrated in Fig. 8. The water control valve can be conveniently located near an existing sink and the temperature sensed at the point of ingress wirelessly.

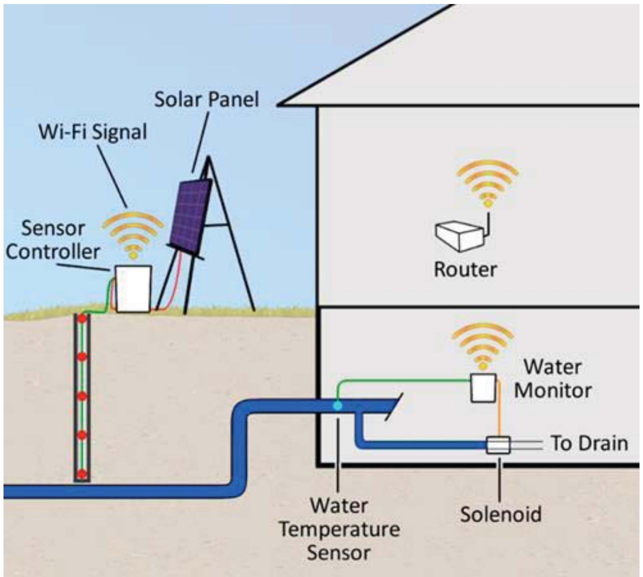


Figure 8: An Automated Flushing System

In an ideal environment a home would be equipped with both a temperature sensor at 8 feet below grade as well as at the point of water ingress. Both would collaborate on controlling the flushing system. For the flushing system, we are interested in the sensor 8 feet underground, because this is often the depth at which the water pipes are located. If the current temperature at that sensor is below the freezing threshold, the flushing controller will flush water for a period of time dependent on how far below the freezing threshold the eight feet underground sensor reads. An inside sensor, which is positioned right at the house water supply ingress, measures the approximate temperature of the water coming into the house. If the incoming water temperature is found to be lower than a threshold, the water will keep running for a period of time. Otherwise, if the inside sensor reads that the incoming water is above the threshold, the water monitor closes the solenoid valve and waits for the next reading from the 8 foot below grade sensor. An advantage of the automated flushing system is that the microcontroller can effectively sample the incoming water temperature by commencing a flow and simultaneously sample the incoming water temperature until it stabilizes.

By flushing the water periodically, the water is prevented from freezing. This is because the movement of the water prevents freezing. A water flow limiting valve may also be installed along the same pipe as the solenoid valve, limiting the flow to that necessary for preventing pipe freezing. This has a twofold benefit. Primarily, waste water is reduced as the water flow can be easily controlled, this reduces the cost for both the owner and the city when they reimburse the costs in the spring. Because the flow is precisely controlled, the volume of water flushed during the winter/spring season can also be tallied and given to the city to more precisely reimburse the property owner. This benefits both the property owner and the city, allowing fairer reimbursement. The current policy is simply based of the history of the customer and the current year change [11], which can be very inaccurate.

6 Summary

The paper presented a distributed temperature sensor array and measurement and analysis of the data to model frost penetration trends in Winnipeg. Data gathered from five temperature sensors placed every two feet forming a vertical temperature sensor array has been used to effectively measure the depth of frost penetration. Data was collected from three different locations within the city using four complete monitoring units.

The temperature sensors are placed inside a PVC pipe that is filled with insulating foam to prevent airflow within the pipe. The sensor data is read by an Atmel microcontroller which uses an Arduino WiFi Shield to transfer the sensor data to the cloud. Each data collecting unit is solely powered by a large battery with solar recharging. ThingSpeak is the cloud service used for storing, displaying and analyzing the sensor data. The data has been collected approximately every six hours since December 06, 2015. With a reading taken every six hours for the five temperature sensors, the EEPROM has the capability to store data readings for approximately 24 days, thus enabling the restoration of data in the case of failure to upload data to the cloud.

The frost penetration depth as of February 17, 2016 was approximately 2.6 feet according to the data collected for this project. On the other hand, the depth of frost penetration as predicted by the empirical model proposed by Soliman [12] should have been 4.7 feet on February 17, 2016 as estimated with data from Environment Canada. Clearly, the data collected for this project does not follow the simplified model proposed by Soliman for frost penetration. Some of the possible reasons for the disagreement between the results obtained and the results expected from the model could be a change in the thermodynamic properties of the soil due to its moisture content, different soil types and different levels of snow coverage on the ground resulting in a change in the insulation of the area of the ground being used for testing. Similar results were recorded for each sensor array location.

To study the impact of the PVC pipe on the sensor readings, five temperature sensors were attached to the outside of the PVC pipe containing the temperature sensors for one of the units. The sensors on the outside were placed at approximately the same level as the sensors inside the pipe. The maximum difference between the readings of the inside sensor and the outside sensor at the time of writing was is 0.6 degrees Celsius, which is below the specified error margin of the sensors (1.0 degrees Celsius). This slight error could be due to other factors like the conductivity of the sand that is used to fill the hole surrounding the sensor pole. Since the error is well within the error margin of the sensors, it is concluded that the PVC pipe has negligible impact on the sensor readings and that the reading from the sensor inside the PVC pipe can be used to keep track of frost penetration.

To compare the difference in frost penetration near the street and at some distance from it, two sensor poles were deployed for one of the property used for measurements. One sensor pole was deployed in the middle of the yard, which was at a significant distance from the street and the other near the street. It was observed that the depth of frost penetration was greater near the street than in the middle of the yard. At eight feet underground, the frost penetration for the street relative to the yard was 116.8% as of February 19, 2016. Hence, it is concluded under road water pipes are more susceptible to freezing than water lines that do not need to traverse a paved roadway. A code suggestion may be to install water pipes deeper than 8 feet for water supplies that cross under a roadway.

The City of Winnipeg estimated the depth of frost penetration to be 3.9 feet whereas temperature profiling estimates the depth of frost penetration to be 2.1 feet near the street and 1.3 feet in the yard as of February 20, 2016. These estimate varies significantly from that of the city. The causes of this variation and the importance of using better data measurement methods were also discussed.

For any future work related to this project, we recommend using a controller with a real time clock to store and upload sensor data. This would ensure that the time stamp of the data is always accurate, even in the case of the data being restored from the EEPROM of the controller.

In summary, paper outlined an ongoing projects oriented to the IoT from a pragmatic perspective and illustrated trends within this consolidating technology. The opportunity for reuse and leveraging existing infrastructure for data collection within the IoT appears unprecedented.

Acknowledgement

We would like to acknowledge the Faculty of Engineering at the University of Manitoba for their financial support for this project.

7 References

[1] Kersten, M. S. "Thermal properties of soils." Highway Research Board Special Report 2, 1952.

[2] Berggren, W. P. "Prediction of temperature-distribution in frozen soils." Eos, Transactions American Geophysical Union 24, no. 3, 71-77, 1943.

[3] Aldrich Jr, H. P., Frost penetration below highway and airfield pavements. Highway Research board bulletin, (135), 1956.

[4] Aldrich Jr, H.P., & Paynter, H.M., Depth of Frost Penetration in Non-uniform Soil (No. CRREL-SR-104). COLD REGIONS RESEARCH AND ENGINEERING LAB HANOVER NH., 1966.

[5] CBC News, "Frozen pipe frustrations boil over with 1,289 still on list", Apr 15, 2014. [Online]. Available: http://www.cbc.ca/news/canada/manitoba/frozen-pipe-frustrations-boil-over-with-1-289-still-on-list-1.2611443 [Accessed: 22- Feb- 2016].

[6] CBC News, "EPC wants pipe-thawing rebates backdated to Dec. 1, 2013", Mar 25, 2014. [Online]. Available: http://www.cbc.ca/news/canada/manitoba/epc-wants-pipe-thawing-rebates-backdated-to-dec-1-2013-1.2585603 [Accessed: 5- Feb- 2016].

[7] [Online]. https://thingspeak.com/[Accessed: 22- Feb- 2016].

[8] Maxim Integrated, "Programmable Resolution 1-Wire Digital Thermometer", 2015. [Online]. Available: https://datasheets.maximintegrated.com/en/ds/DS18B20.pdf [Accessed: Oct- 28- 2015].

[9] Arduino, "Arduino WiFi Shield 101", 2015. [Online]. Available: https://www.arduino.cc/en/Main/ArduinoWiFiShield101 [Accessed: Nov- 20- 2015].

[10] City of Winnipeg, "Cold weather impact on water pipes", February 19, 2016. [Online]. Available: http://winnipeg.ca/waterandwaste/water/frozenPipes/coldweatherimpact.stm. [Accessed: Feb- 22- 2016].

[11] City of Winnipeg, "Billing related to frozen Pipes", December 30, 2015. [Online]. Available: http://winnipeg.ca/waterandwaste/billing/billingRelatedToFrozenPipes.stm. [Accessed: 26- Feb- 2016].

[12] Soliman H., "A Simplified Model to Predict Frost Penetration for Manitoba Soils" 2008 Annual Conference of the Transportation Association of Canada. [Online]. Available: http://conf.tacatc.ca/english/resourcecentre/readingroom/conference/conf2008/docs/x1/soliman.pdf. [Accessed: 18- Feb- 2016].

Cryptography of Things

Cryptography Designed for Low Power, Low Maintenance Nodes in the Internet of Things

Gideon Samid
Department of Electrical Engineering and Computer Science
Case Western Reserve University, Cleveland, Ohio
BitMint, LLC
Gideon@BitMint.com

Abstract Proposing a cryptographic premise where intensive computation is avoided, and security is achieved via non-complex processing of at-will size keys. Memory is cheap, power is expensive throughout the Internet of Things; maintenance may be an issue, and the risk is not less than for the Internet of People. Cryptography must adjust itself: first in its approach and philosophy, then in practical ciphers. The proposed philosophy is to increase the role of randomness, and to build ciphers that can handle any size key without choking on computation. Orthodox cryptography seeks to create a thorough mix between key bits and message bits, resulting in heavy-duty computation. We propose simple, fast ciphers that allow their user to adjust the security of the ciphertext by determining how much randomness to use. We present "Walk in the Park" cipher where the "walk" may be described through the series of visited spots (the plaintext), or, equivalently through a list of the traversed walkways (ciphertext), the "park" being the key, the size of which determines security, but does not affect nominal computation load. We describe a use scenario for the proposed cipher: a drone taking videos of variable sensitivity and hence variable required security – handled by the size of the "park".

Keywords—low-power encryption, randomness, Trans-Vernam Cipher, User-Controlled Security.

I. Introduction

The Internet of Things is comprised of a large number of nodes where mainstay cryptography is overly stressed. Computational power may be limited, data speeds are too demanding, maintenance of remote nodes is a major constraint. We propose a conceptual solution based on the idea that algorithmic complexity may be replaced by large size keys.

We discuss (i) the application environment, and (ii) the principles of the proposed solutions.

A. Application Environment

Our human environment is about to be inundated with sensory nodes, transducers, controllers, invading our living space as they monitor, report, guide and control our modern living space. Swarms of camera fitted drones, and "wall climbers" will watch over us, and help us get out of trouble, as well as exploit opportunities. The communication framework for all this new crop of communicating devices is none other than the Internet of Things. And hence, much as the Internet of People exposed human participants to malware, fraud and abuse, so it will surely happen for the IOT. And much as encryption is the only principled and effective defense for human security, so is the case for security of things.

The IOT connects remote network nodes, which are not easy to access in a physical way. Some may be exposed to the sun, and may generate modest amount of energy on their own, others are not so exposed, and need to be fitted with batteries, which are hard to replace. At either case, there is a strong limit on processing power. Some IOT nodes are tiny flying devices, they cannot 'carry' heavy batteries.

The IOT poses the challenge of saving on computational energy. Alas, modern cryptography is based on algorithmic complexity, which is effected through computational complexity, which in turn drains the device battery. So we need to rethink cryptography with an eye towards computational economy.

Speed: some IOT applications may involve high-speed communication. High speed, high-resolution cameras fitted on flying drones may be required to transmit to an operational center, to serve an important rescue operation, or other social task. Similarly, an isolated device somewhere may be activated with a large stream of commands, most of them should be further transferred to devices down the road. All in all, the IOT may need to accommodate high volume, high speed information exchange. The existing popular ciphers slow down that flow rate, and are not friendly to this requirement. Admittedly, stream ciphers, which are very popular with military applications, are much faster than the standard block ciphers, but they suffer from the vulnerability of synchronization. Losing a couple of bits will create havoc in the entire stream in the downstream.

Maintenance: Quite a few IOT nodes will be placed in hard to access locations, and no physical maintenance will be feasible. Hence the use of any specific cipher, which at any moment may be mathematically breached, is a risky practice. This applies to all algorithmic complexity ciphers. As Prof. Nigel Smith articulates in his book "Cryptography (an Introduction)": "At some point in the future we should expect our system to become broken, either through an improvement in computing power or an algorithmic breakthrough." Normally the IOT gravitates towards very few ciphers considered 'secure'. If one of them is suddenly breached (e.g. GSM communication cipher), then all the IOT nodes which rely on it, have lost their security, and physical attention is not practical.

Magnetic Vulnerability: Many IOT nodes are placed in very harsh environment, and are subject to lightening violence, as well as man made electromagnetic impacts. Software based cipher may be at greater risk.

In summary, IOT nodes are vulnerable both to malicious attack, and to environmental punishment. These vulnerabilities may be remedied to a large extent if we come up with a new cryptographic approach: Cryptography Of Things (COT).

B. Principles of the Proposed Solution

Modern cryptography erects security around data using two parameters: (i) algorithmic complexity, and (ii) randomness. It's generally believed that the more complex an algorithm the more secure the ciphertext, and also the more randomness that is being used (the larger the key), the more secure the ciphertext. Randomness is in a way dull, and of no much interest mathematically (except of course with respect to its definition and to metrics of quality). By contrast, algorithmic complexity is an exciting math dilemma. Academic cryptographers are attracted to this challenge and develop new and newer complex algorithms. Unfortunately in today's state of affairs, we only manage to compare complexities one to the other, not to ascertain their level in an objective mathematical way. And even if it turns out that $P \neq NP$ as most complexity researchers believe, in cryptography complexity is used in combination with randomness, hence one is using a random key selected from a large key space. What is hard to know is how many specific keys when applied with specific plaintexts, offer some mathematical vulnerability, leading to effective extraction of the message. In other words, the de facto complexity, or security of algorithms cannot be ascertained. Worried about this, we come up with increasingly complex algorithms, which require more and more computational effort. They in turn require more and more power -- which many IOT nodes simply don't have.

Randomness, on the other hand is passive memory, and even the smallest and most unsophisticated devices can be fitted with gigabytes of memory, serving as key. These realities lead one to aim to develop cryptography where the role of reliable, passive, manageable, secure randomness is enhanced, while the role of doubtful complex algorithms that are power hogs, is decreased.

This thinking brings to mind the famous Vernam cipher: the algorithm could not have been simpler, and the key could easily be as large as hundreds of gigabytes -- memory is both cheap and light. It may be stored without requiring power. Too bad that Vernam is so impractical to use. Yet, can we re-analyze Vernam as a source of inspiration for security through more randomness and less algorithmic complexity?

Let's envision a Vernam Inspired Cipher (VIC) where at any stage the user can 'throw in a few more key bits' and by that achieve a large increase of cryptanalytic burden, together with a modest increase of nominal processing burden (encryption, and decryption). Let us further demand from the VIC the Vernam property of achieving mathematical secrecy at the minimum key size required by Shannon's proof of perfect secrecy.

To better analyze this vision let's regard any cryptographic key, k, as the natural number represented by binary interpretation of its bit sequence. Accordingly, the Vernam key space associated with n-bits long messages, will be: $1,2,....2^{n+1}-1$ corresponding to $\{00....0\}_n$ to $\{11....1\}_n$. We may further agree that any natural number $N > 2^{n+1}-1$ will be hashed to an n-bits size string. Once we agree on the hashing procedure we have managed to recast Vernam cipher as a cipher that accepts any positive integer as a key, with which to encrypt any message m comprised of n bits to a corresponding ciphertext. We regard this as natural number key representation (NNKR).

We can similarly recast any cipher according to NNKR. We consider a cipher for which the series $n_1, n_2,.....n_{max}$ represents the allowable bit counts for the keys. E.g for DES the series has one member $n_1=n_{max}=56$; for AES the series contains three members: $n_1=128$, $n_2=196$, $n_3=n_{max}=256$. For a cipher where the key is a prime number then the series is the series of primes. For ciphers defined over every bit string of length n_{max} all the natural numbers from 0 to $2^{n+1}-1$ qualify as a n_{max} key. Larger keys will be hashed to a n_{max} bits long hash. For ciphers where the series $n_1, n_2, ... n_{max}$ represents discrete possible keys, we may agree to hash any natural number to highest member of the list $n_1, n_2,....$ which is lower than that natural number. For all natural numbers smaller than n_1, we will "hash" them to the null key ($|k|=0$), and we may formally agree that the case of k=NULL is the case of no encryption (the ciphertext is simply the plaintext).

With the above definition we have recast all ciphers as accepting every natural number as a key.

The basic idea of security is that the larger the key, the better the security. In other words, we "buy" security, and "pay" for it with a choice of a random number -- the larger the number, the higher the price. Let $s_i(k)$ be the security achieved by a user of cipher i, "investing" key k. The metric s, will reflect the average computational effort required of the cryptanalyst for extracting the message m from a captured ciphertext c, computed over the distribution of $m \in M$, where M is the message space from which m is selected. Let $p_i(k)$ be the average combined processing effort (encryption plus decryption) required of a user of cipher i, while using key, k, over the distribution of message $m \in M$.

For any cipher i, using a natural number k as key, we may define the utility of the cipher at this point as the ratio between the cryptanalytic effort and the nominal processing effort:

$$(1)....... U_i(k) = s_i(k)/p_i(k)$$

We can now define a Vernam Inspired Cipher as one where over some range of natural number k (k_1.....k_2) as key, the utility of the cipher will be somewhat stable:

$$(2)......\ U_{K1}, U_{K1+1}, ...\ U_{K2} \sim U$$

In that case a user encrypting with k_1 will be able to increase the security he builds around the data, while still using the same cipher, by simply ratcheting up the key from k_1 to k_2. She will then -- again, using the same cipher -- increase its associated security from $s(k_1)$ to the higher value of $s(k_2)$

$$(3)...........\ S(K_2) = S(K_1) + \Sigma\ (U(K+1) * P(K+1) - U(K) * P(K)\ )\ \text{FOR } K=K_1 \text{ TO } K=K_2$$

which is reduced to:

$$(4).........\ S(K_2) = S(K_1) + U\ \Sigma\ (P(K+1) - P(K)\ )\ \text{FOR } K=K_1 \text{ TO } K=K_2$$

We may now add the Vernam limit for a large enough key. Recasting cryptographic keys as natural number leads to redefinition of the key space, K, as a subset of the natural numbers from 1 (or formally from zero) to the highest natural number to be considered as a key, k_{max}:

$$(5).......\ |K| \leq K_{MAX}$$

And hence, for messages comprised of n bits, a key max of value 2^n ($k_{max} = 2^n$) will allow for a cipher where the user could simply ratchet up the integer value used as key, $k' < 2^n$, to the point of achieving mathematical security. We can define a special case of a Vernam Inspired Cipher, as a **Trans Vernam Cipher (TVC),** being a cipher where increase in the integer value used as key will eventually reach "Vernam Security Levels", or say, Shannon's security, for n-bits long messages:

$$(6).......\ S_{MAX} = S(K_{MAX} = 2^N) = S(K') + \Sigma\ U(K) * (P(K+1) - P(K)\ )\ \text{FOR } K=K' \text{ TO } K=K_{MAX}$$

Existence: It's readily clear that DES, AES and their like will not qualify as Vernam Inspired Ciphers. For DES:

$$(7).......\ S(K < 2^{56}) = 0$$
$$S(K > 2^{56}) = S(K=2^{56})$$

For AES:

$$(8).......\ S(K < 2^{128}) = 0$$
$$S(2^{128} < K < 2^{192}) = S(K=2^{128})$$
$$S(2^{192} < K < 2^{256}) = S(K=2^{192})$$
$$S(\ K > 2^{256}) = S(K=2^{256})$$

A positive example for existence can be shown on the basis of any cipher where the ciphertext, c, is larger than the message, m: $|c| > |m|$. We shall designate such a cipher as EXP (expansion cipher). Let $|k_{exp}|$ be the bit size of the key used by the EXP cipher. Let k be any positive integer larger than $2^{|k}\text{exp}^{|+1}$-1. We may hash k è k_{exp} then use it to iteratively encrypt k to a larger k'> k, k" > k'... until some $k^{(t)}$ is as large as the required Vernam key. ($k^{(t)} \geq 2^n$) for n-bits long messages. . The larger the key k (closer to 2^n) the closer it is to a perfect Vernam, and the greater its security measured. Even this simple procedure highlights the advantage of 'any size key': the cryptanalyst cannot bound his task, as she does with fixed size keys. This procedure may be used with keys that are a tiny fraction of the Vernam size, or very close to Vernam. At the latter case there will remain non-reducible equivocation regardless of the computational ability of the cryptanalyst.

The security of such a procedure depends of course on the nature of the EXP cipher. But it illustrates the underlying notion of a cipher framework where a user can "buy" more security, by simply investing in more randomness, and thereby buying whichever security one desires, up to perfect mathematical (Shannon) security.

II. "WALK-IN-THE-PARK" CIPHER

We present here a Trans-Vernam Cipher (TVC), that runs by the name *Walk-in-the-Park* because both encryption and decryption is taking place by "walking" – charting a path determined by the message, and then describing it through various entities in the "park" where the walk happens. It is based on the idea that a 'walk' can be described either via the places visited, or via the roads taken. One needs the "park" to convert one description to the other.

The cipher is defined as follows:

We employ a four-letter alphabet: X, Y, Z, and W, expressed via 01,10,11,00 respectively. The key is a table (or matrix) of size *m * 2n* bits, which houses some arrangement of the four alphabet letters (m*n letters in total). We regard every letter as a node of a graph, and regard any two horizontally or vertically contiguous letters as connected with an edge. So every letter marked on the graph has between 2 to 4 edges connecting it to other letters on the graph.

We define a path on the graph as a sequence of marked letters such that any two contiguous letters on the graph are connected via an edge.

Informally, the cipher works by mapping the plaintext into a sequence of X,Y,Z, and W; then using this sequence to mark a pathway on the graph. Given the starting point, it is possible to describe the very same graph via denoting the edges traversed by the pathway. Each node, or vertex on the graph has up to four edges; let's mark them Up, Down, Right, Left: U,D,R,L, and assign the bit combinations 01,10,00,11 respectively to them. The translation of the pathway from a sequence of vertices to a sequence of edges amounts to encrypting the plaintext to the ciphertext. And similarly for the reverse (decryption).

Why is this a Trans Vernam Cipher? Because the graph may be large or small. The larger it is the more security it provides. It may be so large that it will be Vernam equivalent, and it may be so small that brute force will extract it relatively easily. The processing effort is not affected by the size of the graph, only by the length of the pathway, which is the size of the encrypted message. By analogy given a fixed walking speed, it takes the same time to walk, say, 10 miles on a straight stretch of a road, or zig-zagging in a small backyard.

Detailed Procedure:

1. Alphabet Conversion: Map a list of symbols to a three letters alphabet: X, Y, Z. By mapping every symbol to a string of 5 letters from the {X,Y,Z} alphabet, it is possible to map 3^5=243 distinct symbols (a few less than the ASCII list of 256 symbols).

2. **Message conversion**: let m=m_0 be the message to be encrypted, written in the symbols listed in the 243 symbols list. Using the alphabet conversion in (1) map m_0 to m_3 - a sequence of the 3 letters alphabet: X, Y, Z.

3. **DeRepeat the Message:**: enter the letter W between every letter repletion in m_3, and so convert it to m_4. m_4 is a no-repeat sequence of the letters {X,Y,Z,W}. Add the letter W as the starting letter.

4. **Construct a key**: construct a n*m matrix with the letters {X,Y,Z,W}. The matrix will include at least one element for each of the four letters. The letters marking will abide by the *'any sequence condition'* defined as follows: Let i, and j represent two different letters of the four {X,Y,Z,W}. At any given state let one of the n*m elements of the matrix be "*in focus*". Focus can be shifted by moving one element horizontally (right or left), or one element vertically (up or down) – reminiscent of the Turing Machine. Such a focus shift from element to an adjacent element is called "*a step*". The '*any sequence condition*' mandates that for any element of the matrix marked by letter i, it will be possible to shift the focus from it to another element marked by the letter j, by taking steps that pass only through elements marked by the letter i. The 'any sequence condition' applies to any element of the matrix, for any pair of letter (i,j).

5. **Select a starting point:** Mark any matrix element designated as "W" as the starting point (focus element).

6. **Build a pathway on the matrix reflecting the message (m_4):** Use the {X,Y,Z,W} sequence defined by the m_4 version of the message, to mark a pathway (a succession of focus elements) through the matrix. The "any sequence condition" guarantees that whatever the sequence of m_4, it would be possible to mark a pathway, if one allows for as much expansion as necessary, when an 'expansion' is defined as repeating a letter any number of times.

7. **Encrypt the pathway** : Describe the identified pathway as a sequence of edges, starting from the starting point. This will be listed as a sequence of up, down, right, left {U,D,R,L} to be referred to as the ciphertext, c.

The so generated ciphertext (expressed as 2 bits per edge) is released through unsafe channel to the intended recipient. That recipient is assumed to have in her possession the following: (i) the alphabet conversion tables, (ii) the matrix, (iii) the identity of the starting point, and (iv) the ciphertext c. The intended recipient will carry out the following actions:

8. **Reconstruct the Pathway:** Beginning with the starting element, one would use the sequence of edges identified in the ciphertext, as a guide to chart the pathway that the writer identified on the same matrix.

9. Convert the pathway to a sequence of vertices: Once the pathway is marked, it is to be read as a sequence of vertices (the matrix elements identified by the letters {X,Y,Z,W}), resulting in an expanded version of the message, m_{4exp}. The expansion is expressed through any number of repetitions of the same letter in the sequence.

10. Reduce the Expanded Message: replace any repetition of any letter in m_{4exp} with a single same letter.

11. **Reduce m_4 to m_3:** eliminate all the W letters from m_4.

12. **Convert m_3 to m_0:** use the alphabet conversion table to convert m_3 to the original message m_0.

Illustration: . Let the message to be encrypted be: m=m_0="love". Let the alphabet conversion table indicate the following:

l	XYZ
o	ZYX
v	XYZ
e	ZYY

Accordingly we map m_0 to m_3 = XYZ ZYX XYZ ZYY.

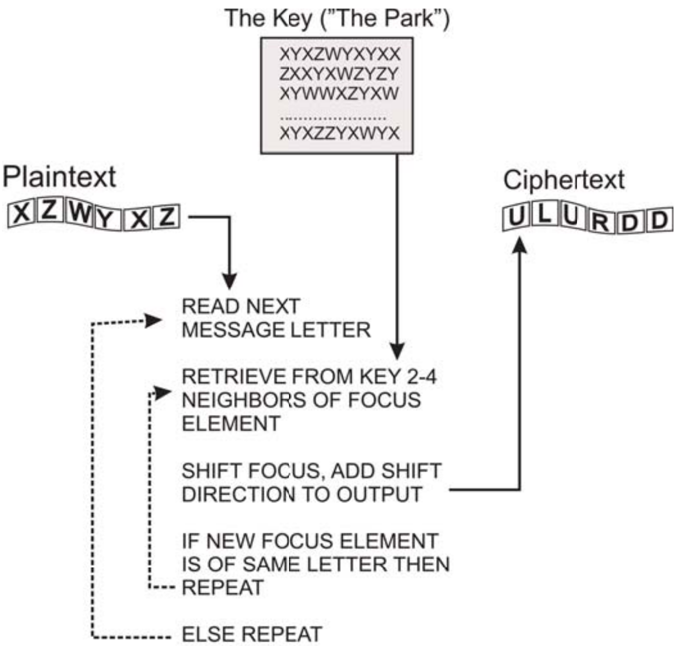
We now convert m_3 to m_4 = WXYZWZYXWXYZWZYWY.

We build a matrix that satisfies the 'any sequence condition':

```
1 2 3   X X Y
4 5 6 = X W Y
7 8 9   Z Z Z
```

Using m_4 as a guide we mark a pathway on the matrix: 5,2,3,6,9,6,5,8,9,6,3,2,5,2,3,6,9,8,5,8,9,6,5,6

The pathway is now defined through the traversed edges, and construct the ciphertext c = URDDULDRUULDULDDLUDLULR. In order to decrypt c, its recipient will have to use the matrix (the graph, the key, or say, “the walking park”), and interpret the sequence of edges in c to the visited vertices: 5,2,3,6,9,6,5,8,9,6,3,2,5,2,3,6,9,8,5,8,9,6,5,6. This is the same pathway marked by the plaintext. Once it is marked on the matrix it can be read as a sequence of the visited vertices: m_{4exp} = WXYYZYWZZZYYXWXYYZZWZZYWY.Which is reduced $m_{4exp} > m_4$: WXYZWZYXWXYZWZYWY; Which, in turn, is reduced to the three letters alphabet: $m_4 > m_3$ = XYZ ZYX XYZ ZYY, which is converted to m = “love”



“Walk in the Park” Encryption Algorithm

“Turing Machine” algorithmic simplicity.
Security achieved via at-will randomness

Walk-in-the-Park as a TVC: There are various procedures, which would translate the matrix (the key) into a natural number and vice versa. Here is a very simple one. Let k be a square matrix (key) as described above, comprised of n^2 letters. Each letter is marked with two bits, so one can list the matrix row by row and construct a bit sequence comprised of $2n^2$ bits. That sequence corresponds to a non-negative integer, k. k will be unambiguously interpreted as the matrix that generated it. To transform a generic positive integer to a matrix one would do the following: let N be any positive integer. Find n such that $2(n-1)^2 < N \leq 2n^2$. Write N in binary and pad with zeros to the left such that the total number of bits is $2n^2$. Map the $2n^2$ bits onto a n^2 matrix, comprised ot 2 bits element, which can readily be interpreted as n^2 letters {X,Y,Z,W}. If the resultant matrix complies with the 'any sequence' condition, this matrix is the one corresponding to N. If not, then increment the $2n^2$ bit long string, and check again. Keep incrementing and checking until a compliant matrix is found, this is the corresponding matrix (key) to N.

It is clear by construction that Walk-in-the-Park is a TVC: the key (the map) gets larger with larger integer keys, and for some given natural number k_{Vernam} a message m will result in a pathway free of any revisiting of any vertex. The resultant ciphertext can then be decrypted to any message of choice simply by constructing a matrix with the traversed vertices fitting that message.

Cryptanalysis: A 9-letters key as in the illustration above will be sufficient to encrypt any size of message m. Alas, a cryptanalyst who is aware of the size of the key will readily apply a successful brute force analysis. Clearly, the larger the size of the key the more daunting the cryptanalysis. As long as the pathway revisits a vertex twice, the resultant cipher is not offering mathematical security, but for a sufficiently large map (key) the pathway may be drawn without revisitation of same vertices -- exhibiting Vernam, (or say, perfect) security. The critical feature of this cipher is the fact that the size of the map (the key) is integral part of its secrecy, so a cryptanalyst has no clear end point within which to conduct brute force cryptanalysis. For further depth and background see Samid 2002, Samid 2004.

III. USAGE SCENARIOS

We describe here a use case that is taken from a project under evaluation. The IOT node in this case is a micro drone equipped with a versatile video camera. The drone is extremely light, it has a small battery, and a solar cell. It is designed to land on large objects like trees and roofs. The camera streams to its operators a life video of the viewable vista. The drone requires encryption both for interpretation of commands, and for transmitting videos. The high-powered multi mega pixel camera may be taping non sensitive areas like public roads; it may stream medium sensitive areas, like private back yards, and it may also stream down highly sensitive areas, like industrial and military zones. The micro drone is dropped in the vicinity of operation, with no plans of retrieval, It should operate indefinitely.

Using Walk-in-the-Park the drone will be equipped with three keys (matrices, graphs): 1. a small hardware key comprised of square flash memory of 500x500 {X,Y,Z,W} letters. This will amount to a key comprised of 500,000 bits. 2. A flash memory holding 1000x1000 {X,Y,Z,W} letters, comprising 2,000,000 bits. 3. A flash memory holding 2500x2500 {X,Y,Z,W} letters comprising 12,500,000 bits. The latter key should provide perfect secrecy for about 6 megabytes of data, and will be used to communicate to the drone transposition keys to be used on another TVC transposition based cipher, not described here.

The determination of the security sensitivity of the photographed area (and the corresponding security level used) may be determined onboard the drone, or communicated from the reception center based on the transmitted pictures.

IV. Summary Notes

We presented here a philosophy and a practice for 'Cryptography of Things' (CoT) -- means to facilitate data security associated with things-nodes in the IP protocol. The CoT is mindful of processing parsimony, maintenance issues, and security versatility. The basic idea is to shift the burden of security away from power-hungry complex algorithms to variable levels of randomness matching the security needs per transmission. This paper presents the notion of Trans-Vernam Ciphers, and one may expect a wave of ciphers compliant with the TVC paradigm. It's expected that the IoT will become an indispensable entity in our collective well being, and at the same time that it should attrack the same level of malice and harmful activity experienced by the Internet of People, and so, despite its enumerated limitations, the IoT will require new horizons of robust encryption to remain a positive factor in modern civil life.

References

1. Auguste Kerckhoffs, « La cryptographie militaire », Journal des sciences militaires, vol. IX, pp. 5–38, Janvier 1883, pp. 161–191, Février 1883.
2. M. Hellman. 1977 “An extension of the Shannon theory approach to cryptography”. IEEE Transactions on Information Theory, V. 23 , 3 1977 , pp. 289 - 294
3. Ma´t´e Horva´th, 2015 “Survey on Cryptographic Obfuscation” 9 Oct 2015 International Association of Cryptology Research, ePrint Archive https://eprint.iacr.org/2015/412
4. Masanobu Katagi and Shiho Moriai "Lightweight Cryptography for the Internet of Things" Sony Corporation 2011 https://www.iab.org/wp-content/IAB-uploads/2011/03/Kaftan.pdf
5. Rein Canetti, Cynthia Dwork, Moni Naor, Rafail Ostrovsky “Deniable Encryption” CRYPTO '97Volume 1294 of the series Lecture Notes in Computer Science pp 90-104Date: 17 May 2006
6. S. Zhou (ZTE Corporation) Z. Xie (ZTE Corporation) 2011 "On Cryptographic Approaches to Internet-Of-Things Security" http://www.lix.polytechnique.fr/hipercom/SmartObjectSecurity/papers/ZhouSujing.pdf
7. Samid, (A) 2015 “Equivoe-T: Transposition Equivocation Cryptography” 27 May 2015 International Association of Cryptology Research, ePrint Archive https://eprint.iacr.org/2015/510
8. Menezes, A. J., P. van Oorschot and S.A. Vanstone. The Handbook of Applied Cryptography. CRC Press, 1997.
9. Samid, 2004 "Denial Cryptography based on Graph Theory", US Patent #6,823,068
10. Samid, 2015 “The Ultimate Transposition Cipher (UTC)” 23 Oct 2015 International Association of Cryptology Research, ePrint Archive https://eprint.iacr.org/2015/1033
11. Samid, 2016 “Shannon’s Proof of Vernam Unbreakability” https://www.youtube.com/watch?v=cVsLW1WddVI
12. Samid, G. "Re-dividing Complexity between Algorithms and Keys" Progress in Cryptology — INDOCRYPT 2001 Volume 2247 of the series Lecture Notes in Computer Science pp 330-338
13. Samid, G. 2001 "Anonymity Management: A Blue Print For Newfound Privacy" The Second International Workshop on Information Security Applications (WISA 2001), Seoul, Korea, September 13-14, 2001 (Best Paper Award).
14. Samid, G. 2001 "Encryption Sticks (Randomats)" ICICS 2001 Third International Conference on Information and Communications Security Xian, China 13-16 November, 2001
15. Samid, G. 2002 " At-Will Intractability Up to Plaintext Equivocation Achieved via a Cryptographic Key Made As Small, or As Large As Desired - Without Computational Penalty " 2002 International Workshop on CRYPTOLOGY AND NETWORK SECURITY San Francisco, California, USA September 26 -- 28, 2002
16. Samid, G. 2003 "Intractability Erosion: The Everpresent Threat for Secure Communication" The 7th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2003), July 2003.
17. Samid, G. 2003 "Non-Zero Entropy Ciphertexts (Stochastic Decryption): On The Possibility of One-Time-Pad Class Security With Shorter Keys" 2003 International Workshop on CRYPTOLOGY AND NETWORK SECURITY (CANS03) Miami, Florida, USA September 24 - - 26, 2003
18. Shannon, Claude, 1949 “Communicaiton Theory of Secrecy Systems” http://netlab.cs.ucla.edu/wiki/files/shannon1949.pdf
19. Smart, Nigel “Cryptography (an Introduction)” 3rd Edition http://www.cs.umd.edu/~waa/414-F11/IntroToCrypto.pdf
20. Stallings Williams, 2002 “Introduction to Cryptography” http://williamstallings.com/Extras/Security-Notes/lectures/classical.html
21. Vernam 1918; Gilbert S. Vernam, US Patent 1310719 Filed 13 September 1918.

Mesh-based Cloud Data Storage & Management Framework for Internet of Things

Hnin Yu Shwe[1] **and Peter Han Joo Chong**[2]
[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Department of Electrical and Electronic Engineering, Auckland University of Technology, New Zealand

Abstract— *The Internet of Things (IoT) is the network of physical objects, devices, vehicles, buildings and other items that enables these objects to collect and exchange data. One of the application domains of IoT is smart building where a number of things are efficiently interconnected in order to combine and integrate the physical world to the information space and to provide the services in an efficient way. In the smart building management systems, the data of building conditional monitoring are increasing rapidly. Facing with these massive, distributed, heterogeneous and complex state data, conventional data storage and management will encounter great difficulties. Conventional infrastructure uses centralized approach, expensive large-scale server, disk array storage hardware and relational database management system; which leads to poor system scalability, higher costs, and essentially, is difficult to adapt to the requirement of higher reliability on real-time state data of the smart building applications. Accurate, fast, open, shared information system is the basis for the IoT applications. In this paper, we proposed a mesh-based data storage and management framework which provides reliable data storage and management of the vast amount of information in the IoT applications. We have evaluated and validated our approach by applying our cloud data storage and management framework in our cyber-physical test-bed, a facility which is intended to test innovative technologies for urban sustainability.*

Keywords: Mesh, Cloud, Data Storage, Smart-building, Internet of Things

1. Introduction

Smart building systems are becoming popular and significant due to the improvement they provide to the quality of life. Urban IoT is aimed to support the concept of smart cities, whose objective is to exploit advanced control and communication technologies to support high quality services [1]. In the scope of the IoT [2], smart building management system requires architecture that is able to deal with large amount of information from its building components, extract valuable building information in a timely fashion, interpret data in real-time or near-real-time to allow for further improvements in reliability. In order to be part of the IoT, all the information and resources of smart building must be accessible from everywhere. These requirements for smart building can be met by utilizing the cloud computing model. Cloud computing efficiently uses distributed resources and thus it has the benefits of high reliability, flexible, scalable and large-scale computing capability [3] [4]. In order to achieve a reliable smart building infrastructure, cloud computing solutions and services must be incorporated.

Addressing the core requirements of smart building architecture and utilizing the modern advanced technologies, in this paper, we proposed a mesh-based cloud data storage and management framework for IoT which offers better architecture than conventional centralized scheme to accommodate a large scale data, analyze timely information and make rapid decisions. Our idea is to have all the computational power as well as control and monitor capabilities in the cloud.

The remaining of this paper is organized as follows. We first discuss the related works in Section 2. We then briefly present our proposed mesh-based cloud data storage and management framework for internet of things in Section 3. Finally, we finally conclude our paper in Section 4.

2. Background

Internet of Things is becoming so persistent that everything is going to be connected to the Internet and its data will be used for various progressive purposes, creating not only information from it, but also, knowledge and even wisdom. In smart building, all the appliances and devices are geographically distributed throughout the building environment and, as mentioned in the introduction, each building component should be able to access and exchange data via certain communication protocol. Data store and management system allows the data to be stored in a systematic manner and enable them to be retrieved, processed and analyzed either immediately or later. Generally, data are stored on centralized magmatic hard disk drives.

In recent years, cloud computing and cloud database become popular and people are interested to get benefits of cloud technologies in the smart applications [5]. A cloud database can be in the form of either a virtual machine instance which can be purchased for a limited time or a database as a service in which the service provider installs and maintains the database, and application owners pay according to their usage.

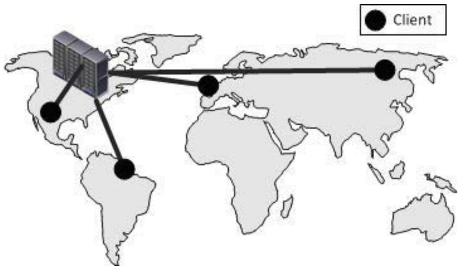


Fig. 1: An example of a centralized topology.

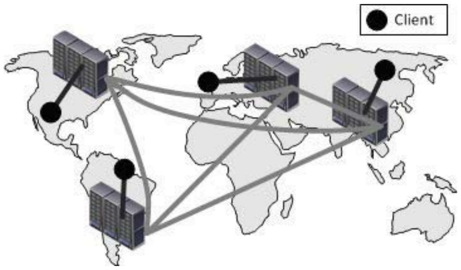


Fig. 2: An example of a distributed topology.

Cloud data storage can either be centralized or distributed. In a centralized cloud data storage system, as shown in Fig. 1, the data are stored at a central stand-alone data storage center which is physically in one location and users typically use an Internet connection to access it. Cloud data storage provider, eg., Dropbox [6], uses centralized topology in which all clients have to connect to the same data center [7]. On the other hand, in a distributed cloud data storage system as in Fig. 2, the data are stored in multiple data centers which are spread over a geographical area and users can access to a different data center closest to it. Google Drive [8] follows distributed topology and uses multiple cloud storage data centers [9] [10].

In addition, cloud data storage can also be classified as either public or private. A public cloud is one in which the services and infrastructure are provided off-site over the Internet by a third-party provider. In contrast, a private cloud is one in which the services and infrastructure are maintained on a private network. Both can offer advantages over traditional data center in the areas of performance and scalability. However, the advantage of private cloud over public cloud is that the private cloud provides greater levels of control and security. Another great benefit of private cloud is the ability to customize the compute, storage and networking components to best suit with specific information technology requirements.

To the best of our knowledge, existing data center in smart building applications adopts traditional centralized data storage system. Some of the challenges in centralized data storage system for smart building applications are as follows:

- A centralized database tends to create bottlenecks if a large number of users access it simultaneously and their needs are substantial since all the data physically reside in one place.
- Centralized data server is not good enough for various types of data formats.
- Difficult to maintain huge amount of data.
- Data availability is not efficient in terms of scalability, recoverability and accessing time.

Our idea is to develop a wireless mesh-based cloud data storage and management platform which is an integration of two modern technologies; wireless mesh network and cloud service. Our proposed data center platform allows end users to make use of the data storage center efficiently and effectively in terms of accessibility, data storage, analysis and data processing and better performance.

3. Proposed System

3.1 System Architecture

Our distributed cloud data storage and management system is mainly composed with two-tier network and it supports hierarchical scheduling or multi-level decision making. High level architecture of our proposed data storage platform can be seen in Fig. 3.

The first-tier is wireless mesh backbone network in which wireless mesh assess points (APs) are used to distributedly store the real-time data in the distributed cloud system. The mesh APs can be installed regionally either in each zone or room to store time-critical or continuous changing type of data. The second-tier is central data storage center which stores the long-term or historical information of the entire building system. The main objective of our approach is to allow the end users to make use of the data storage center efficiently and effectively in terms of accessibility, data storage, analysis and data processing for better performance.

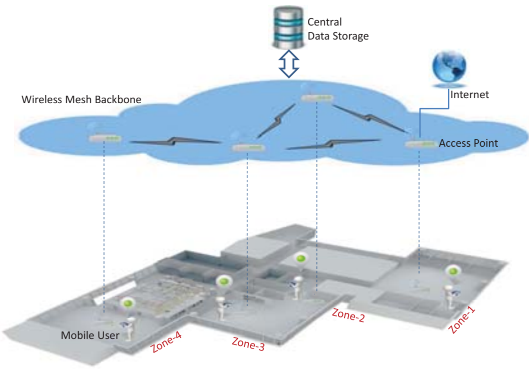
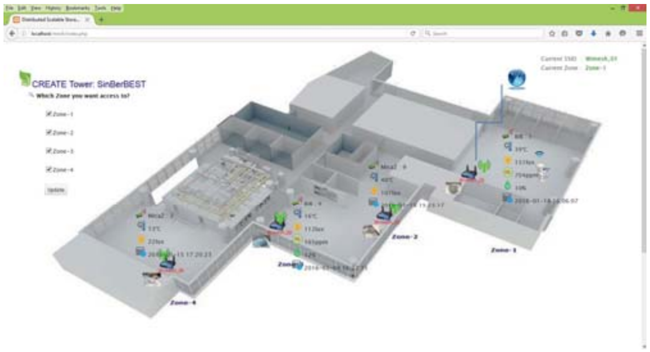


Fig. 3: Mesh-based cloud data storage system for IoT.

3.2 Mesh-based Cloud Data Storage Platform

A mesh network is a network topology in which each node, called mesh AP, relays data for the network. Each mesh AP is configured with internal storage capacity and build up private distributed data storage network to efficiently store the building data. All nodes cooperate in distribution of data in the network. Gateways on the edge collect raw data, store it and perform a first-level of translation and/or analysis. Mesh-based cloud data storage can accommodate large scale data interactions that take place on smart building environments.



CREATE Tower: SinBerBEST

Which Zone you want access to?

☑ Zone-1

☑ Zone-2

☑ Zone-3

☑ Zone-4

Update

Fig. 4: Data distribution between different zones.

Fig. 4 shows data distribution between the different regional mesh clients. Through the proposed platform, different systems are allowed to upload the data to the mesh-based cloud data center through the associated gateways. In Fig. 4, wireless sensor network, one of the IoT components, uploads the sensor data to cloud data center. First-level data processing is performed with the help of cloud computing and then those data are stored in its regional mesh AP.

Proposed platform allows to exchange information between different systems in different zones. The data are available to the other mesh client who wants to obtain the real-time information of the other region in the building. As in Fig. 4, the user (mesh client) in one zone can easily obtain the real-time sensor information of other zone through wireless mesh backbone. In this testing, we used four-radio dual-band wireless mesh AP (MeshRanger MN4300) which is compatible in both indoor and outdoor environment and the real deployment of the mesh APs in our test-bed can be seen in Fig. 5.

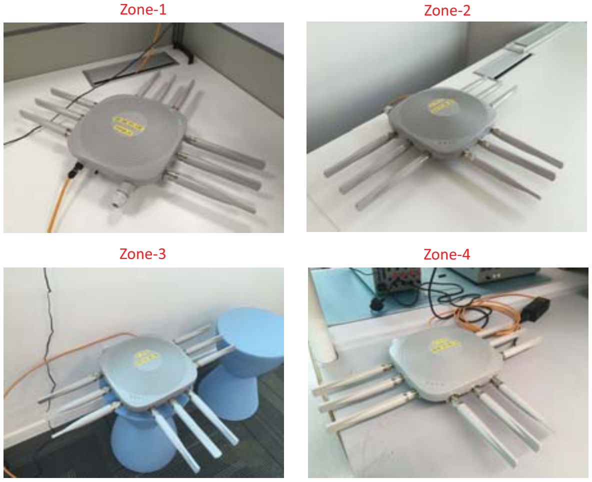


Fig. 5: Mesh-AP installation in different zones.

Cloud platforms are intrinsic to creating a software architecture to drive effective use of smart IoT applications. Cloud data storage has better architecture than centralized systems to process huge, persistent streams of data generated across the utility value chain. In smart building, on-going data streams from multiple sources need to be continuously monitored, verified, and processed given changing conditions, with near-zero latency. With the help of our proposed mesh cloud storage platform, multiple tiers of storage can be linked together to keep data in the most efficient way, transparently to the user.

3.3 Distributed Cloud Computing

Our proposed data storage platform allows users to benefit from the power of the cloud. Proposed system can perform first-level data processing such as aggregation, analysing real-time data securely, easily, assuring data integrity and can scale seamlessly and efficiently as the system requires. We provide an IoT cloud platform to ease and speed up the development and deployment of solutions based on real time stream data. The platform provides a wide range of cloud services, including connectivity, to vertical applications (smart appliances, smart metering devices, smart vehicles, etc.) working with remote devices (sensors, tablets, etc). It allows the interoperability of multiple systems and devices, offering a semantic platform to make real world information available to smart IoT applications as shown in Fig. 4.

Cloud data storage platform provides functionalities to gather, integrate, store and analyze data from the building and enables the building users to understand what is happening in real time in order to proceed with solutions immediately. By integrating with an efficient decision making service, the proposed platform can provide more value and power to all end-user.

3.4 Hierarchical Data Composition

One of the features of our proposed platform is to support efficient hierarchical scheduling and/or decision making by providing the data to the users hierarchically. Our data storage platform provides distributed database service to users. In our platform, we generally classify the data into two types: Type-1 and Type-2.

We define the global and historical data as Type-1 and which are stored in central data storage system. Type-2 data are real-time recurrent or time-critical data and which are distributively stored in mesh APs. Information sharing is available between mesh APs through wireless mesh backbone network. Data can be processed by accessing regional mesh APs in efficient time manner and make decision timely and feedback locally rather than go through the central decision server for every decision making process.

3.5 Database as a Service (DaaS)

Data as a Service (DaaS) model will enable the building management system to continuously monitor the condition of the building and support energy management decisions quickly and accurately. Building management authorities will also be able to use this aggregate, real time data to perform analysis, building maintenance and to improve any performance issues.

3.6 Advantages of the Proposed System

Dynamic Cloud Topology: Cloud topology changes dynamically depending on the request of users.

Improved Throughput: It avoids network congestion at central data center since some of data can be obtained from regional APs and the users do not need to go through the central stand-alone data center for every decision making process which may lead to heavy traffic congestion and bottlenecks.

Data Transparency: It provides data transparency which makes users the transparent use of data storage, analysis and exchange data among building components or different zones.

Lower Data Communication Cost: Data communication costs are lower with our distributed data storage platform since some decisions can be performed locally by providing access to distributed mesh APs in each zone.

On-the-go Scaling: Another advantage of our proposed platform is on-the-go scaling. With the help of the cloud data storage system, the users do not need to worry about running out of storage space or increasing the current storage space availability since cloud provides almost unlimited storage capacity. This is very important feature to store huge amount of data for the large-scale IoT system.

Easy-access-to-information: It provides easy access to geographically distributed information. The users can access the information from anywhere via wireless mesh backbone. This feature makes users more reliable and convenient since the users do not need to care about the geographic location issues.

Massive System: Our proposed platform is a unified data collector for various data types of different sub-systems and helps the other sub-systems to work under one management engine for better performance.

4. Conclusion

We have presented a mesh-based data storage and management framework for internet of things in a smart building environment. Our work addresses two important issues, data scalability and data efficiency. We have implemented our algorithm in our cyber-physical test-bed and validate its effectiveness. From the testing, we could see our proposed platform provides efficient and effective data storage service in terms of data accessibility, transparency and scalability. In the future, we are going to deploy our data storage and management system in the large-scale and real environment.

Acknowledgement

This research is funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

References

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, "Internet of Things for Smart Cities," in the IEEE Internet of Things Journal, vol. 1, no. 1, February 2014.

[2] D. Miorandi, S. Sicari, F. D. Pellegrini and I. Chlamtac, "Internet of Things: Vision, Applications and Research Challenges," in Ad-Hoc Networks, vol. 10, no. 7, pp. 1497-1516, September 2012.

[3] M. N. O. Sadiku, S. M. Musa and O. D. Momoh, "Cloud Computing: Opportunities and Challenges," in IEEE Potentials, vol. 33, no. 1, pp. 34-36 February 2014.

[4] F. Tao, Y. Cheng, L. D. Xu, L. Zhang and B. H. Li, "CCIoT-CMfg: Cloud Computing and Internet of Things-Based Cloud Manufacturing Service System," in IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 1435-1442, May 2014.

[5] W. Hu, T. Yang and J. N. Matthews, "The Good, The Bad and The Ugly of Consumer Cloud Storage," in ACM SIGOPS Operating Systems Review, vol. 44, no. 3, pp. 110-115, July 2010.

[6] https://www.dropbox.com/

[7] I. Drago, M. Mellia, M. M. Munafo, A. Sperotto, R. Sadre and A. Pras, "Inside Dropbox: Understanding Personal Cloud Storage Services," in Proceedings of the 2012 ACM conference on Internet Measurement Conference, Boston, Massachusetts, USA, pp. 481-494, 2012.

[8] https://www.google.com/drive/

[9] E. Bocchi, I. Drago and M. Mellia, "Personal Cloud Storage: Usage, Performance and Impact of Terminals," in 2015 IEEE 4th International Conference on Cloud Networking (CloudNet), Niagara Falls, ON, pp. 106-111, 2015.

[10] I. Drago, E. Bocchi, M. Mellia, H. Slatman and A. Pras, "Benchmarking Personal Cloud Storage," in Proceedings of the 2013 conference on Internet measurement conference (IMC'13), pp. 205-212, 2013.

Compelling Use Cases for the Internet of things

Henry Hexmoor

Computer Science Department, Southern Illinois University, Carbondale, Illinois, USA
Alqithami@gmail.com, hexmoor@cs.siu.edu

Keywords: Social network: virtual team: virtual organization: non-traditional organization: spontaneous and networked organization.

Abstract: Internet of agents are an emerging technological development lacking sufficient teleology. This paper aims to delineate compelling categories of application. We wish to promulgate environments for fostering research in socially aware agency for eclectic teams of humans and machines.

1 INTRODUCTION

Google has launched a campaign called Brillo, designed to allow other companies to program and compute their devices using Google's software technology. Google's Weave is designed for devices that use Weave to also communicate with android devices. Weave uses less ram and will take up less space than its android counterpart. Whereas in the cellular device market the push right now is for more powerful devices with more space, Google Brillo is meant to make tiny applications, highly efficient, very fast.

Recently, it has been suggested that things in the internet of things (IoT) framework be modeled as agent entities (Yu, et. al., 2013). In sharp contrast to passive view entities of things, agent things are active and may take action proactively. Although there are reported architectures [3][4][5] [7], they are at a high level and much of current literature in this area is a call to arms to develop agent based platforms and technologies in order to accommodate seemless interaction between things and humans.

Thing agents must be aware of their environment and must reason about others as peer residents of IoT. Part of this awareness must be when agents account for humans in this inevitably mixed teams of humans and active things. There are conceptual suggestions for accounting for sociality [3][7]. However, a more in depth exploration of sociality is lacking.

2. Social Network of Things

There have been numerous suggestions that things in proximity form social links creating social networks. Minimally, things provide profiles that include goods and services relevant to other things. [6]. Hence, we will refer to them as Social networks for IoT (SIoTN).

SIoTN are predominantly formed for a few common purposes. Chiefly, they are (a) to expedite access and use of goods and services by members of the network (e.g., locating and using a printer), (b) maintain and monitor a pattern of interactivity (e.g., caravan travel of a fleet of automobiles), (c) to achieve a shared goal or a common charter (e.g., detecting faulty components in a complex system). Invariably, multiagent protocols and algorithms are immediately useful.

In the case the beneficiary is a single node, a variant of contract network is applicable where the node initiates a call for assistance, collects bides for help, and chooses assistance from appropriate sources. When the concern involves multiple nodes, a system of work needs to be established that best address the needs. The requirements (or desiderata) can be expressed by one or many nodes that spawn a working group we'll generically call an organization. Such an electronic organization will dwell on the SIoTN. The socially networked IoTs can be considered to be the background fabric on which the in situ organization will appear as a splatter satin

(Figure 1). We have developed conceptual frameworks for these electronic organizations [1][2].

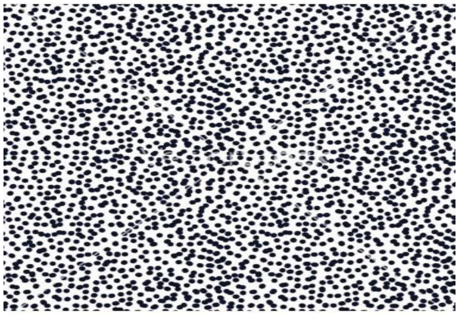


Figure 1. A depiction of prototypical IoT nodes that are formed over the background social network of IOTs.

3. Compelling IoT Use cases

A case for the three Ds—

Many routine tasks are dirty, dull, or dangerous (dubbed D3). Cleaning and maintenance of heavy machinery in a factory and tracking the public restroom cleanliness are dirty task example. Replenishing pet food and water bowls are dull tasks. Opening food can lid can be a dangerous task. In the household, feeding and cleaning after pets and children, taking out trash are other D3 examples. Surely, smart Wifi enabled gadgets will simplify D3 tasks. Things need to be social mainly to meet the possibly changing required D3 standards and report them as specified. Things will communicate the latest required demands installed by the authorized user.

A case for contingencies—

Unlike D3 that happens routinely and expectedly, security and emergency plans are often unscheduled and unexpected. Contingencies are actions designed for what if scenarios. Things must be vigilantly ready. For example, a secure door must detect unauthorized entry by a person entering a house or an animal such as a pest entering a room and must alert the authorities as needed. Rapid evacuation plan for fire safety is another example. Surely, smart Wifi enabled gadgets will be useful for actuation in such tasks. Things need to be social mainly to meet the possibly changing required contingency plans and deploy them as specified. Things will continually communicate the status prior to activation and execution steps of a contingency plan once activated.

A case for improved efficiency—

Many routine tasks such as a commute to work are inefficient on road traffic. Time is wasted on congested roads, in parking lots, excessive driving, excessive braking, hunting for parking spaces, avoiding walking people and pets. Wifi enabled vehicles, roads, parking spaces, pets and pedestrians are useful components when they form a socially communicating network with competing multiple goals of safety and efficiency each paired with specific measures. Each thing would announce its measure corresponding to its goal while others will moderate their actions in order to meet those needs. This will contribute toward a city where public transport and the grid are smart. The efficiency for this case is derived from social cooperation of things that belong to the public and not a personalized gadget or productivity app for one person featured next.

A case for personal productivity—

Each individual's personality creates patterns of interaction that at times are at odds with that person's long term objectives. If all of a person's wifi enabled devices (e.g., smart phone, car, household appliances) and apps (e.g., apps on wifi enabled devices such those provided by mydevices.com) shared personal preferences along with the person's health, well-being, and productivity, it is possible to automate and streamline routine functions for a more coherent and efficient outcome.

4. IoT versus Open Systems Interconnect (OSI) Model

Open Systems Interconnect (OSI) model is a seven layer hierarchy describing how data is transmitted in a computer network. Things in IoT are simultaneously two entities: (a) real world objects possibly interfacing with human users and other things, (b) communication nodes of a computer network and abide by the OSI layers. As real world objects, things are enhanced with with contextual awareness and decision making as well as qualities such as autonomy that qualify them to the proactivity status of agents. The agent ascription of a thing is akin to mind of a machine. Agents (aka minds) are intangible (i.e., virtual). An IoT system is both physical and comprised of agents. A smart city can be considered to be a collection of IoT systems servicing many goods and utilities for occupants of a city, which is a population of humans, machines, (and their agent counterparts). Agent consideration is a logical overlay link layer atop other OSI consideration of communication. The logical interpretation can be extended to a group of agents engaged in interaction with common enduring objectives forming an organization, which can be considered to be virtual as well. Institutions and societies are yet larger in scope and virtual. Virtual entities will never surpass real world physical things but only extend them with meaningful conceptions.

IoT for VANET is a perfect use case that combines prior use cases. Cars and human drivers are all physical entities accompanied by driverless agents and the vehicle ad hoc networks forming VANET communication networks. There are multitude of transportation, traffic, and driving objectives shared among VANET agents. Avoiding collision and near miss among cars is an objective. Obeying all driving rules and regulations is another objective. Efficiency of traffic and avoiding congestion is a transportation and traffic objective. An IoT of VANET may have dirty, dull, or dangerous objectives as well as increased safety objectives as is the case with driverless cars. Contingency and productivity could also be objectives for a VANET.

5. An Envisioned Vehicular IoT Case

An envisioned vehicular interaction is to install wifi enabled devices in vehicles, in road segments, in road intersections, and in train carriages. Vehicles will report to their inhabiting road segments about whether they are entering or exiting from the specific section of the road. Cars will report their direction and speed of movement. At the intersections, roads segments will report to corresponding nearest intersection with the number of cars on that road segment and the vehicle speeds. The road segments will also report lane closures or other abnormalities that would alter “normal” traffic flow. Similarly, a railroad crossing will report to nearest crossroad intersection if there is a train using that crossing. Intersection stations will communicate with other intersections about the number of vehicles headed in their direction.

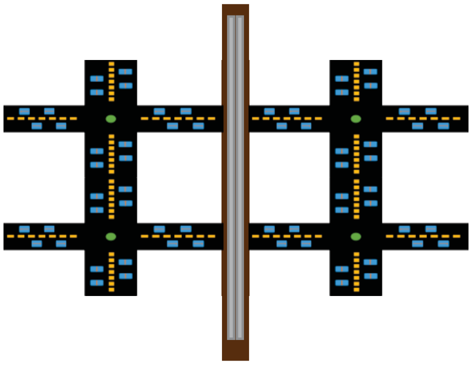


Figure 2. A typical set of intersections, road, and rail segments

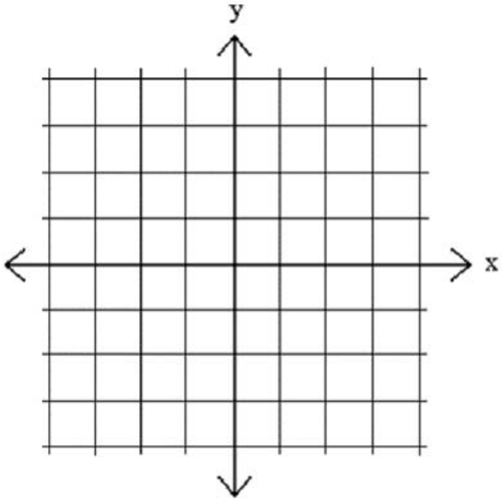


Figure 3. A typical roadway intersection

A scheduling algorithm will track request to use the intersection and will sequentially permit intersection use. This is the vision depicted in Figure 4.

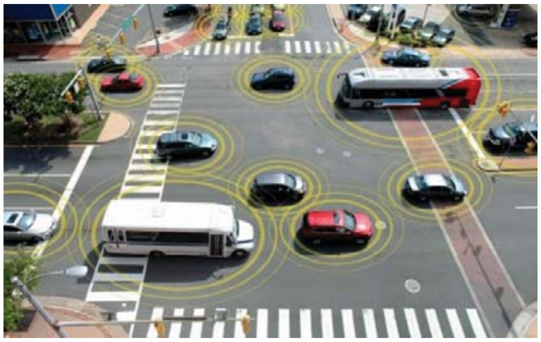


Figure 4. A typical busy intersection full of vehicles.

6. Conclusions

Clearly, Internet of things is a technological leap and with that requires a sober roadmap for categories of research and development spanning allied disciplines including computer networks, social networks, smart electronic devices, and multiagent systems

Initial stages of IoT must establish standards and protocols for communication and interaction. Beyond that, we will need to incorporate methods from varied engineering, social, and applied sciences. We have delineated a few compelling use cases chief among them vehicular ad hoc networks. New York City has over half of its 12,460 intersections controlled by a centralized computer network, and other cities such as Toronto are closer to 83% coverage. Intersections control is one of the most promising application areas for IoT.

Acknowledgement. A group of my students implemented prototypes of vehicular use case in Python described in section 5 in a course during spring 2016.

References

1. Alqithami, S., Hexmoor, H., 2014. Modeling Emergent Network Organizations, Web Intelligence and Agent Systems, Vol. 12, No. 3, pp. 1570-1263, IOS.
2. Alqithami, S., Hexmoor, S., 2015. Ubiquity of Network Organizations: Paradigmatic Perspective and Synergistic Effect, In proceedings of CTS, Atlanta, IEEE.
3. L. Atzori , A. Iera , G. Morabito, M. Nitti, 2012. The Social Internet of Things (SIoT) – When Social Networks meet the Internet of Things: Concept, Architecture and Network Characterization, Journal of Computer Networks 56 (2012) 3594–3608, Elsevier.
4. P. Chamoso, F. Prieta, J. Francisco de Paz, J. Corchado, 2015. Swarm Agent-Based Architecture Suitable for Internet of Things and Smartcities. DCAI 2015: 21-29
5. Giancarlo Fortino, Antonio Guerrieri, Wilma Russo, Claudio Savaglio, 2013. Integration of Agent-based and Cloud Computing for the Smart Objects-oriented IoT, Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design, IEEE.
6. Michahelles, F., Probst, P. 2014. Object Circles: Modeling Physical Objects as Social Relationships, In International Conference on Mobile and Ubiquitous Multimedia, ACM.
7. H. Yu, Z. Shen, C. Leung, 2013. From Internet of Things to Internet of Agents, IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, IEEE, Singapore.

Water Quality Running On Droids

Elias Klassen, Ryan Wieler, Maxim Krivoshchekov, Ziang Wang, Chunzi Jiang, Kelsey Wiens, Ken Ferens, Robert Mcleod, Shamir Mukhi

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba, Canada
{Ken.Ferens@umanitoba.ca}

Abstract—This paper reports on a cost-effective, modularized, cloud-based, water quality monitoring system. The system architecture consists of a portable water quality measuring device (with temperature, pH, and turbidity sensors), an Android phone application, an Internet connected database, and a web interface. The system provides a functional proof of concept for large scale implementation of a public network of water quality measuring devices.

Keywords—Water quality, Internet of Things; crowd sourcing; mobile sensor node; cloud server, phone application, networking software.

I. Introduction

Water is an essential resource in all parts of the world, and water quality directly affects the health of a population. While water quality is tested in designated areas around the city, some rural and remote areas lack water quality testing systems. Without a way to ensure the adequate quality of a water source, many people could be exposed to harmful contaminants. Implementing a system that relays real-time information about water quality at the point-of-use would assist in locating and detecting contaminants faster. Inexpensive equipment able to collect water quality data would ease some of the burden of monitoring a city's water supply and would assist in better protecting the public.

The goal of the Water Quality Running on Droids (WaQRoD) project is to demonstrate the proof of concept for a water quality testing node. The node must be inexpensive with the ability to relay basic water quality data to a user. With a proof of concept, companies or organizations will be able to implement upgrades to the node to suit different purposes. It is possible for the node to be upgraded to support a variety of more complex sensors. Additionally, it could be upgraded to support real time monitoring which could be used to implement an early warning detection system. This would help provide health organizations with wide spread data, decreasing the response time when poor water quality is detected.

Some of the current methods of testing involve devices that cost hundreds to thousands of dollars [1] and often involve trained professionals to operate. The WaQRoD system would use non-intrusive data gathering methods to gather large amounts of data with minimal overhead costs using portable water quality measuring devices that consumers would purchase to monitor their own water quality.

The WaQRoD project aims to provide low cost water analysis to diverse rural and urban settings. The solution is a cost-effective, modularized, water quality monitoring system. The implemented WaQRoD system architecture (Figure 1) is a combination of five main components, grouped into three subsystems:

1. Sensor Node: Water quality sensors grouped with a portable data acquisition device.
2. Phone Application: A smartphone application serving as an interface between the Sensor Node and the Central Server.
3. Central Server: The Web Database grouped with the Web Interface, providing cloud data storage and an interface to the database.

This architecture is modular and scalable, making it a candidate for large scale implementation. The software components used in the WaQRoD project are either created by the WaQRoD team, or provided to the public as open source.

The paper reports on (i) the design and implementation of the WaQRoD project, (ii) the testing and verification of the project, and (iii) the results and future development. The project is implemented in small scale for prototyping purposes for a large scale development. The project's scope includes the design and implementation of a portable sensor node, the design and implementation of an Android phone application, the design and implementation of a web database, and the design and implementation of a web interface.

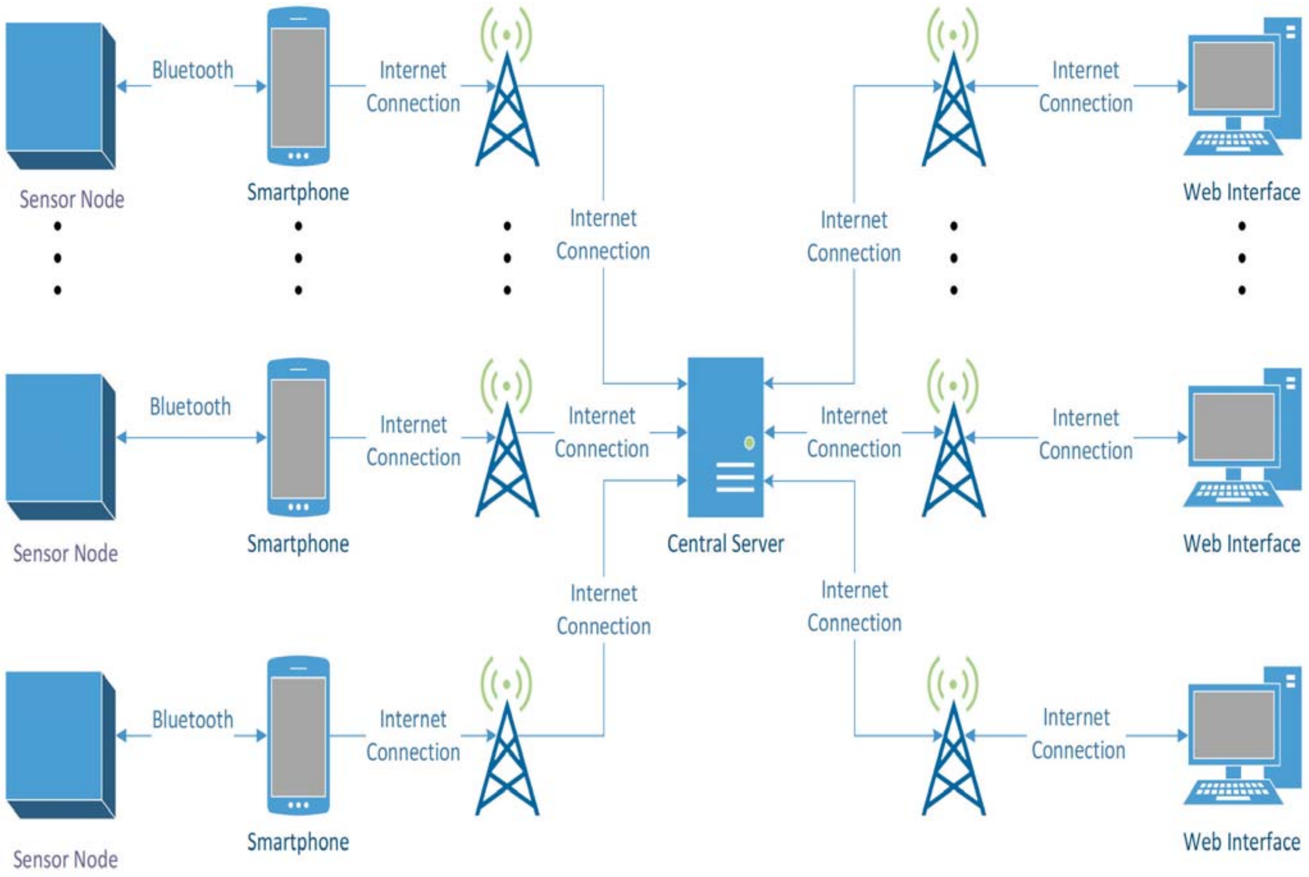


Fig. 1 WaQRod architecture.

The remaining parts of this paper include: Section 2: Water Quality Sensors– Details of the design process, implementation, and verification of the sensors used in the Sensor Node. Section 3: Sensor Node – Details of the design process and implementation of the Sensor Node. Section 4: Phone Application – Details of the design process and implementation of the Phone Application. Section 5: Web Interface and Database (Central Server) – Details of the design process and implementation of the Web Interface and Web Database. Section 6: System Integration and Test – Descriptions of how the system was integrated and tested, and the results. Section 7: Results – Discussions on the final project implementation, and future work. Section 8: Conclusion – A summary of the project and final remarks.

II. System Design

1) *Sensor Considerations and Design*

Multiple characteristics of water need to be tested to determine the quality of water. Characteristics chosen should give proper indication of the quality of the water. Some commonly used characteristics are: Aesthetic Objective Coliforms; Fecal Conductivity; Total Dissolved Solids Sodium; Turbidity; Calcium and Magnesium; Sulphate; Hardness; Escherichia coli; Temperature; and pH [2].

In this paper, water quality sensors were chosen considering (i) the diverse range of outputs (Analog, Digital, Serial), (ii) cost efficiency, and (iii) availability. As such, the sensor types chosen for this project are pH, temperature, and turbidity.

To determine acceptable ranges of these water characteristics, the City of Winnipeg's water quality guidelines were adopted. The test results of pH, temperature, and turbidity of Winnipeg's water distribution system in 2014 are listed in Table 1 [3].

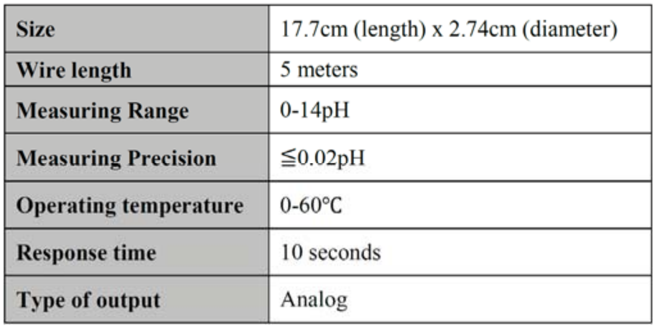
Table 1 Winnipeg, 2014, water quality test results.

What is being measured	How it is measured	Guideline and regulation	Winnipeg average	Winnipeg range	Comment
pH	Based on relative amount of free hydrogen and hydroxyl ions	Between 6.5 and 8.5	7.57	7.33 to 7.74	Meets the guideline
Temperature	Degrees Celsius	No more than 15	10.4	1.3 to 25.5	Does not always meet the guideline
Turbidity	Nephelometric Turbidity Units (NTU)	No Guideline	0.28	0.09 to 18.1	-

B. *pH Sensor*

A pH measurement of water indicates the alkalinity/acidity of the water. A higher value of pH is a sign of alkalinity, which may cause bad taste and odour, and a lower value indicates acidity, suggesting possible presence of heavy metals. In this work, the pH sensor was required to measure pH levels between 6 pH and 9 pH, with a resolution and accuracy of 0.1 pH. The Haoshi Analog pH Meter Pro has been chosen for this project. The sensor is portable and waterproof. The pH sensor selected provides an analog voltage as an output, which is then converted in software to a pH value between 0 and 14. The specifications of the pH sensor are listed in Table 2. Verification of the pH sensor involved using a calibrated pH meter and solutions with known pH values. The calibration values were determined by testing the known solutions using both the calibrated meter and pH sensor at the same time and then comparing the results.

Table 2 Specification of the Haoshi pH sensor.



Size	17.7cm (length) x 2.74cm (diameter)
Wire length	5 meters
Measuring Range	0-14pH
Measuring Precision	≦0.02pH
Operating temperature	0-60℃
Response time	10 seconds
Type of output	Analog

C. *Temperature Sensor*

According to the United States Geological Survey, the temperature of water can “affect the ability of water to hold oxygen as well as the ability of organisms to resist certain pollutants” [4]. Knowing the temperature of water provides an indication of potential bacterial growth. A temperature sensor informs consumers of possible harmful changes in the water. The range requirement of the temperature sensor has been set to above and below the Winnipeg water distribution system temperature range (Table 1). Research shows that the temperature at point of use should not exceed 25℃ [5]. In this work, the temperature sensor was required to record with a resolution of 0.5 °C, with an accuracy of ±0.5 °C. The Dallas Semiconductor DS18B20 [6] digital temperature sensor has been chosen for this project. An analog mercury thermometer was used to test the sensor. Both the temperature sensor and thermometer were placed into the same solution and their readings compared to verify the sensor’s accuracy and resolution.

D. *Turbidity Sensor*

Turbidity is a measurement in NTUs which represents the transparency of a liquid. The level of the transparency depends on the suspended particles in the solution; the more particles suspended in the solution, the higher the turbidity value. Suspended particles in water suggest possible traces of harmful contaminants, such as bacteria. Research shows that the turbidity value of drinking water should be ideally lower than one NTU. Water needs to be filtered at least once before consumption if the turbidity value is between one to 20 NTU. If the turbidity value falls in the range of 50 to 100 NTU for more than 24 hours, this could indicate a serious issue with the water, and the water supply should be analyzed before further consumption. When the turbidity is higher than 100 NTU, the water must not be consumed [7].

In some instances it is possible for the color of tap water to be yellow or even brown. In these cases, the turbidity of water may exceed the range given in Table 1. To account for this, the range of the turbidity sensor was selected to be larger than the range listed in Table 1. In this work, the turbidity sensor was required to detect the range of 0-100 NTU of the solution, with subranges as follows: 0 NTU, 0-20 NTU, 20-50 NTU, 50-100 NTU, and >100 NTU. The Lagoele Electronics D031 turbidity sensor was chosen for this project. Once calibration had begun with the selected sensor, it was found that the accuracy and stability of the turbidity sensor were less than expected because if it’s high sensitivity to light. The microcontroller will indicate a range of turbidity values instead of a precise NTU value based on the turbidity sensor reading. The thresholds of turbidity values were chosen to be 0 NTU, 20 NTU, 50 NTU and 100 NTU. In order to test the water turbidity, the sensor required waterproofing. The area around the wire connections was waterproofed by using hot glue and liquid silicon. The turbidity sensor tests the water turbidity by testing the percentage infrared light intensity received by one side of the turbidity sensor. As shown in the Figure 2, one side of turbidity sensor generates infrared light, the other side receives the light and generates a voltage output [8]. If the water is clear, the light intensity is unaffected. If the water is turbid, the light deflects on particles suspended in the water reducing the intensity. A circuit is required between the sensor and microcontroller to ensure that the reading of a solution between zero and 100 NTU outputs a reasonable and stable value. The wiring circuit includes 470 Ω and 4.7 kΩ resistors which provide a stable output voltage and a detectable voltage difference between each range. The circuit contains a filter design with a 1 kΩ resistor and a 1μF capacitor to further stabilize the output. Due to the instability of the turbidity sensor, the sensor is required to be zeroed before testing. The output of the turbidity sensor is converted to a number from 1-5 which represents the range of the turbidity value.

Due to the sensitivity of the turbidity sensor, any external light source might affect the output reading of the sensor. In order to minimize the error of the turbidity sensor, the sensor must be tested in an absolute dark environment.

To verify the performance of the turbidity sensor, a solution’s turbidity is measured using both calibrated turbidity meter and the turbidity sensor, and the results are compared. This is repeated with multiple solutions with different turbidity values (see Fig. 2).

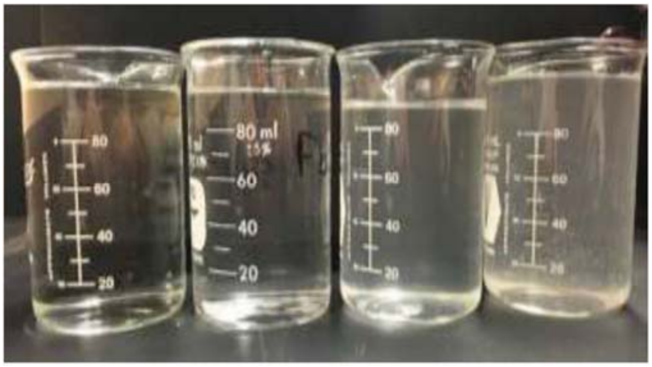


Fig. 2 Solution of 0NTU, 20 NTU, 50 NTU and 100 NTU.

E. Sensor Summary

The pH, temperature and turbidity water characteristics were chosen for measuring the water quality. The pH and temperature sensors were successfully connected to the microprocessor and verified (**Error! Reference source not found.**). The output of the turbidity sensor was changed from specific NTU values to ranges because the accuracy and stability of the turbidity sensor was less than expected.

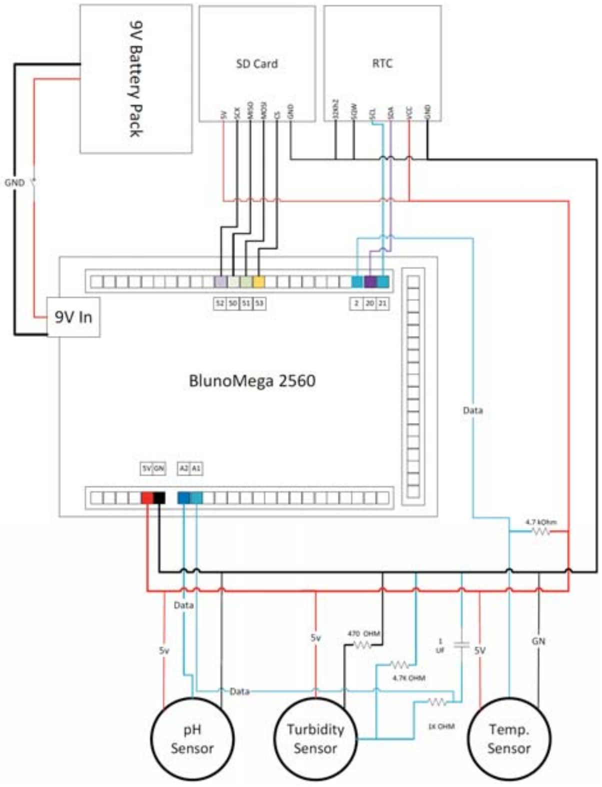


Fig. 3 Sensor/microcontroller wiring diagram.

For future revisions, the types of sensors used can be expanded to measure other characteristics of water quality. It not recommended to use the turbidity sensor selected for this project, as it is too unstable to be used without software, electrical, and hardware modification. A higher quality turbidity sensor would provide more accurate and stable readings. Additionally, the sensor wiring circuits can be implemented in a printed circuit board (PCB) or other topology to reduce the noise, lower the power consumption and provide stronger connections.

1) Sensor Node Design

The Sensor Node component of the project exists to collect and store measurements from the water quality sensors, and transmit the data to a phone application. In this work the sensor node was required to store and delete data locally in non-volatile memory; locally stored data wirelessly over Bluetooth; operate on battery power; manually turned on and off; and provide a timestamp for each measurement.

The Sensor Node was designed with the following characteristics in mind:

1. Portability: the ability of the Sensor Node to operate in different locations while functioning at full capacity only using the internal power supply.
2. Protection: the ability of the Sensor Node to be transported and operated in non-optimal environments without the risk of being damaged.
3. Data Collecting: the ability for the sensor node to read in data from sensors.
4. Storage: the ability for the data to be preserved in memory while powered off.
5. Communication: the ability for the Sensor Node to interface and exchange data with the phone application and water quality sensors.
6. Time Stamping: the ability for the Sensor Node to time stamp data collected.
7. Cost Efficiency: a minimal to no cost software implementation and a minimal cost hardware implementation.

To meet the desired characteristics above, the Sensor Node was designed using a microcontroller, water quality sensors, a real time clock (RTC), a secure digital (SD) card and SD card reader, a battery pack, and a case (Fig. 4).

The microcontroller interfaces and communicates with the sensors and the Phone Application. The battery pack allows the Sensor Node to operate in different locations without access to the power grid. The case provides a water resistant environment and also helps absorb shock to protect the electrical components. The RTC provides a way for the microcontroller to time stamp collected data measurements. Lastly, the SD card and SD card reader provide a non-volatile memory source to ensure data is not lost in the event of power loss.

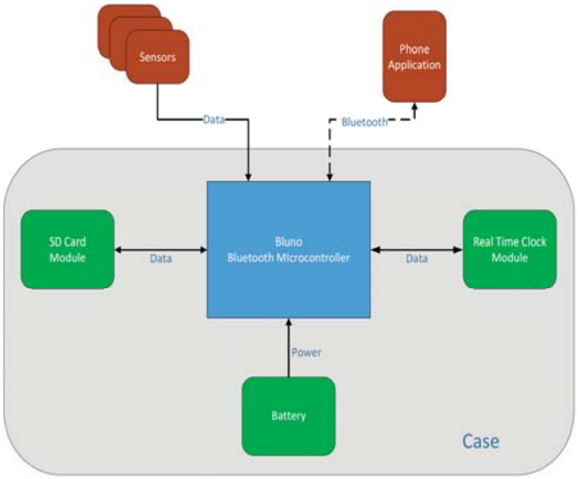
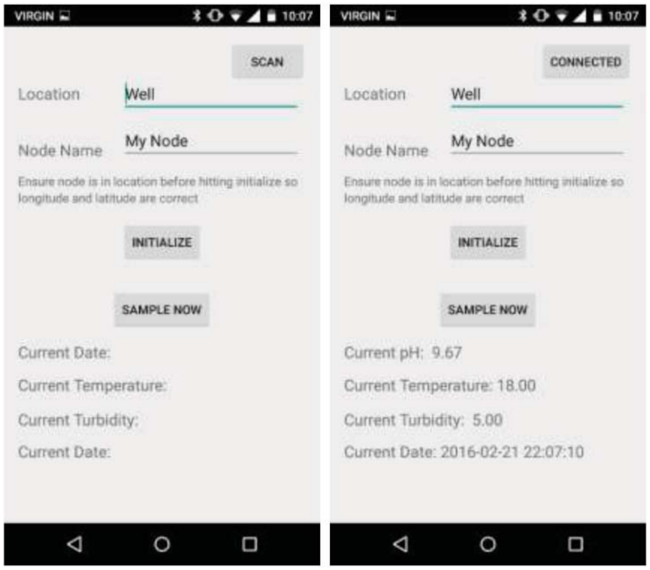


Fig. 4 Sensor node architecture.

2) *Phone Application*

In this work, the phone application was required to connect and send/receive data wirelessly over Bluetooth to the Node; store data received from the Node locally on the phone in non-volatile memory; connect and insert data wirelessly over HTTP to the Central Server; retrieve the phone's GPS location; command the Node to sample data immediately, and show the results; and log in credentials when submitting data to Central Server. The Phone Application was developed and verified for the Android OS. The Fig. 5 shows example activity screens of the phone application.



Location Well

Node Name My Node

Ensure node is in location before hitting initialize so longitude and latitude are correct

Current Date:
Current Temperature:
Current Turbidity:
Current Date:

Location Well

Node Name My Node

Ensure node is in location before hitting initialize so longitude and latitude are correct

Current pH: 9.67
Current Temperature: 18.00
Current Turbidity: 5.00
Current Date: 2016-02-21 22:07:10

Fig. 5 Sensor node gathering acitivy screens.

The application connects to the Sensor Node via Bluetooth communication to command the node and retrieve stored measurements. The measurements are uploaded from the Phone Application to the Web Database via the Internet.

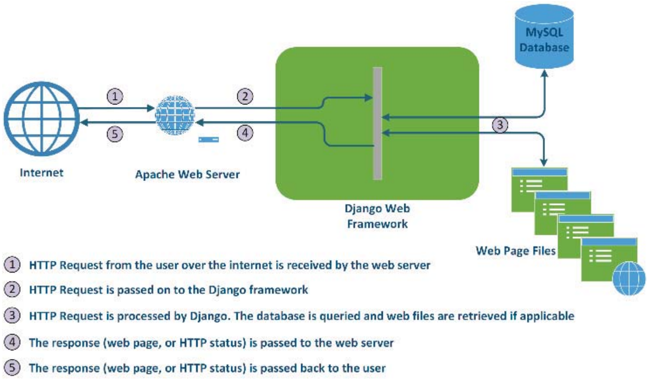
The Phone Application is a stand-alone piece of the WaQRoD system and future development on a large scale would only involve increasing the appeal of the application and developing the application for more phone platforms.

3) *Web Interface and Database Design*

The Web Interface and database component of the project exists to store collected data from the distributed Sensor Nodes, and provide a user interface to the data. The Web Interface and Web Database implementation of the WaQRoD project required the following characteristics:

i. Accessibility: The ability for data to be submitted and viewed from anywhere with an Internet connection.
ii. Storage: The ability to store data in non-volatile memory in a scalable and efficient manner.
iii. Data Presentation: The ability for the data to be presented in a convenient manner.
iv. Security: Protection against unprivileged access to the measurements or user data.
v. Cost Efficiency: A minimal to no cost software implementation.
vi. Flexibility: The ability for the storage, data presentation, and security to be configured.

The present architecture is a single cohesive system integrating the required characteristics; its foundation is made up of a web server, a web design framework, and a local database, as shown in Fig. 6.



1 HTTP Request from the user over the internet is received by the web server
2 HTTP Request is passed on to the Django framework
3 HTTP Request is processed by Django. The database is queried and web files are retrieved if applicable
4 The response (web page, or HTTP status) is passed to the web server
5 The response (web page, or HTTP status) is passed back to the user

Fig. 6 Web server and database architecture.

The Apache web server provides open Internet accessibility to the web design framework, and was integrated with Django. The Django web development framework provides flexible front end web design for data presentation and a convenient interface to the database storage. The MySQL database provides flexible mass scalable data storage, and is integrated with the Django framework. These three main components are implemented on a single host running Ubuntu (Linux), and together make up the Central Server. All software components of the Central Server are open source and contain security functionality that can be used to protect the data against intrusion.

F. Web Interface

The Web Interface is what the user sees when the project's website is accessed. The purpose of the interface is to provide the user with presentation of the collected data in an effective manner. In order to accomplish this purpose, the Web Interface is designed with three main components:

i. Index View: numerical view of each measurement log.
ii. Map View: a location view of where the measurement logs were taken geographically.
Graph View: a graphical view of the trending data.

Another component of the Web Interface is security, which is handled by the Web Interface and the web framework. The framework allows the designer to restrict access to users, and the interface requires the user to provide authentication to access appropriate portions of the data.

The map page provides the user with a geographical view of the location of the Sensor Nodes which are logging the data. Each location alias displayed on the Index page maps to a point on the map (see Fig. 7).

A graph displays a single sensor's values over a given date range for a specified location. Each of the three sensor values can be displayed as a chart. The date range of the displayed values can be specified using the calendar date range selector. Additionally, the graph allows the user to zoom in and pan over the values. Administer access of this page gives the option to filter the locations that belong to a specific user.

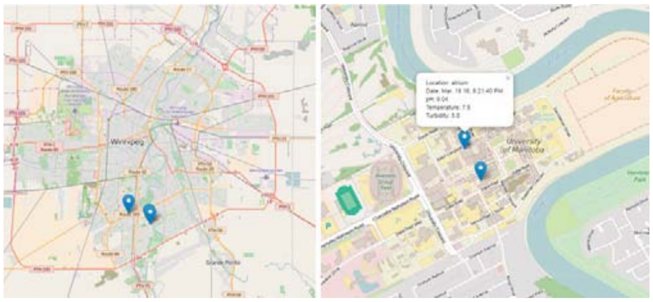


Fig. 7 Example map view of the web interface.

III. Experimental Results

The finished project is a functioning system consisting of two fully functional WaQRoD nodes, the WaQRoD Phone Application, the WaQRoD Database, and the WaQRoD Web Interface. The system was run several times to obtain performance measures, which were compared with the original proposed metrics (Table 3).

The SD Card Reader was added to the project after the proposal as a solution to the need for large non-volatile memory. The real time clock was a part of the original design, and its performance metric was omitted in the proposal by mistake. The GPS location performance metric was added to the project after the proposal as a solution for the need for accurate geographical positioning of the Node.

The temperature sensor accuracy metric was originally a specification of the hardware. When interfacing with the temperature sensor, the maximum resolution output by the sensor observed to be 0.5 °C. This resolution was determined to be sufficient for the project, and the metric was modified.

The turbidity sensor was observed to be very sensitive to external light, which affected the accuracy of the measurement. The solution was to reduce the accuracy to threshold ranges rather than specific NTU values, and extend the measuring range.

The cellphone Android OS version level was reduced, exceeding the original metric, without compromising the features or performance of the WaQRoD Phone Application.

The extended battery life of the WaQRoD node was pushed out of the scope of the project because the focus changed to simply proving operation on battery power alone. The microcontroller used in the node was not selected for long-term battery use. Informal battery observations were recorded.

The microcontroller analog to digital converters and operating temperature were incorrectly included as a performance metric in the proposal, as they are simply a component specification.

1) Potential Applications and Discussion

The WaQRoD system would be useful in a large water distribution system, as well as remote locations. The system could be applied to the large distribution network to monitor the many branches of the network to detect any issues in specific areas of the distribution network. One advantage with the WaQRoD system is that an organization would not need to pay the entire cost of the Sensor Nodes if the public were to cover a portion of the cost.

This paper conveys that the technology is possible as demonstrated by student project, and that it can be expanded as new sensors become available, but it will need further work for it to be operationalized.

Current monitoring systems are very expensive [1] and may require trained technicians to setup and use, whereas the WaQRoD system is inexpensive and only requires an Android Smartphone and Internet connection to use. Chosen water quality sensors for this project limit the accuracy and diversity of data compared to a higher end water monitoring system, but with mass use the WaQRoD system can provide early detection to more dangerous water conditions (for example, lead causes the water to become more acidic, which the WaQRoD system could detect).

If the WaQRoD system was deployed in a large distribution environment, public health authorities would be able to see the current water quality over the deployed area. This would allow public officials to react sooner to a water issue that is past their testing stations and let users know if the water quality issue is an isolated incident or the entire area is being affected. Deployment in remote areas would allow public health officials to monitor the ground water to determine if there have been any unknown disturbances in the water.

Table 3 System performance measurements.

Item		Feature	Original Requirement	Additional and Updated Requirements	Outcome of Updated Requirement
Node	Microcontroller	Bluetooth	Bluetooth 2.0 or higher	N/A	Success
		Operating temperature	-30°C to +40°C	Removed. Part specification.	N/A
		Analog-to-Digital Converters	At least four channels, At least 10bit resolution	Removed. Part specification.	N/A
		Interface Type	GPIO, I2C, SPI, Serial	N/A	Success
	SD Card Reader	Non-volatile memory	N/A	Store the data from the Node with or without the main battery power	Success
	Real Time Clock	Power-interruptible Real-time	N/A	Keep track of the current time with or without the main battery power	Success
	Battery	Voltage	7V to 12 V	N/A	Success
		Current Draw Active	>100 mA at 0.5 hour/day nominal	Pushed out of project scope	N/A
		Current Draw Standby	< 3 mA	Pushed out of project scope	N/A
		Lifetime	30 Days (3600 mAh)	Pushed out of project scope	N/A
	pH Sensor	Working Voltage	3.3V to 5V	5V	Success
		Measuring Range	6pH to 9 pH	N/A	Success
		Accuracy	± 0.1pH	N/A	Success
	Temperature Sensor	Working Voltage	3.3V to 5V	5V	Success
		Measure Range	0°C to 30°C	N/A	Success
		Accuracy	±0.1°C	±0.5°C	Success
	Turbidity Sensor	Working Voltage	3.3V to 5V	5V	Success
		Measure Range	0NTU to 50 NTU	0NTU to 100NTU	Success
		Accuracy	0.1 NTU	Convert sensor output to threshold values	Success
Cellphone		Bluetooth	Bluetooth 2.0 or higher	N/A	Success
		System	Android 4.4 or higher	Android 4.0.3 or higher	Success
		WiFi	Required	N/A	Success
		GPS Location	N/A	Required	Success
Database Server		Storage	At least 100GB	N/A	Success
		Operating System	Ubuntu 14 or greater	N/A	Success
		Network	Network connection required	N/A	Success

IV. Conclusions and Recommendations

The purpose of the WaQRoD project was to design, implement, and verify the proof of concept of a large scale network of water quality measuring devices. As part of this process:

i. Two Sensor Nodes were designed and implemented with pH, temperature, and turbidity sensors to gather water quality measurements.

ii. The Phone Application was designed and implemented for the Android OS which successfully controls the Sensor Node and provides an interface between the Sensor Node and the Central Server.

iii. The Central Server (Web Interface and Web Database) was implemented on a single local host which successfully gathers measurement data submitted by the Phone Application and provides a user interface to the collected data.

The WaQRoD project was completed successfully as samples from different sensors were taken and stored on the Sensor Node, transmitted to the Phone Application, and then transmitted to the Central Server, from where they could be viewed over the Internet.

2) Recommendations

The successful completion of the WaQRoD project represents a strong base for future revisions, and it can be used as a starting point for future development. However, further work is needed for it to become operational, and there are several areas needed for improvements. Future extensions of this project could include the following:

i. Applying more reliable sensors, and different types of sensors, such as higher quality pH sensors, and the addition of E.coli sensing and testing.

ii. Acquiring guidance, support, and approvals from water quality regulatory bodies and government agencies.

iii. Developing a Sensor Node with a better battery lifetime, and a stronger electrical circuit.

iv. Developing plug and play sensors that could be detected in real time and added to the list of sample data (this would be a heavy improvement as all three subsystems would have to be modified to incorporate plug and play sensors).

v. Expanding the Phone Application to other software platforms.

vi. Allowing the Sensor Node to push data to the Central Server directly to allow real time data gathering.

vii. Improving the user experience of the Web Interface.

REFERENCES

[1] Grainger, "Multi-Parameter Meters - Water Testing Equipment and Meters - Grainger Industrial Supply," [Online]. Available: https://www.grainger.com/category/multiparameter-meters/water-testing-equipment-and-meters/lab-supplies/ecatalog/N-kvg?ssf=3. [Accessed 27 February 2016].

[2] Alberta Agriculture and Forestry, ""Water Quality Testing: Common Water Quality Terms/Parameters," 7 January 2014. [Online]. Available: http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/wqe11091. [Accessed 29 February 2016].

[3] City of Winnipeg, "2014 Winnipeg distribution system water quality test results," 2015. [Online]. [Accessed 29 September 2015].

[4] Netcraft, "September 2015 Web Server Survey," 16 September 2015. [Online]. Available: http://news.netcraft.com/archives/2015/09/16/september-2015-web-server-survey.html. [Accessed 21 February 2016].

[5] American Water Works Assiciation, 14 September 2015. [Online]. Available: http://www.awwa.org/. [Accessed 19 March 2016].

[6] M. I. Products, "Maxim Integrated," 2015. [Online]. Available: https://datasheets.maximintegrated.com/en/ds/DS18B20.pdf.. [Accessed 19 March 2016].

[7] H. S. B. RPBio, [Online]. Available: http://www.env.gov.bc.ca/wat/wq/BCguidelines/turbidity/turbidity.html.. [Accessed 19 March 2016].

[8] Wikipedia, "Flint Water Crisis - Wikipedia, the free encyclopedia," [Online]. Available: https://en.wikipedia.org/wiki/Flint_water_crisis.. [Accessed 24 February 2016].

A Study on Middleware for IoT

A comparison between relevant articles

Carlos Albuquerque[1], **Aércio Cavalcanti**[1] , **Felipe S. Ferraz**[1, 2] **and Ana Paula Furtado**[2]
[1]CESAR – Recife Center for Advanced Studies and Systems, Recife, Pernambuco, Brazil
[2]Federal University of Pernambuco Informatics Centre – CIn, Recife, Pernambuco, Brazil

Abstract – *In this paper, we present initial concepts of Internet of Things, whose technology combines Internet, sensors and smart objects; and Middleware, that is software that interconnects hardware and software in different layers. We analyzed several scientific publications covering many visions of Internet of Things using Middleware to integrate objects into different applications and networks. Using this knowledge base, we made summaries of the main articles, aiming to bring the main points addressed in each one, and performed a comparative analysis between them, highlighting their similarities and points of greatest relevance. In the end, we bring conclusions about the current state of the use of Middleware for IoT and the main challenges for combining IoT and marketing applications.*

Keywords: Middleware; Internet of Things; Comparative Analysis

1 Introduction

The Internet of Things (IoT) has become an increasingly constant topic in everyday conversations. It is a concept that not only has the potential to affect the way we live, but also the way we work.

The term "Internet of Things (IoT)" was created by Kevin Ashton at the end of the 90s, more precisely in 1998, in a presentation to executives of a major international brand that had the objective to gather the use of RFID's (Radio - Frequency IDentification) with the product supply network in retail markets and wholesale [1]. Then the Massachusetts Institute of Technology (MIT) presented their IoT vision in 2001 [2] and later, the Internet of Things was formally established by the International Telecommunication Union (ITU) by the ITU Internet Report in 2005 [3].

The concept of IoT it is defined as the connection of any electronic device to the Internet. This includes cell phones, coffee makers, washing machines, lamps, portable devices, refrigerators, and many others a multitude of devices built with sensors and connection capabilities. This also applies to machine components, e.g., a jet aircraft engine, an automobile press mill or an oil platform drill. The connection between these devices generate a significant amount of data, which in turn must be stored, processed and presented efficiently and easily interpretable way. [4]

When a large number of sensors are implemented, and the generation of data is initiated, the approach based on the traditional application (i.e., the connection of sensors directly to the applications individually and manually) becomes unfeasible. In order to address this inefficiency, significant amounts of Middleware solutions are reported by researchers. Each Middleware solution focuses on different aspects of the Internet of Things, such as device management, interoperability, platform portability, security and privacy, among others. Even so, some solutions address several aspects, and an ideal Middleware solution that addresses all aspects required by the IoT is yet to be developed [5].

This paper presents an analysis of a few researches that addresses Middleware applied to IoT, together with the understanding of the work done by different authors and draws a comparative analysis of the selected studies. Specifically in Section II, it presents an overview of the concepts of the Internet of Things (IoT) and in section III, about a review of the concepts of Middleware. Section IV provides a summary of selected articles and their comparative analysis of the articles are presented in Section V. Finally, Section VI presents the conclusion and final considerations.

2 Internet of Things (IoT)

Internet of Things (IoT) is the convergence between Internet and RFID, sensors and smart objects. Internet of Things can be defined as "things that belong to the Internet" for the supply and access to all real-world information. According to Gullemin and Friess [6], IoT allows people and things to be connected anytime, anywhere, with anything or anyone, preferably using any path and any service. Figure 1 illustrates this connection between the "things" related by Gullemin and Friess more easily.

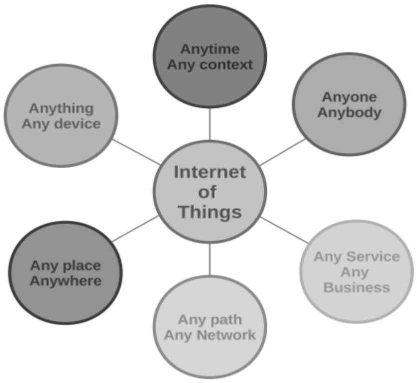


Fig. 1. Connection between the "things" according to the definition of IoT for Gullemin and Friess [6].

According to Atzori [7], Internet of Things can be accomplished in three paradigms: Middleware, sensors and Knowledge Base, which interact with each other and fulfill oriented visions of the Internet for Things. One paradigm will not meet all the connectivity vision. The intersection of these visions will be the main focus to make and connect objects on the network. For any network, the things will be active participants in business, information, and social processes.

The wireless sensor has brought technological advances in hardware domain for circuits and communications bringing effective and robust devices for sensing applications. This has led to diverse environments using wireless communication devices as described in Atzori [7]. Sensor data is collected and sent to a processing data centralized, distributed or hybrid processing module. Therefore, there are several challenges that a wireless sensor network has to face to develop a successful communication network of Internet of Things.

The Internet of Things, the interconnection and communication between everyday objects, allow many applications in several fields. Many of the attributes that can and should be considered when developing an application relate to network availability, bandwidth, coverage area, redundancy, user involvement and impact analysis. Some studies separate these areas in different applications, which are divided into two broad categories: first, Information and Analysis and in the second, Automation and Control [8]. In the following Tables 1 and 2, the six different applications are presented:

Table I. IoT Application Categories

Information and analysis		
Track behavior	*Enhanced situational awareness*	*Decision analysis guided by sensors*
- To monitor the behavior of people, things and data through space and time.	- Using the data infrastructure or environmental conditions to make real-time decisions.	- Help in making decisions about the use and analysis of data and visualization.

Table I. Information and Analysis Categories.

Table II. IoT Application Categories

Information and analysis		
Processes improvement	*Optimize resource consumption*	*Complex autonomous systems*
- Automatic control of industrial systems. - Continuos adjustment of factory lines.	- Consumption control to optimize the use of resources throughout the chain of generation of products, services, and natural resources.	- Automated in open environments. - Cleaning of hazardous materials through the use of robots

Table II. Information and Control Categories.

3 Middleware

The Middleware is a layer or set of software sub-layers interposed between the levels of application, operational and communication [9]. Its characteristic is to hide the details of different technologies, protocols, network environment, data replication, parallelisms, among others, it is essential to exempt the programmer about issues that are not directly relevant to its purpose, which is the development of the specific application. Furthermore, the Middleware masks the heterogeneity of computer architectures, operating systems, programming languages and network technologies to facilitate programming and application management.

The main features of a Middleware are:

- Hide the information distribution, which means hiding the fact that an application is usually composed of many interconnected parts running in distributed locations;

- Hide the heterogeneity of various hardware components, operating systems and communication protocols;

- Provide uniformly high level of standard interfaces for developers and application integrators, so that applications can be easily built, reused, ported and made to interact;

- Provide a set of common services to perform various functions general used in order to avoid efforts duplicated and facilitate the collaboration between applications [10].

There are different types of Middleware and the best knows are:

- Object Oriented: focuses on the receiver of the information and the introduction of reference tools to remote objects or proxies, to preserve the "look and feel". An advantage of this approach is that the coding can be used by anyone because the code will be done in the same way whether the system is distributed or not. An example of Middleware: CORBA.

- Services Oriented: similar to the object-oriented Middleware, except that there is less focus on the target object invoked and more emphasis on the operation to be performed. It's simpler to develop because the identity and life of remote objects do not need to be completely resolved. An example of Middleware: Thrift.

- Focused on data: this Middleware does not focus on the receiver, but in information propagated, i.e., the purpose and the meaning of what is transmitted. More effort is spent in ensuring the effective encapsulation of information and its transmission quality, controlling the physical aspects, time and bandwidth. Examples of Middleware are DDS and JMS implementations.

- Oriented message: this Middleware differs in its focus from all the previous ones, because either the receiver, the operation or the data are the communication paradigm, but the communication, with its physical aspects, such as size and time. The main idea of this Middleware is that the communication is not hidden or encapsulated, but exposed at least to the extent that allows it to be managed. When sending a message the receiver receives an identifier that allows you to track the progress of the delivery of the message. This is the identifier that is the focus on the delivery process. An example of Middleware: YAMI4 [11].

Middleware has gained more importance in recent years because of its role in simplifying the development of new services and integration of old and new technologies [12]. Businesses and organizations are increasingly integrating applications and systems, which they were independent, with new technology and development, building information systems for the entire company. This integration process involves legacy applications and outdated data bank. Many of these applications can only be used for its old interface and modifications are expensive or even prohibitive. The use of Middleware can connect this information within the entire company, various departments, and systems by placing them in a centralized environment with easier operation and maintenance. But with all this ease, there are many technical challenges for scale use of Middleware, such as interoperability, scalability, abstraction, spontaneous interaction between the "things", distributed infrastructure, security and privacy and a variety of types of Middleware [13].

4 Analysis of Relevant Articles

In this section, we summarize the most relevant articles from the point of view of citation numbers, year of publication and relevance for this research.

In the article "*Flexible IoT Middleware for Integration of Things and Applications*" [14], Bowman argues that while the Internet of Things is certainly a long way from becoming ubiquitous, it becomes increasingly closer to every day. And the future of the Internet of Things will consist of a variety of sensors connected to a network that will send the data to some types of cloud storage service, which will be available for all users, or authorized users. The authors also indicate that data should be shared between applications.

In this sense, the article summarizes that Internet of Things (IoT) must be supported by a Middleware that enables consumers and IoT developers to interact in a friendly manner, regardless of the different perspectives of use of IoT systems. The authors propose software that is bridge the gap between consumers and developers. The authors propose what they called the first step toward a ubiquitous Middleware for Internet of Things.

Charith Perera – 2014 in his article "*Context-Aware Computing for The Internet of Things: A Survey*" [5], it emphasizes that in our walk towards the Internet of Things (IoT), the number of sensors in operation around the world is growing at an accelerated rate. Market surveys have shown significant growth of new sensors in the last decade and predict an incremental growth rate in the future. These sensors have generated continuously, a huge amount of data. However, the use of such data is not trivial - it is necessary to understand them better. For Charith the use of context-aware computing is extremely important to the challenge of collecting, modeling and distribution of data generated.

The authors studied a subset of projects, representing research and commercial solutions proposed in the Context-Aware Computing area conducted over the last decade (2001-2011), based on a taxonomy presented by the researchers. The survey covers a wide range of techniques, methods, models, features, systems, applications, solutions and Middleware related to awareness of context and IoT. One of the article's focuses was to discuss the applicability of Computer Context-Aware applied to Internet of Things. The results obtained by researchers presented positive conclusions regarding the use of the Context-Aware Computing.

Zhen Peng – 2012 in his article titled "*Message Oriented Middleware Data Processing Model In Internet of Things*" [15] argues that with the development of Internet of Things, more and more devices are connected to the network. Devices have generated many different types of applications, however, they also bring new challenges for the maintenance and management of the network. One of the significant differences between the Internet and the Internet of Things pointed out by the authors, is that embedded devices represent an important category of devices on the Internet of Things, single function, poor performance and in large quantities. When divided by their functions, all devices are distributed in a monitored network and in data collection all the time.

The authors point out that even with the new challenges, the efficient data transmission and how to meet the needs of new applications that will arise. At work, the researchers detailed the data transmission characteristics of these applications. Moreover, the authors presented a new data processing model using a Message Oriented Middleware. With the new model Peng points out that transmission and data processing will be more convenient, efficient, easy to share and secure.

Hiro Gabriel Cerqueira Ferreira – 2014 in his article "*Proposal of a Secure, Deployable and Transparent Middleware for Internet of Things*" [16] proposes a security architecture for Middleware for IoT, focused on bringing real-life objects into the virtual world, the architecture proposed by the author is implementable and includes protective measures based on existing technologies for security such as AES, TLS and OAuth . He further says that privacy, authenticity, integrity and confidentiality of the data exchange services are integrated to provide security to smart objects generated for users and services involved, reliably and implementable.

In a previous [17] work, Hiro proposed a Middleware for IoT and detailed usage scenarios, while in the above article he specifies the security architecture based on existing technologies. The Middleware resulting of Hiro's article allows the implementation of the Internet of Things environments. In relation to the specific security aspects, the article details how communication and commitment must be to provide privacy, authenticity, integrity and confidentiality of users, applications, and devices. The model proposed by the author analyzes and resolves questions about scalability, a variety of devices and applications, simplicity, robustness and it keeps the possible interactions with low computational resources and energy entities. In his conclusion, the author draws attention to future works, including the creation of wrapper of existing devices that already communicate using other technologies. These devices should be placed under the master domain controller in order to become available under Middleware API. It addresses the issue to be resolved is the mobility for end devices and identification and how to automatically carry your settings and permissions to other controllers.

Jayavardhana Gubbi - 2013 in his article "*Internet of Things (IoT): A vision, architectural elements, and future directions*" [18] addresses that ubiquitous sensing enabled by Wireless Sensor Network (WSN) technologies are presented in many areas of modern life. He claims that it offers the ability to measure, infer and understand environmental indicators, ecologies and natural resources in urban environments. The proliferation of these devices in this communication network sets the creation of the Internet of Things, in which sensors and actuators combine perfectly with the environment around us, and the information is shared across platforms, in order to develop a common operational picture (COP).

Fueled by recent adaptation of a variety of enabling wireless technologies such as RFID, sensors and actuators embedded, IoT has gone reality and it is the next revolutionary technology to transform the Internet into a fully integrated Internet. He also states that as we move from the Web page (static web pages) to web2 (web social networking) to web3 (web ubiquitous computing), the need for data on demand using sophisticated intuitive queries increases significantly. The article also presents a vision-centered in cloud computing for the implementation of the Internet of Things.

Gubbi's article resulted in the proposal of a cloud-based model, centered on the user, allowing the flexibility to meet the diverse needs and conflicts from different sectors. The authors proposed a framework capable of using the Internet of Things. The structure allows that network themes, computing, storage and visualization allow separated independent growth in all sectors, but complementing each other in a shared environment.

In the article "*Middleware for IoT-Cloud Integration Across Application Domains*" [19], Chengjia Huo presents a Middleware capable of connecting to the Internet of Things to the Clouds to provide a more robust security. The implementation shown in the article is a proof of concept of this approach. It is an example that shows how to integrate embedded systems and cloud computing. In the article, the author proposed IoT architecture with the inclusion of what he called "Rimware", a Middleware that is integrated with the cloud and the gateway. The authors assumed that the data service for a particular application would already be available in the cloud as well as configuration and command devices belonging to authenticated users via the gateway. In the study, the gateway consisted of an application running on a smartphone that performs user authentication and connects to Bluetooth devices. They identified some problems with this approach, as limits iteration of users and applications. With this in mind, they proposed a plugin allowing multiple gateways can be used at the same time.

In their work the authors had as a result, the concept of rimware for IoT integration to the clouds to form a powerful cyber-physical system. This Middleware, called BlueRim is scalable globally. They assert that the effectiveness of the characteristics contained in the Middleware proposed has been validated in real applications with different types of access.

5 Comparative Analysis Between Articles

In this section, we will raise aspects that were considered relevant to the thematic context and applicability in the real context.

The result of comparative analysis these articles reveal that the integration between objects is still the great challenge since the IoT conception. In near future people and objects relate autonomously and directly. To achieve this, researches are directed to the integration of objects through brokers or Middlewares, where the location and integration of objects must be automatic. Because the scale is not feasible there is human intervention, this has to be a process, at least semi-automated or automated, to connect them. Another important challenge for this theme will be the creation of a pattern of communication between objects where in this pattern will be possible an automatic analysis of objects and their integration [5].

Overall, the articles that address this study attack the problem at different bias. The majority of studies propose a Middleware architecture [12] [17] [20] [22] [23] [24] [25] [26] [27] [34] to integrate the "things" with networks and systems, other devices, applications and social networks, and other articles, try to address the security issue [21] [30] [31] [32], which is one aspects of most concern because it involves issues of privacy, confidentiality and data integrity throughout the IoT architecture; and in this aspect, the solution usually is pointed out in the cloud computing [19] [21] - Middleware integrated to cloud computing, where it is possible to increase the demand for network, hardware, storage and systems to meet growing number of IoT applications. The results show that softwares and applications will be essential in developing solutions for IoT, in the coming years. As the developers plan to interconnect objects with multiple systems, including mobile applications, desktop, database, cloud services, enterprise applications, Middleware and other devices to the Internet of Things.

6 Conclusion and Future Works

The current state of the art of Middleware for the Internet of Things explores different approaches to supporting for some of the features provided in IoT. In our study, we found enough research evidence related to information security in this area. It is a major concern of researchers to develop a usable architecture, integrated into the clouds to provide certain abstractions like security and scalability. In most of the researches, it has a marketing nature, and the authors of the studies analyzed are indicating future works in one-way: the handling and processing of the data generated are and will remain a major challenge for the IoT to establish fully.

Although some of the articles we have covered in this study did not deal directly of Middlewares, we feel favorable the analysis since the topics discussed other solutions to answer the challenges to adoption of the Internet of Things.

For future researches, there is a great deficiency to find out a stable architectural solution that they are able to meet the demand for data, keeping them private, fair and accessible. Based on our analysis, perhaps the most appropriate way to go, it is to an architecture that integrates Middleware and cloud computing in order to use existing services, facilitating the devices integration in the worldwide network of things.

7 Acknowledgments

Authors would like to acknowledge the support from CESAR in the development of this research.

8 References

[1] K. Ashton, That ''Internet of Things'' thing, RFiD Journal, 2009.

[2] D. L. Brock, "The electronic product code (epc) a naming scheme for physical objects," Auto-ID Center, White Paper, January 2001, http://www.autoidlabs.org/uploads/media/MIT-AUTOID-WH-002.pdf [Accessed on: 2015-07-1].

[3] International Telecommunication Union, "Itu internet reports 2005: The internet of things," International Telecommunication Union, Workshop Report, November 2005, Available at https://www.itu.int/osg/spu/publications/internetofthings/InternetofThings_summary.pdf [Accessed on: 2015-07-01]

[4] D. Giusto, A. Iera, G. Morabito, L. Atzori (Eds.), The Internet of Things, Springer, 2010. ISBN: 978-1-4419-1673-0

[5] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context-aware computing for the internet of things: A survey," IEEE Commun. Surv. Tutorials, vol. 16, no. 1, pp. 414–454, 2014.

[6] P. Guillemin and P. Friess, "Internet of things strategic research roadmap," The Cluster of European Research Projects, Tech. Rep., September 2009, Available at http://www.researchgate.net/publication/267566519_Internet_of_Things_Strategic_Research_Roadmap [Accessed on: 2015-07- 01].

[7] L. Atzori, A. Iera, G. Morabito, The Internet of Things: A Survey,Computer networks, 54(16), pp. 2787–2805, 2010.

[8] C. Michael, L. Markus, R. Roger. "The internet of things." Fobers Magazine, 2002.

[9] IGILL, C. D.; SMART, W. D. Middleware for robots? In: AAAI SPRING SYMPOSIUM ON INTELLIGENT DISTRIBUTED AND EMBEDDED SYSTEMS. Stanford. Proceedings... Stanford: 2002.

[10] K. Sacha, "Introduction to Middleware", 2003. Available at http://Middleware.objectweb.org/index.html. [Accessed on: 2015-07- 01].

[11] Available at http://www.inspirel.com/articles/Types_Of_ Middleware.html. [Accessed on: 2015-07- 01].

[12] S. De Deugd, R. Carroll, K. Kelly, B. Millett, J. Ricker, SODA: Service Oriented Device Architecture, IEEE Pervasive Computing 5, pp. 94–96, 2006.

[13] Chaqfeh, Moumena a. Mohamed, Nader, Proceedings of the 2012 International Conference on Collaboration Technologies and Systems, CTS 2012.

[14] J. Boman, J. Taylor, and A. Ngu, "Flexible IoT Middleware for Integration of Things and Applications," Proc. 10th IEEE Int. Conf. Collab. Comput. Networking, Appl. Work., no. CollaborateCom, pp. 481–488, 2014.

[15] Z. Peng, Z. Jingling, and L. Qing, "Message-oriented Middleware data processing model in Internet of things," Proc. 2012 2nd Int. Conf. Comput. Sci. Netw. Technol., pp. 94–97, 2012.

[16] H. Gabriel, C. Ferreira, R. Timóteo, D. S. Júnior, F. Elias, and G. De Deus, "Proposal of a Secure, Deployable and Transparent Middleware for Internet of Things," Information Systems and Technologies (CISTI), 2014.

[17] Ferreira, H. G. C.; Canedo, E. D.; Sousa Júnior, R. T. "IoT Architecture to Enable Intercommunication Through REST API and UPnP Using IP, ZigBee and Arduino." 2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp.53,60, 2013.

[18] Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. Internet of things: Vision, applications and research challenges. Ad Hoc Networks, 10(7), 1497–1516, 2012.

[19] Huo, C., Chien, T. C., & Chou, P. H. (2014). "Middleware for IoT-cloud integration across application domains." IEEE Design and Test, 31(3), 21–31, 2014.

[20] L. Atzori, A. Iera, G. Morabito, and M. Nitti, "The social internet of things (SIoT) - When social networks meet the internet of things: Concept, architecture and network characterization," Comput. Networks, vol. 56, no. 16, pp. 3594–3608, 2012.[12] J.

[21] Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," Futur. Gener. Comput. Syst., vol. 29, no. 7, pp. 1645–1660, 2013.

[22] S. Gusmeroli, S. Piccione, and D. Rotondi, "IoT@Work automation Middleware system design and architecture," IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA, 2012.

[23] P. Maló, B. Almeida, R. Melo, K. Kalaboukas, and P. Cousin, "Self-organised Middleware architecture for the internet-of-things," Proc. - 2013 IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput. GreenCom-iThings-CPSCom 2013, pp. 445–451, 2013.

[24] V. H. Rocha, F. S. Ferraz, H. N. De Souza, and C. A. G. Ferraz, "ME-DiTV : A Middleware Extension for Digital TV An Architectural Proposal of A Middleware Extension based on Dynamic Context Changes for Distributed System." The Seventh International Conference on Software Engineering Advances (ICSEA), 2012.

[25] W. Wang, K. Lee, and D. Murray, "Building a generic architecture for the Internet of Things," Proc. 2013 IEEE 8th Int. Conf. Intell. Sensors, Sens. Networks Inf. Process. Sens. Futur. ISSNIP 2013, vol. 1, pp. 333–338, 2013.

[26] B. Guo, D. Zhang, Z. Wang, Z. Yu, and X. Zhou, "Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things," J. Netw. Comput. Appl., vol. 36, no. 6, pp. 1531–1539, 2013.

[27] S. Kang, Y. Lee, S. Ihm, S. Park, S. M. Kim, and J. Song, "Design and implementation of a Middleware for development and provision of stream-based services," Proc. - Int. Comput. Softw. Appl. Conf., pp. 92–100, 2010.

[28] J. Al-Jaroodi and N. Mohamed, "Service-oriented Middleware: A survey," J. Netw. Comput. Appl., vol. 35, no. 1, pp. 211–220, 2012.

[29] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," Comput. Networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[30] S. Unger, S. Pfeiffer, and D. Timmermann, "How much security for switching a light bulb the SOA way," IWCMC 2012 - 8th Int. Wirel. Commun. Mob. Comput. Conf., pp. 1034–1039, 2012.

[31] T. Xu, J. B. Wendt, and M. Potkonjak, "Security of IoT Systems : Design Challenges and Opportunities," pp. 417–423, 2014.

[32] K. Geihs, "Middleware challenges ahead," Computer (Long. Beach. Calif)., vol. 34, no. 6, pp. 24–31, 2001.

[33] P. Guillemin and P. Friess, "Internet of things strategic research roadmap," The Cluster of European Research Projects, Tech. Rep., September 2009, http://www.internet-of-things-research.eu/pdf/IoT Cluster Strategic Research Agenda 2009.pdf

[34] Y. Hong, "A resource-oriented Middleware framework for heterogeneous internet of things," Proc. 2012 Int. Conf. Cloud Comput. Serv. Comput. CSC 2012, pp. 12–16, 2012.

IoT is SoS

Parisa Mahya[1], **Hooman Tahayori**[2]
[1]Dept. of Computer Science and Information Technology, International Division, Shiraz University, Shiraz, Iran
[2]Dept. of Computer Science and Information Technology, School of Electrical and Computer Engineering, Shiraz, Iran

Abstract- *The rise and evolution of the Internet and complex systems in recent decades has led to the emergence of new technologies such as Internet of Things and System of Systems. The focus of this paper is to study the relationship between Internet of Things (IoT) and System of System (SoS) and will discuss that IoT can be considered as a subcategory of SoS. Since system of system is designed to analyze and solve problems in complex and large systems and facilitates decision making in the related situations. IoT-as a subcategory of SoS- would benefit SoS toward a systematic growth.*

.Keywords: Internet of Things (IoT), System of Systems (SoS), e- Health, complex systems, emergent behavior.

1 Introduction

System is an omnipresent concept in daily life that usually refers to a group or complex of parts (such as people, machine, etc.) interrelated in their actions toward fulfilling some goals [1].

A system may itself be a part of a larger system, called complex system. Different complex systems can be identified in various areas like aerospace, manufacturing, industry, environmental systems, etc. [2]. The presence and importance of complex systems in recent decades, has led to the establishment of new research threads and technologies like "System of Systems" (SoS). Among various definitions that are proposed for SoS, a practical one says that SoS is a super system comprised of other elements which themselves are independent complex operational systems that interact among themselves to achieve higher goals [2].

Internet of Things (IoT) is another well-known technology that is related to SoS. To clarify the relation between the two mentioned technologies, imagine a collection of sensors that are wearable i.e. fitness wearable that has widespread applications in security, healthcare and sports. Fitness wearable as an example is an independent system that allows people monitor their – vital - signs when doing any kind of exercises. If it connects to a Wi-Fi or smart phone, vital signs would be saved in a cloud or shared with healthcare centers. Connecting to a comprehensive health care system enables achieving better insight into the people's health condition and enables providing services like heart attack alerts. As seen with fitness wearable sometimes a complex system can be a part of a larger network or be an object in the environment. The new problem that arises though is related to the objects and how to enable them act intelligently and share observations in real world. As a result, a novel paradigm called Internet of Things (IoT) has been introduced. The idea of IoT is based on the pervasive presence of objects around us that are able to interact with each other (via smart sensors, RFIDs, etc.) to achieve common goals. Existing technologies such as RFID tagging, has posed IoT in variety of fields and areas such as healthcare, transportation, smart home, etc. Despite the fact that applications of IoT are among researchers and organizations, yet IoT is not well-established.

In this paper we demonstrate that IoT is a subcategory of SoS by providing evidences, examining cases and providing comprehensive overview of IoT and SoS. Addressing this issue would be helpful from different aspects. First, the scopes of IoT and SoS are determined and second, one's application can be extended to the other. The paper is organized as follows. Section 2, reviews definitions and characteristics of SoS. Section 3, explains various definitions of IoT and section 4, is devoted to the relation between IoT and SoS. Section 5 concludes the paper.

2 System of Systems Definitions and Characteristics

In order to be able to contrast and compare SoS and IoT in this section we will study different definitions and characteristics of SoS. However, next section is dedicated to IoT. As it is mentioned earlier, many definitions are proposed for SoS. In the following, we will review some of these definitions.

Definition 1: System of Systems integration is a method to pursue development, integration, interoperability and optimization of systems to enhance performance in future battlefield scenarios [2].

Definition 2: Systems of Systems are large-scale concurrent and distributed systems that are comprised of complex systems [2][3].

Definition 3: Enterprise System of System Engineering(SoSE) is focused on coupling traditional systems engineering activities of strategy planning and investment analysis.

Definition 4: System of System is a multiple, independent systems that interact for the purpose of a global goal [1][4].

Definition 5: SoS is defined as a set or arrangement of systems that results when independent and useful systems are integrated into a larger system that delivers unique capabilities [5][6].

Definition 6: SoS [is a] a collection of trans- domain networks of heterogeneous systems that are likely to exhibit operational and managerial independence, geographical distribution, and emergent and evolutionary behaviors that would not be apparent if the systems and their interactions are modeled separately [5][7].

Definition 7: Systems of systems are large-scale concurrent and distributed systems the components of which are complex systems themselves. [5].

As mentioned, there is no unique definition for SoS that all researchers agreed on. However, to distinguish a normal system from SoS, several characteristics are identified for SoS.

In general, there are five common characteristics known as Maier's criteria: operational and managerial independence, geographical distribution, emergent behavior, evolutionary development [5][2]. We will discuss them in the following.

Operational independence indicates independency and usefulness of subsystems. Managerial independence indicates that "a system is able to operate independently and is operating independently." [2][5]

Geographical distribution refers to the ability of systems to communicate and exchange information in any local scale; in other words, a system in any geographical location should be able to communicate with other systems [5].

In emergent behavior, the behavior of system is based on the relationship among the parts- than the functionality of individual parts. A collection of complex system's properties is called emergent property that is not available for an individual system [5].

Evolutionary behavior indicates that SoS should evolve over time based on the modification and changes.

3 Internet of Things: Definitions

IoT focuses on objects and entities in environment and their relations with each other.

The very first idea of IoT was presented in 1982 [8], however Kevin Ashton in 1999 in the context of supply management first used the term "IoT". The main motivation of IoT was explained by Ashton[8]:

"Today computers-and, therefore, the Internet-are almost wholly dependent on human beings for information ... The problem is, people have limited time, attention and accuracy... We need to empower computers with their own means of gathering information, so they can see, hear and smell the world for themselves..."

In 2001 Auto-ID center in MIT represented a viewpoint of IoT [8] as follows.

"The Electronic Product Code (EPC) was conceived as a means to identify all physical objects. The primary purpose of the EPC was to serve as a reference to networked information. Used in conjunction with Object Name Service, the EPC associates the physical object with information about the object. Together, these components allow physical objects to be networked, creating essentially an "Internet of Things"" [9].

In 2005, IoT was officially recognized as a field in International Communication Union [8]. The first conference on IoT was held in Zurich in 2008 [8]. A year later, in 2009, the government of China, supported the idea of sensing China and consequently the city of Wuxi was determined as pioneering center in IoT. In the same year, IoT European Research Cluster (IERC) proposed an IoT research roadmap till 2020 and a year later, published a comprehensive document for IoT challenges and perspectives [10].

IoT is changing rapidly with the aim of creating an integrated Internet that evolves people's lives, thoughts and works.

However, like SoS, there is no comprehensive definition for IoT.

IEEE, has defined IoT as: *A network of items – each embedded with sensors – which are connected to the internet*." [11]

ETSI[1] - , a standardization organization in the telecommunication industry, has introduced Machine-to-Machine (M2M) concept which is similar to IoT:

"M2M is a communication between 2 or more entities that do not necessarily need any direct human intervention. It services intend to automate decision and communication processes." [10]

Another definition is presented by ITU[2], the United Nations specialized agency for information and communication technologies. ITU uses the phrase "ubiquitous network" while defining IoT, which means the availability of networks everywhere and anytime [10].

[1]. European Telecommunication Standard Institute

[2]. International Telecommunication Union

OASIS[3], a nonprofit, international consortium defines IoT as a system where the Internet is connected to the physical worlds via ubiquitous sensors.

4 Internet of Things vs. System of Systems

We described the main features and characteristics of SOS and reviewed the definitions of IoT respectively in the sections 2 and 3.

In this section, we will compare IoT and SoS and will define IoT as a subcategory of SoS.

Elaborating on the definitions presented in section 2, and considering the five characteristics enumerated, SoS can be considered as independent systems that are interoperating with each other with common goals.

IoT, however, consists of two parts; Internet and things. According to the IoT definitions, objects should have a minimum intelligence. In other words, they are not intelligent by themselves and they cannot think and make decisions, however, technologies such as sensors and RFIDs provide intelligence for objects by sensing and gathering data from environment. These technologies make things think, enable them to solve their own problems and decide. Hence, objects with sensors are independent systems.

Based on the discussions in the sections 2 and 3, a collection of objects equipped with sensors interact in complex manner with each other or they can integrate with other complex systems. Furthermore, data from remote sensors in an IoT-system can be integrated into decision-making support systems, power grids, telecommunication networks and clouds and construct a complex large-scale system [12][13]. Like any system, such a system has to be managed and monitored from different aspects like security, reliability, etc. which are not easy through IoT it can be considered as SoS to help properly in different aspects from SoS perspectives. [14].

In order to map IoT concepts to SoS definitions, key characteristics of SoS which distinguish a regular system from SoS should be satisfied. In the following, two cases on transportation and healthcare are examined to better understand IoT applications and their relevance to SoS characteristics.

Case 1: Nowadays, in most of the cities, different modes of public transportation like metro, bus, etc. are available that are aimed to ease people's life while, at the same time vehicles have destructive effects on the environment. Studies have revealed that air pollution causes by vehicles are as dangerous as traffic accidents [15]. In this case, IoT helps creating a smart environment, smart transportation system and smart cities with the goals of reducing pollutions and controlling traffic. In order to implement IoT in transportation, vehicles, roads and traffic lights should be equipped with sensors. Vehicles and devices are systems themselves that operate independently. For example, a bus operation is based on its schedule, however it should cooperate with other buses and vehicles or even connects to transportation data centers and clouds toward controlling the traffic. Table-1 contrasts this IoT case with SoS characteristics.

Table-1- Smart Transportation as a SoS.

SoS Characteristics	IoT Smart Transportation
Managerially and operationally independent	Vehicles such as metro or bus are independent systems and they operate independently helping people to commute.
Geographical distribution	Metros and buses in any geographical location and distances can share their information.
Emergent behavior	Targets like traffic control is possible through the collaboration of all vehicles such as metros, buses, etc.
Evolutionary behavior	Measurements in transportation such as traffic status and number of vehicles are dynamic and based on these information and measurement, SoS evolves over time.

As is shown in table-1, in smart transportation, devices are managerially and operationally independent. Devices can exchange information via GPS regardless of their location, so geographical distribution characteristic is supported in this case.

Undoubtedly reaching the ultimate goal of IoT is only possible through the cooperation of all devices. Each individual system cannot achieve the goal associated with IoT, in other words, minimizing pollution as a goal is an emergent property of this system. Ref. [15] has proposed a system for transportation based on IoT that utilizes information to learn from experiences and evolve over time.

Case 2: People's health status has always been noteworthy and as lifestyle has changed, healthcare has become more important. Consequently, many successful studies and researches have been done in this field and one of the improvements is known as e-Health. Many definitions have been specified for e-Health and some of them are related to the use of Internet in healthcare and

[3]. Organization for the Advancement of Structured Information Standard

many attempts have been made in this direction [16][17]. Ref. [18] has proposed an e-Health framework based on IoT with the focus on increasing quality of life especially for people with chronic diseases and emergency situation. In this paper, the author illustrates a day of a diabetic person. Based on this scenario, low-cost medical devices equipped with sensors, communicates wirelessly, given a connection be available. They monitor the person's health status, since a diabetic person needs special cares. In emergency situation, medical centers are alerted and will send services to the patients as soon as possible.

In this case, devices for general health monitoring work independently, measuring blood pressure and monitoring other vital signs, so they are managerially and operationally independent. With the help of communicators, it is possible to share data with doctors, health care centers in any geographical distribution – which coincides with the geographical distribution characteristic in SoS. Communications among devices, equipment and medical centers are the only way of achieving the aim of this platform which follows the emergent behavior characteristics of SoS. Decisions about person's health status are made based on information in databases and new data can be added in the database that causes gradual changes and improvements in decisions – which is evolutionary behavior characteristics. Table-2 contrasts this IoT case with SoS characteristics.

Table-2- e-health as a SoS

SoS Characteristics	IoT e-health
Managerially and operationally independent	General health monitoring devices are independent. They work independently with the goal of measuring blood pressure, monitoring vital signs, etc.
Geographical distribution	Healthcare equipment in any geographical location can submit their data to data centers and share it with others.
Emergent behavior	Goals like monitoring people's health is possible through the collaboration of different healthcare equipment and an independent system alone cannot achieve this goal.
Evolutionary behavior	Information gathered from measurements such as temperature, blood pressure are added to data centers gradually. Due to these changes, the system evolves over time.

As we mentioned above, objects in IoT are independent and they operate independently. Furthermore, IoT is based on the Internet and nowadays the Internet is public, cooperative facility accessible to millions of people worldwide and is evolving over the years. So it is easy to connect objects everywhere, anytime over the Internet. Also, as IETF has mentioned while defining IoT, objects cooperate with each other in order to make an accessible service from anywhere and anytime.

An essential concept in IoT is system thinking; which means that IoT as a complex system should be able to think and sensors obtain such ability. But the presence of sensors is not enough for a system to think and there should be some units for processing data and learning. IoT on the other hand has four basic hardware units namely: sensors/actuators, processing unit, storage unit and communication unit. Data collection, data processing, decision making and learning are possible via processing unit, storage unit and two others. Moreover, an operating system like TinyOs is designed for IoT to fulfill requirements such as reliability and high-level programming [10].

The forth characteristics of SoS, i.e. emergent behavior, is satisfied in IoT since the heart of system thinking is that the behavior of a system is an emergent property of its structure [19]. In the previous cases, goals such as controlling traffic is only possible through interaction and cooperation among appliances and this is correspondent with emergent properties definitions.

In IoT, various learning algorithms are designed and implemented and the most important feature of these algorithms is their adoption with environment over time. This depicts that IoT supports evolutionary development characteristic. An example of these characteristics have studied in [10] in which the proposed system adjusts itself to new behavior after a while.

As a result, handling a large-scale IoT system is difficult because the barriers of system are immense. Considering these kinds of system as SoS is a promising solution since SoS is specified to deal with complex systems and it can offer appropriate solutions for problems in complex systems. We expect this paper be helpful in designing large-scale IoT system based on SoS definitions, characteristics and applications

5 Conclusions

System of System (SoS) and internet of Things (IoT) are emergent new technologies. SoS helps modeling and dealing with complex systems and IoT creates a broad network in which objects are able to think, decide and operate. In this paper, IoT is studied as a subcategory of SoS based on the definitions and applications. This assortment is beneficial from different aspects. First, the applications and characteristics of IoT can be viewed from the SoS window. Also, in systems with great complexities especially in industry and government, decision-making, analyzing and modeling are difficult. In such cases, SoS approach is a good solution for handling and managing complex systems and additionally it is possible to use this approach in IoT cases.

6 References

[1] W. C. Baldwin and B. Sauser, "Modeling the Characteristics of System of Systems," in *System of System Engineering*, pp. 1–6, 2009.

[2] M. Jamshidi, "system of System Engineering," vol. Mandarin V, no. China Machine Press, 2013.

[3] P. Carlock and J. A. Lane, "System of Systems Enterprise Systems Engineering, the Enterprise Architecture Management Framework, and System of Systems Cost Estimation," in *SoS ESE, EAMF, and Cost Estimation*, pp. 1–12, 2006.

[4] W. Crossley, "System of systems: An introduction of Purdue University schools of engineering's signature area," in *Proceedings of the Engineering Systems Symposium*, 2004.

[5] D. Firesmith, "Profiling Systems Using the Defining Characteristics of Systems of Systems (SoS)," pp. 1–74, 2010.

[6] C. Keating, R. Rogers, R. Unal, D. Dryer, A. Sousa-Poza, R. Safford, W. Peterson, and G. Rabadi, "Systems Engineering Guide for Systems of Systems," *IEEE Engineering Management Review*, vol. 36, no. 4. 2008.

[7] D. DeLaurentis and R. K. Callaway, "A system-of-systems perspective for public policy decisions," *Rev. Policy Res.*, vol. 21, no. 6, pp. 829–837, 2004.

[8] Q. Wu, G. Ding, Y. Xu, S. Feng, Z. Du, J. Wang, and K. Long, "Cognitive Internet of Things: A New Paradigm Beyond Connection," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 129–143, 2014.

[9] D. L. Brock, "white paper The Compact Electronic Product Code A 64-bit Representation of the Electronic Product Code," in *Mit Auto-Id Center Massachusetts*, pp. 1–12, 2002.

[10] D. Singh, G. Tripathi, and A. J. Jara, "A survey of Internet-of-Things: Future vision, architecture, challenges and services," *2014 IEEE World Forum Internet Things, WF-IoT 2014*, pp. 287–292, 2014.

[11] R. Minerva, A. Biru, and D. Rotondi, "Towards a definition of the Internet of Things (IoT)," *IEEE Internet Things*, pp. 1–86, 2015.

[12] M. M. Rana and L. Li, "Kalman Filter Based Microgrid State Estimation Using the Internet of Things Communication Network," in *2015 12th International Conference on Information Technology - New Generations*, pp. 501–505, 2015.

[13] V. G. T. N. Vidanagama, D. Arai, and T. Ogishi, "Service environment for smart wireless devices: An M2M gateway selection scheme," *IEEE Access*, vol. 3, pp. 666–677, 2015.

[14] P. Maia, E. Cavalcante, P. Gomes, T. Batista, F. C. Delicato, and P. F. Pires, "On the Development of Systems-of-Systems based on the Internet of Things," in *Proceedings of the 2014 European Conference on Software Architecture Workshops - ECSAW '14*, pp. 1–8, 2007.

[15] D. Kyriazis, T. Varvarigou, D. White, A. Rossi, and J. Cooper, "Sustainable smart city IoT applications: Heat and electricity management & Eco-conscious cruise control for public transportation," in *2013 IEEE 14th International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM*, 2013.

[16] T. Gong, H. Huang, P. Li, K. Zhang, and H. Jiang, "A Medical Healthcare System for Privacy Protection Based on IoT," in *2015 Seventh International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)*, pp. 217–222, 2015.

[17] F. Fernandez and G. C. Pallis, "Opportunities and challenges of the Internet of Things for healthcare: Systems engineering perspective," in *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare - "Transforming Healthcare Through Innovations in Mobile and Wireless Technologies", MOBIHEALTH 2014*, pp. 263–266, 2015.

[18] N. Bui and M. Zorzi, "Health care applications: A Solution Based on The Internet of Things," in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies - ISABEL '11*, pp. 1–5, 2011.

[19] "Emergent Behavior - Tool_Concept_Definition." [Online]. Available: http://www.thwink.org/sustain/glossary/EmergentBehavior.htm.

A Smart Pipe Indicator Management Method under IoT Environments

KeeHyun Park[1], In Sung Kim[1]
[1] Computer Engineering Dept., Keimyung University, Rep. of Korea
{khp, epqlfqmffn}@kmu,ac,kr

Abstract. This paper proposes an underground utility management scheme using smart pipe indicators (SPIs) based on the IoT technologies. A SPI is a pipe indicator installed on the ground right above underground gas pipes to indicate the type and the management authority of pipes, burial depth and so on. Efficient management of underground gas pipes is very important because unplanned ground excavation without exact knowledge about the underground environment could cause an enormous gas explosion and possible death and injury. The oneM2M communication protocol, one of the standard IoT communication protocols, is used to manage the SPIs. Smartphones act as Application Entities in the oneM2M based IoT system in order to sense the data stored on SPIs. A SPI management server acts as an Infrastructure Node to store the information of SPIs. Gateways, located between the smartphones and the server for communication efficiency, act as a Middle Node.

Keywords: Smart Pipe Indicator, oneM2M, IoT

Type of Submission: Short research paper

1 Introduction

Many utilities are laid underground every year and the number of underground utilities will continue to increase [1-3]. Gas, water and electricity are examples of such underground utilities. Efficient management of underground utilities has become one of the more serious problems that we face because underground damage to or accidents at such utilities could lead to enormous disasters.

The IoT (Internet of Things) technologies [4-7], which makes objects connected with each other, have become increasingly popular. The objects can be sensors, home equipment, meters, and other similar devices. The IoT technologies can be applied in many areas, including remote metering, traffic control, home appliances, and healthcare. Application areas that utilize the concept of IoT can be broadened to the management of underground utilities. Some communication protocols including oneM2M [7-10] was proposed as standards for IoT environments.

This paper proposes an underground utility management scheme using smart pipe indicators (SPIs) based on the IoT technologies. A SPI is a pipe indicator installed on the ground right above underground gas pipes to indicate the type and the management authority of pipes, burial depth and so on. Efficient management of underground gas pipes is very important because unplanned ground excavation without exact knowledge about the underground environment could cause an enormous gas explosion and possible death and injury. The oneM2M communication protocol, one of the standard IoT communication protocols, is used to manage the SPIs under IoT environments. Smartphones act as Application Entities (AEs) in the oneM2M based IoT system in order to sense the data stored on SPIs. A SPI management server acts as an Infrastructure Node (IN) to store the information of SPIs. Gateways, located between the smartphones and the server for communication efficiency, act as a Middle Node (MN).

The remainder of this paper is organized as follows. Section 2 describes some related studies, and Section 3 explains the structure of the oneM2M-based IoT system for SPI management proposed in this study. Finally, Section 4 draws some conclusions and discusses some possible directions for future research.

2 Related studies

2.1 oneM2M-based IoT system

The oneM2M communication protocol is an international standard for the IoT system. Figure 1 shows the structure of the oneM2M-based IoT system [4-7]. In an IoT system, a (program installed on) a sensor or device represents an Application Dedicated Node – Application Entity (ADN-AE) that gathers surrounding data and transmits it to the system's Middle Node – Common Service Entity (MN-CSE). An MN-CSE controls or monitors ADN-AEs that belong to the MN-CSE; moreover, it performs processing that is necessary to achieve efficient communication between ADN-AEs and the Infrastructure Node – Common Service Entity (IN-CSE). A manager or a user can access data stored in IN-CSE via ADN-AE.

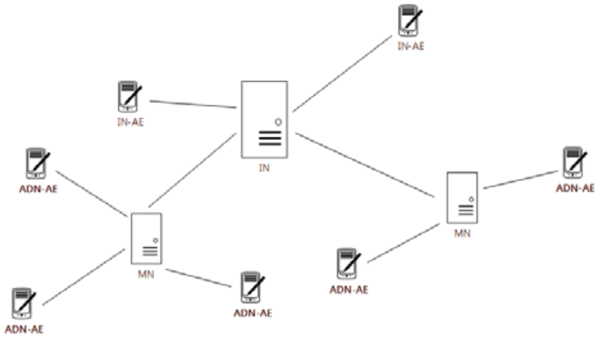


Figure 1 oneM2M-based IoT system structure

3 Structure of a oneM2M-based IoT system for SPIs

3.1 SPI (Smart Pipe Indicator)

Figure 2 shows a SPI (Smart Pipe Indicator) used in this study. SPIs are installed on the ground right below buried gas pipes. Smartphones can be read the data about underground gas pipes using NFC communication. From a SPI, the following data can be read by using NFC.

- Type/size/direction of the gas pipe
- Management ID and Management authority
- Burial depth
- Emergency contact numbers

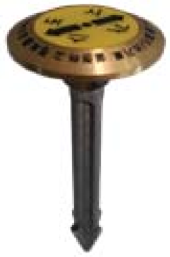


Figure 2 Smart Pipe Indicator

3.2 System structure

Figure 3 shows the structure of the proposed oneM2M-based IoT system for SPIs. SPIs have the data about gas pipes buried right below the SPIs. The data can be read by a smartphone using NFC (Near Field Communication) protocol [11]. Smartphones are acts as Application Entities (AEs) in the oneM2M based IoT system in order to sense the data stored on SPIs. A SPI management server acts as an Infrastructure Node (IN) to store the information of SPIs. Gateways, located between the smartphones and the server for communication efficiency, acts as Middle Node (MN). The dotted circle in the figure represents the oneM2M based IoT system and therefore the oneM2M communication protocol is used in the circle.

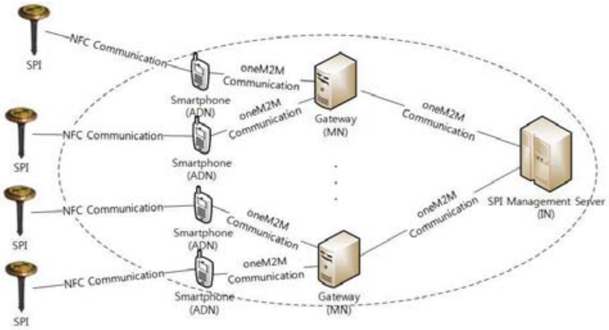


Figure 3 Structure of the oneM2M based IoT system for SPIs

Table 1 shows the hardware specifications used for the SPI management system proposed in this study.

Table 1 Hardware specification for the SPI management system

	ADN	MN	IN
CPU	Exynos 7420	Intel Core i5-650(3.2GHz)	Intel Core i7-4770(3.4GHz)
Main Memory	3GB	4GB	8GB
HDD	UFS 2.0	SSD	SSD
Operating System	Android	Windows 7	Windows 7

4 Conclusion

This paper proposes an underground utility management scheme using SPIs based on the IoT technologies. The oneM2M communication protocol, one of the standard IoT communication protocols, is used to manage the SPIs under IoT environments. Smartphones act as Application Entities (AEs) in the oneM2M based IoT system in order to sense the data stored on SPIs. A SPI management server acts as an Infrastructure Node (IN) to store the information of SPIs. Gateways, located between the smartphones and the server for communication efficiency, act as a Middle Node (MN). The system proposed in this study is being developed.

Acknowledgments. This research was supported by the Basic Science Research Programs through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (No. NRF-2015R1D1A3A03019278). Also, this work (Grants No. C0350317) was supported by Business for Academic-industrial Cooperative establishments funded by Korea Small and Medium Business Administration in 2015.

References

1. I. Heywood, S. Cornelius, S. Carver, An Introduction to Geographical Information Systems (3rd ed.). Essex, England: Prentice Hall, 2006.
2. K. T. Chang, Introduction to Geographical Information Systems. New York: McGraw Hill, 2008.
3. Vida Malienea, Vytautas Grigonisb, Vytautas Palevičiusb and Sam Griffiths, “Geographic information system: Old principles with new capabilities”. Urban Design International vol. 16, no. 1, pp. 1–6, 2011. doi:10.1057/ udi.2010.25.
4. ITU, *The Internet of Things*, 2005.
5. European Commission, *Internet of Things-An action plan for Europe*, 2009.
6. CISCO, *How the Internet of Everything Will Change the World*, 2012.

7. oneM2M, *Functional Architecture (TS-0001-V1.6.1)*, http://www.onem2m.org, 2015.
8. oneM2M, *Requirements (TS-0002-V1.0.1)*, http://www.onem2m.org, 2015.
9. oneM2M, *Service Layer Core Protocol Specification (TS-0004-V1.0.1)*, http;//www.inem2m.org, 2015.
10. oneM2M, *Published Specifications*, http://www.onem2m.org/technical/published-documents, 2015.
11. NFC Forum, http://nfc-forum.org/, 2016.

SESSION

SMART CITIES AND RELATED ISSUES

Chair(s)

TBA

Embedded Intelligence in Smart Cities through Urban Sustainable Mobility-as-a-Service: research achievements and challenges

George Dimitrakopoulos, George Bravos

Harokopio University of Athens, department of Informatics and Telematics

9, Omirou str., 17778, Athens, GREECE, e-mail: gdimitra@hua.gr

***Abstract*— Economic growth in Europe has been, strongly associated with urbanization, overwhelming cities with vehicles. This renders mobility inside cities problematic, since it is often associated with large waste of time in traffic congestions, environmental pollution and accidents. Cities struggle to invent and deploy "smart" solutions in the domain of urban mobility, so as to offer innovative services to citizens and visitors and improve the overall quality of life. In this context, the paper discusses on the fundamental challenges that cities face when trying to become smarter, focusing on the particular area of mobility and presenting some sets of mobility-as-a-service ideas, as well as 3 indicative case studies that showcase the effectiveness of the quest for sustainable mobility in smart cities.**

***Index Terms*—smart cities, smart city operations, mobility, car pooling, parking management, emergency management**

I. Introduction

It is widely accepted that citizens inside large cities at a worldwide level are "bombed" by large amounts of uncorrelated and non-synchronized data, from innumerable sources and through various devices in a complex manner. Citizens are thus not in position to efficiently handle them, this resulting in severe inefficiencies associated with their mobility, such as (i) fragmented travel solutions / lack of door-to-door solutions, especially when dealing with multimodal transportation, as well as (ii) inadequateness in providing real-time, whilst individualized services. Those drawbacks often result in losses of time, decrease in the level of safety in mobility, pollution, degradation of life quality, and huge waste of non renewable fossil energy. Moreover, they affect not only citizens, but all relevant stakeholders, such as also public authorities and businesses.

At the same time, cities keep on becoming smarter and smarter, trying to offer traditional services with unconventional methods (e.g. via Information and Communication Technologies – ICT), as well as completely novel services, often enabled again by ICT. This trend is reflected on a concept coined by IBM, namely the "smart cities" concept [2][3].

Considering that transportation inside large cities is rapidly increasing, alongside with the addition of new transport media (car pooling, car sharing, etc.), it is among a city's priorities to improve the quality of living inside them, providing smart services to their citizens and visitors. As such, it would be of great interest to place a special focus on a "smart" city and try to revolutionarize mobility in the aforementioned context. Further, the above necessitate research towards improving novel mobility practices for citizens/policymakers/businesses. This can be done only by engineering innovative strategies for aggregating large amounts of data from versatile sources (conventional and new ones), intelligently processing it and providing accurate directives associated with actual mobility status and potentials, in a multimodal and concurrently individualized fashion [1].

The contribution of this paper is manifold:

a) It gathers and summarizes all fundamental challenges that arise towards the implementation of Smart City Operations (SCOs);

b) It provides an insight specifically for SCOs focusing on mobility-as-a-service;

c) It defines three representative use cases in order to demonstrate the importance of sustainable mobility services in smart cities and

d) It creates the basis for the design and implementation of such services.

The rest of this paper is organized as follows. Section II provides some fundamental definitions in the smart cities domain, namely Smart City Operations (SCOs), as well as an overview on the relevant research challenges that arise. Section III discusses on sustainable (smart) mobility in smart cities, focusing on the main research achievements and challenges. Then, section IV presents some indicate case studies, used for exemplifying the provision of smart mobility services in urban environments. Concluding remarks are drawn in section V, along with an outlook on future research activities.

II. Smart Cities and Smart City Operations (SCOs)

A. Basic definitions

There is no unique definition of the term "smart city". Instead, there have been several attempts to provide descriptive definitions of the term. As such, according to [11], smart city is a city well performing in a forward-looking way in economy, people, governance, mobility, environment, and living, built on the smart combination of endowments and activities of self-decisive, independent and aware citizens. Likewise, the authors of [12] define as smart a city connecting the physical infrastructure, the IT infrastructure, the social infrastructure, and the business infrastructure to leverage the collective intelligence of the city.

In this respect, SCOs constitute an important development that is expected to have a profound impact on the socioeconomic future of Europe. ICT is a strong enabler for cities to turn "smarter" and thus offer their citizens the opportunity of a better quality of life. This can be achieved through better decision making about a variety of domains within a city.

Particular areas where SCOs find fertile ground for development include light and traffic optimization, energy consumption, public and private transport, health care, environmental protection and citizen empowerment.

Those indicative areas are outlined in the figure below.

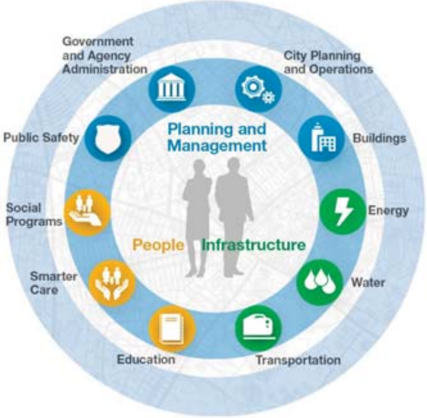


Figure 1: Fundamental SCO [3]

B. SCOs challenges

Smart cities worldwide are becoming increasingly smarter, through capitalizing on new technologies and insights to transform their systems and operations delivery to citizen-centered useful service delivery [1]. To be able to continue advancing in this area and consolidate a solid "smart" background", several fundamental requirements need to be addressed from an operational point of view [4]. To extract those requirements, a set of smart city operations challenges, as identified in the international literature, are detailed below.

1) Level of intelligence ("smartness") required

Intelligence ("smartness") might be a difficult concept to sketch from various viewpoints. As such, a city should appropriately consider a priori the desired levels of smartness to be achieved at short, medium and long time scale. This depends of course to a number of services that a city wants to provide to its citizens, so as to be considered "smart". Moreover, to do so, a city should consider the needs, plans and opinions of all stakeholders involved in its operations, such as (i) citizens, (ii) service providers, (iii) businesses, (iv) municipal authorities and (v) national standards. At the same time, all economic, environmental and people oriented viewpoints should be considered. Last, scalability of the smart operations to be provided, should also be considered. This means achieving a balance not just between the interests of a particular city's stakeholders, but also taking into account relationships with neighboring cities. The above seems a complex algorithmic process with multiple variables [8][11][12].

2) Technology

Undoubtedly technology constitutes the primary driver towards the transformation of a city from a conventional one to a smart one. A smart city relies, among others, on a collection of smart computing technologies applied to critical infrastructure components and services. Smart computing refers to a "new generation of integrated hardware, software, and network technologies that provide IT systems with real-time awareness of the real world and advanced analytics to help people make more intelligent decisions about alternatives and actions that will optimize business processes and business balance sheet results [15].

In this respect, numerous challenges can be identified, such as

(i) Lack employees with integration skills and culture, (ii) Lack of cross-sectoral cooperation, (iii) Lack of interdepartmental coordination, (iv) Unclear vision of IT management, (v) Politics, as well as (vi) Culture issues.

3) Scalability of smart solutions

Smartness should be scalable enough, in that a city should appropriately design the objectives to be achieved at various scales.

First come the minimum objectives that will attribute "smart" characteristics to the city and will be able to provide its citizens the minimum levels of quality needed to live a civilized life. At this high-level stage, the values of a city and its residents include many qualitative concepts and things of an emotional nature, such as lifestyle values and a sense of attachment to the neighborhood [4].

Second come the fundamental objectives that will enhance the level of smartness of the city towards a desired level, such as e.g. the reduction of carbon emissions. Such objectives could be agreeable at a local / regional / national level.

Last come some longer term objectives that will further advance the smartness level already achieved, which are usually set at a local level, albeit being negotiable also at an international level in the context of organizations and fora.

4) *Formulation of city-specific objectives*

A city usually sets at a local level some standards to be achieved at various time scales. Then, some Key Performance Indicators (KPIs) are monitored, so as to evaluate the achievement of those standards. Those KPIs are nothing less than city-selected criteria / benchmarks. Moreover, KPIs should be adaptive enough to respond to new (external) requisitions [6][7][8].

In order to formulate city specific objectives, factors that need to be taken into account are (i) the people and cultural diversity, as well as (ii) the environment [9][10].

5) *Economic growth*

From a high level, economics viewpoint, a city can be thought of as an entity that enables internally operating business groups to obtain income from outside its geographical region, and then enables the obtained revenues to circulate within its region. This of course can function the other way round (extroversion).

Accordingly, the economic performance of a city can be viewed from two viewpoints: its industrial competitiveness relative to other regions, and the soundness of the finances within its region.

In this respect, it is essential that when planning and designing the provision of smart city operations, one must take a holistic, long term approach. In particular, the assessment of strengths, weaknesses, opportunities and threats needs to look 10 or even 20 years ahead. Such a process will allow a city to continue attracting immense attention for businesses, whilst being comfortable and secure for its citizens [6][7].

6) *Management and organization*

There are only a few studies in the academic literature on smart city initiatives that adhere to address issues related to managerial and organizational factors of a city. In contrast, a wide array of previous research on IT initiatives and projects has highlighted these issues as important success factors or major challenges [13][14]. Thus managerial and organizational concerns in smart city initiatives need to be discussed in the context of the extensive literature on e-government and IT projects success.

In this respect, the authors of [13] suggested several challenges, namely (i), Project size, (ii) Manager's attitudes and Behavior, (iii) Users or organizational diversity, (iv) Lack of alignment of organizational goals and project, (v) Multiple or conflicting goals, (vi) resistance to change, as well as (vii) Turf and conflicts.

III. Smart Mobility in Smart Cities

The latest mass transit and e-mobility technologies match with city infrastructures from monorail and metro systems running through buildings at-grade, elevated or underground, to new solutions for electric vehicles. These solutions support a better way, which helps us thinking from traditional transport modes to electric public transport.

Smart mobility is a key challenge in the world. The huge increase in urban population and the growing environmental topic find prosper ground to the concept of smart mobility, which proposes solutions for greener, safer and more efficient transfer of humans and freight.

Historically, mobility has been seen as a product. That includes the vehicles, physical infrastructure and fuels which used people to mobilize. But, mobility is approached as a service also. This means that mobility is a method by which we provide food, engage in economic activity, access entertainment or meet with friends and family, all through ideal movements from place to place. When we use mobile phones, web and video to manage our lives on the go, the ways in which we discharge these tasks are changing. These new capabilities rely on physical and digital infrastructure whose potential is only beginning to be carried out. By supplementing urban planning and management practices with digital technologies, there is an opportunity to improve mobility services for citizens, while managing demand on physical transport networks and generating wider economic and environmental value [17].

In this way, the challenges in smart mobility are [18]:

i) To develop a system that can communicate with the vehicle and so the user is able to receive information from the surrounding environment, which can have influence in the vehicle performance (traffic information, internet-connected vehicles, parking management, car pooling, etc);

ii) To make the best effective use of the trip planning and routing of fully electric vehicles, using information from these sources including alternatives from other transport modes adapted to user's needs;

iii) To set efficient and optimum charging strategies which match to user and fully electric vehicles needs and grid conditions, as well as

iv) Using energy saving methods (as driving modes and In-Car Energy Management Services) within the fully electric vehicles interaction with the driver.

An example is a new mobility model, the Mobility-as-a-Service (MaaS). MaaS bridges the gap between public and private transport operators, envisaging the integration of all the fragmented tools (planning, booking, real time information, payment and ticketing) a traveler needs to conduct a trip. This model reduces the dependence on private vehicles and allows modern travelers in urban areas to plan and manage their transit quickly and safely using their smart phones. The key to successful uptake of such services is the effective integration among different technologies and tools.

Several EU funded initiatives are relevant to the smart mobility in smart cities in general, as well as MaaS in particular. The InSMART [20] concept brings together cities, scientific and industrial organizations in order to establish and implement a comprehensive methodology for enhancing sustainable planning addressing the current and future city energy needs through an integrative and multidisciplinary planning approach. READY4SmartCities [21] operates in a European context where other initiatives are currently running in order to create a common approach on Smart Cities. STEP-UP - Strategies Towards Energy Performance and Urban Planning [22] takes an integrated approach to energy planning, integrated project design, and implementation by addressing 3 vital themes together: energy and technology; economics; organisation and stakeholders.

TRANSFORMation Agenda for Low Carbon Cities [23] supports cities to meet the 20-20-20 targets by the integration of energy in urban management. PLEEC Planning for energy efficient cities [24] gathers cities with innovative planning and ambitious energy saving goals. SMART-ACTION [25] supports the development of strategic research agendas and serves as an enabler for the dissemination and further integration of results into future research and industrial developments, while coordinating international efforts.

Also, smart mobility can improve the economic gains. One way is to improve intermodal transport for better efficiencies and economic activity, for example, in Osaka, Japan, high-speed rail is blended with connections to local public transport networks to encourage public transport use and drive down the demand for roads. This reduces congestion, travel times and supports greater productivity. In another case, the efficient, integrated intelligent transport system in Medellin, Colombia, is cutting travel time into the city from many hours to 30 minutes, linking residents to jobs and education and driving economic gains.

In accordance with the above, the next section presents 2 indicative case studies that fall in the realm of smart mobility services offered to the citizens and visitors of smart cities.

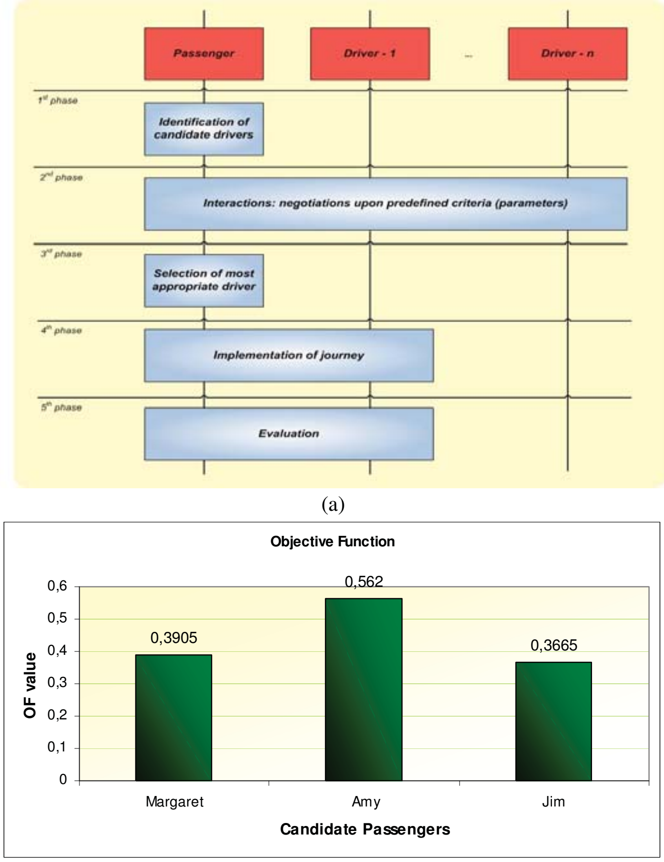
IV. Indicative Case Studies

A. *Car pooling*

Car pooling is an idea that has been proposed by several researchers since many years, mainly for reducing traffic [26][27]. However, with the advent of ICT, this idea has turned into a smart mobility service offered by some cities/regions, to their citizens and visitors [28].

The service assumes a set of candidate drivers, namely the drivers' pool, as well as a set of prospective passengers, namely the passengers' pool. Drivers and passengers are associated with context information parameters, i.e. data on their current positions and itineraries. Furthermore, they are associated with personal profile parameters, as well as with service related parameters. Last, a set of overarching policies reflects driver/passenger preferences, in the form of weights (importance) attributed to the aforementioned parameters. The above are depicted on

In general, personal profile and service parameter values can change from time to time. The authors of [29] propose a functionality that can interact, on behalf of the prospective passenger, with all candidate drivers and find and propose an optimum match, taking into account the request, the available context, personal and service profile information, the policies, as well as previous knowledge turned into experience, which increase the degree of reliability of the decisions reached.



(a)

(b)

Figure 2: Car pooling, (a) indicative business case, (b) selection of the most appropriate passenger

The SCO functionality presented in [29] uses previous knowledge in proposing valid car pooling matches. Knowledge is obtained through the exploitation of Bayesian networking concepts and specifically the Naive-based model. Results showcase its effectiveness, the advantage of which lies in that the reliability of the knowledge-based selection decisions is higher. This means that there is higher probability of satisfying the drivers' and passengers' preferences through the selected matches.

Overall, car pooling is a representative case of an SCO that more and more cities tend to offer, often combined with most modern ideas/services, such as car sharing.

B. *Intelligent parking management*

A major contribution toward the improvement of the quality of transportation in large cities would be the introduction of a system that would reside inside vehicles or even consist in a smart phone application, communicate with the transportation infrastructure using IP, obtain information on "white parking spaces", and then issue the appropriate directives to the driver so as to drive the vehicle to the desired parking space.

The information acquisition is based on a parking management system on the infrastructure side, which disposes a database of white parking spaces through information received from sensors (cameras) that constantly updates the database on the location and size of white parking spaces.

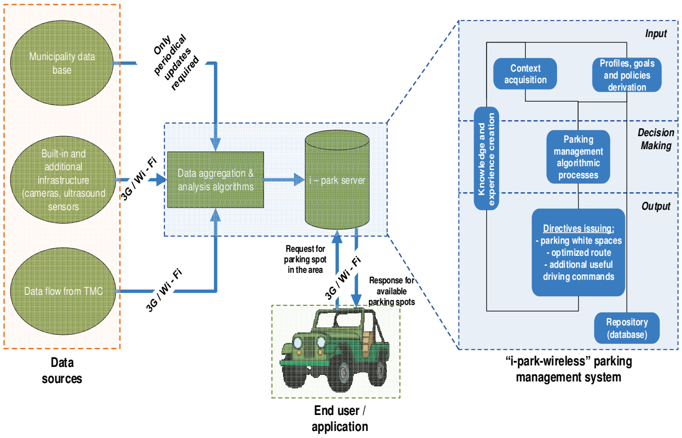

Municipality data base
Built-in and additional infrastructure (cameras, ultrasound sensors
Data flow from TMC
Data sources
Only periodical updates required
3G / Wi - Fi
3G / Wi - Fi
Data aggregation & analysis algorithms
i – park server
Request for parking spot in the area
3G / Wi - Fi
Response for available parking spots
End user / application
Input
Context acquisition
Profiles, goals and policies derivation
Knowledge and experience creation
Decision Making
Parking management algorithmic processes
Output
Directives issuing:
- parking white spaces
- optimized route
- additional useful driving commands
Repository (database)
"i-park-wireless" parking management system

Figure 3: Intelligent parking management

Cities tend to study more and more systems like the one shown in Figure 3, since, in doing so, several impressive achievements could arise, namely:

- minimization of the time consumed in searching for a white parking space,
- improvement of the quality of the driver's life through a most productive utilization of time,
- increase in the mobility efficiency through the identification of the most appropriate route towards the white parking space,
- adoption of "green transportation" techniques through the real time adaptation to changes (e.g. sudden occupation of a parking space) and minimization of the GHG (Greenhouse Gas) levels in terms of the energy required / consumed from the moving vehicles.

C. Emergency management: early warning system for vehicles

As mentioned also above, by enabling vehicles to communicate with each other, as well as with roadside base stations via Vehicle-to-Infrastructure (V2I) communication, ITS can contribute to safer and more efficient roads and cities can offer smarter and smarter mobility services to their citizens and visitors.

In the light of the above, [30] proposes a mobility SCO, targeted at proactively managing vehicles and the surrounding transportation infrastructure quickly and efficiently, in a way that guarantees significant improvements in traffic / safety / emergency management.

The proposed approach is a smart city operation indeed, as it combines (i) wireless sensors placed on the vehicles and on specific parts of the transportation infrastructure (traffic lights, road signs), (ii) Wireless Sensor Networks (WSNs) formed by neighboring vehicles and parts of the infrastructure, thus referred to as "vehicular sensor networks" (VSNs) and (iii) a computationally efficient heuristic for evaluating the available information and proactively issuing directives to the drivers and the overall transportation infrastructure, which may be valuable in context handling.

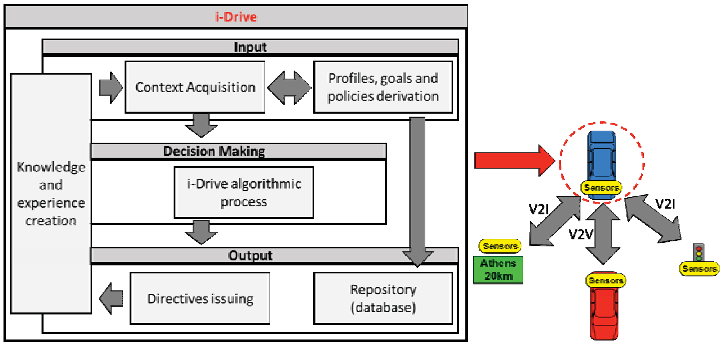


Figure 4: Warning system for vehicles as an SCO

The particular contribution of this SCO mainly lies in the utilization of a knowledge-based decision making algorithm, which can increase the overall levels of safety through recognizing potential emergencies a priori, improving thus the total transportation quality. Moreover, it laterally also addresses the integration of the advantages of vehicular sensor networks in ITS through the description of a whole framework that can incorporate various services/applications that can improve the quality of transportation.

V. Conclusions and Future Directions

This paper discussed on smart mobility services offered in the context of SCOs. As such, it first provided some fundamental definitions in the smart cities domain, as well as some basic challenges that cities face when designing SCOs. Then it focused on smart mobility that falls in the realm of SCOs, presenting its main research achievements and challenges. Then it went through some indicative case studies for exemplifying the provision of smart mobility services as SCOs.

Overall, smart cities are continuously getting smarter. This naturally requires capital expenditure and calls for novel solutions in various areas. Transportation is an area where SCO find prosperous ground since it can increase the quality of living in large cities.

Several exciting areas are yet to be explored in the area of mobility offered in the context of SCOs. In particular, the further exploitation of intelligent transport systems principles in SCOs can lead to a 100% real-time assessment of traffic congestions, a priori identification of forthcoming dangers, as well as to the provision of open APIs and interfaces for intermodal MaaS inside cities/regions. Moreover, city-wide services can inform drivers on city-specific events (cultural, etc.), as well as on city-specific incidents (e.g. protests, works, etc.) and offer also targeted/focused ads and infotainment. Last, the exploitation of modern mobile communication infrastructures (e.g. 5G D2D) with which cities are more or less equipped, can naturally reduce deployment costs and provide low-latency emergency management services.

VI. Acknowledgment

The authors wish to acknowledge the Qatar National Research Fund project i-Doha (Proj. No. NPRP 7-662-2-247). project,

under the auspices of which the work presented in this paper has been carried out.

VII. References

[1] G. Dimitrakopoulos, P. Demestichas, “Intelligent Transportation Systems based on Cognitive Networking Principles”, IEEE Vehicular Technology Magazine (VTM), March 2010.

[2] Toppeta, D., The Smart City Vision: How Innovation and ICT Can Build Smart,“Livable”, Sustainable Cities, 2010, The Innovation Knowledge Foundation. Available from http://www. thinkinnovation. org/file/research/23/en/Toppeta_Report_005_2010. Pdf .

[3] http://www.ibm.com/smarterplanet/us/en/smarter_cities/overview/ , accessed February 26th, 2015

[4] http://www.hitachi.com/products/smartcity/download/pdf/whitepaper.pdf , accessed March 16th, 2015

[5] BIS, The Smart City Market: Opportunities for the UK, 2013, Department of Business, Innovation and Skills.

[6] Naphade, M., Banavar, G., Harrison, C., Paraszczak, J., Morris, R., “Smarter Cities and Their Innovation Challenges”, IEEE Computer, Volume: 44 , Issue: 6 , 2011, pps. 32 - 39

[7] Hogan, J., Meegan, J., Parmar, R., Narayan, V., Schloss, R.J., “ Using standards to enable the transformation to smarter cities”, IBM Journal of Research and Development, Volume: 55 , Issue: 1.2, 2011, pps. 4:1 - 4:10

[8] Difallah, D.E., Cudre-Mauroux, P., McKenna, S.A., “Scalable Anomaly Detection for Smart City Infrastructure Networks”, Internet Computing, IEEE, Volume: 17 , Issue: 6 , 2013, pps. 39 - 47

[9] Benouaret, K., Valliyur-Ramalingam, R., Charoy, F., “CrowdSC: Building Smart Cities with Large-Scale Citizen Participation”, Internet Computing, IEEE, Volume: 17 , Issue: 6 , 2013 , Page(s): 57 - 63

[10] Walravens, N., Ballon, P., “Platform business models for smart cities: from control and value to governance and public value”, Communications Magazine, IEEE, Volume: 51 , Issue: 6 ,, 2013, pps. 72 - 79

[11] Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanović, N., & Meijers, E., “Smart Cities: Ranking of European Medium-Sized Cities”, Vienna, Austria: Centre of Regional Science (SRF), Vienna University of Technology. Available from http://www.smartcities.eu/download/smart_cities_final_report.pdf

[12] Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., & Williams, P., “Foundations for Smarter Cities”, IBM Journal of Research and Development, 54(4).

[13] Gil-García, J. R., & Pardo, T. A., “E-government success factors: Mapping practical tools to theoretical foundations”, Government Information Quarterly, 22(2), 187-216.

[14] Scholl, H. J., Barzilai-Nahon, K., Ahn, J-H., Olga, P., & Barbara, R., “E-commerce and e-government: How do they compare? What can they learn from each other?”, in Proc. 42nd Hawaiian International Conference on System Sciences (HICSS 2009), Koloa, Hawaii, January 4-7.

[15] Washburn, D., Sindhu, U., Balaouras, S., Dines, R. A., Hayes, N. M., & Nelson, L. E., “Helping CIOs Understand "Smart City" Initiatives: Defining the Smart City, Its Drivers, and the Role of the CIO.Cambridge, MA: Forrester Research, Inc. Available from http://public.dhe.ibm.com/partnerworld/pub/smb/smarterplanet/forr_help_cios_und_smart_city_initiatives.pdf.

[16] Nam, T. & Pardo., T. A., “Conceptualizing Smart City with Dimensions of Technology, People, and Institutions. In Proceedings of the 12th Annual Digital Government Research Conference, College Park, Maryland, June 12-15

[17] Charbel Aoun, “Urban mobility in the smart city age”

[18] “Smart mobility in smart city”. Available from http://www.mobincity.eu/about_mobincity/project_overview

[19] Government & Public Sector Insights, “Routes to prosperity, How smart transport infrastructure can help cities to thrive”

[20] http://cordis.europa.eu/project/rcn/186975_en.html, accessed December 2015

[21] http://cordis.europa.eu/project/rcn/110042_en.html, accessed December 2015

[22] http://cordis.europa.eu/project/rcn/186983_en.html, accessed December 2015

[23] http://cordis.europa.eu/project/rcn/186978_en.html, accessed December 2015

[24] http://cordis.europa.eu/project/rcn/186984_en.html, accessed December 2015

[25] http://cordis.europa.eu/project/rcn/109708_en.html, accessed December 2015

[26] R. F. Teal, “Carpooling: who, how and why”, Transportation Research, Part A: General, 21 A (3), pp. 203-214, 1987.

[27] G. Giuliano, W. Douglas, D. Levine and R. Teal, “Impact of high occupancy vehicle lanes on carpooling behavior”, Transportation 17 (2), pp. 159-177, 1990.

[28] Darm - Division of Resources Management. Carpooling and You. Public domain document, Florida, U.S., January 2005. Available online: http://www.dep.state.fl.us/Air/publications/airpubs/carpool.pdf

[29] G. Dimitrakopoulos, P. Demestichas, V. Koutra, "Intelligent Management Functionality for Improving Transportation Efficiency by means of the Car Pooling Concept", IEEE Transactions on Intelligent Transportation Systems, vol 13, issue 2, pp. 424-436, June 2012.

[30] G. Dimitrakopoulos, G. Bravos, M. Nikolaidou and D. Anagnostopoulos, “A Proactive, Knowledge-Based Intelligent Transportation System based on Vehicular Sensor Networks”, IET Intelligent Transport Systems journal, vol. 7, Issue:4, pp 454 - 463, December 2013.

Sentiment Analysis for Smart Cities: State of the Art and Opportunities

Kaoutar Ben Ahmed[1], **Atanas Radenski**[2], **Mohammed Bouhorma**[1] **Mohamed Ben Ahmed**[1]
[1]Abdelmalek Essaâdi University, Tangiers, Morocco
[2]Chapman University, Orange, CA 92866, USA

Abstract - *Advances of information and communications technologies in general and of social media platforms in particular have changed the way people communicate and express themselves. Citizens are now using smartphones and other mobile devices to share, on an unprecedented scale, their experiences and views in blogs, micro-blogs, comments, photos, videos, and other postings in social sites. The smart city research community has already recognized that sentiment analysis can contribute to a better understanding of, and timely reactions to, public's needs and concerns by city governments. Yet, relatively little is known about how to best harness the potential benefits for smart cities of opinion mining and sentiment analysis. The objective of this article is to help fill the void by reviewing the state of the art, challenges and opportunities of sentiment analysis platforms, architectures and applications for the smart city application domain.*

Keywords: sentiment analysis, opinion mining, smart cities, big data, text mining

1 Introduction

Smart (or smarter) cities "are urban areas that exploit operational data, such as that arising from traffic congestion, power consumption statistics, and public safety events, to optimize the operation of city services" [1] A smart city's main objective is to increase the quality of life for its citizens and to make the city more attractive, lively and greener. To achieve this goal, smart city technologies (SCTs) are fused with the traditional city's infrastructure. SCTs refer to all the information and communications technologies (ICTs) that enable cities to harness big data gathered and analyzed in order to become connected and sustainable. The smart city concept emerged during the last decade as a fusion of ideas about how ICTs might improve the functioning of cities, enhancing their efficiency, improving their competitiveness, and providing new ways in which problems of poverty, social deprivation, and poor environment can be addressed. Urban communities worldwide are planning, developing and adopting digital systems and technologies to improve efficiency and quality of life for the citizens. According to a 2014 forecast, in 2025 there will be at least 26 global smart cities, about 50 percent of which will be located in North America and Europe [2]. Open data initiatives around the world make public data available online to wider audiences. Such initiatives have provided in recent years support for innovative projects, aiming the design of SCTs that enhance cities' smartness. A number of technological developments illustrate this trend. For example, real time big data analyses are used to boost public safety effectiveness by integrating smart solutions in disaster and emergency management, and in law enforcement systems. Smart mobility solutions based on road sensors and intelligent transportation systems are employed in transport and environment fields aiming to reduce urban congestion and CO2 emission. Websites and cloud services make a number of city government services accessible for anyone over the Internet; they can also become smarter due to tracking and analytics technologies that can discover usage and access patterns and respond to citizen's individual needs and preferences. Recommendation systems, mobile location-based applications, and other advanced ICTs offer city visitors novel and personalized experiences during which they are assisted by smartphone apps acting like electronic tourist guides updated with real time information about accommodation, dining options, weather, currency rates, nearby points of interest based on the visitor's personal preferences or geographical location or other strategies and touch-sensitive maps are found in all stations to help tourist easily find their way. Social media triggered the raise of sentiment analysis which brings new possibilities to city governance in general and decision making in particular.

Sentiment analysis is the examination of people's opinions, sentiments, evaluations, appraisals, attitudes, emotions, and personal preferences towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [3].

Traditionally, sentiment analysis mines information from various text sources such as reviews, news, and blogs then classifies them on the basis of their polarity as positive, negative or neutral. An important preliminary task of sentiment analysis is to evaluate the subjective or objective nature of source texts. Subjectivity indicates that the text bears opinion content whereas objectivity indicates that the text is without opinion content. Recently, sentiment analysis aims to exploit audio, video, location, and other non-traditional data sources.

An essential issue in sentiment analysis is to identify how sentiments are expressed in texts and whether the expressions indicate positive (favorable) or negative (unfavorable) opinions toward the subject. Thus, sentiment analysis involves identification of sentiment expressions, polarity and strength of the expressions, and their relationship to the subject [4].

Historically, sentiment analysis has been exclusively approached as a natural language processing task at many levels of granularity. Starting from being a document level classification task [5], it has been handled at the sentence level ([6]; [7]) and more recently at the phrase level ([8]; [9]).

Over the years, scholars and developers have introduced a number of terms that refer to tasks that are very similar to sentiment analysis, such as opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, and others. Recently, such diverse terms are converging under the umbrella of either sentiment analysis, or opinion mining [3]. To our knowledge, the two terms were first published at the same time, in 2003: the term sentiment analysis appeared first in [4] and the term opinion mining was first published in [10]. Currently, the two terms seem to represent the same field of study. In industry, the term sentiment analysis is more commonly used than opinion mining; in academia both sentiment analysis and opinion mining are frequently employed. In this paper, we choose to use the term sentiment analysis because it is well adopted in both industry and academia. We also acknowledge that sentiment analysis now expands beyond its traditional text sources to aim the analysis of non-text data, such as image and video data from social networks, and human vitals as measured by wearable computers.

In this paper, we focus on sentiment analysis as a promising smart city technology and offer a review of the state of the art, the challenges, and the opportunities in the areas of tools and techniques (section 2) and system architectures (section 3). We limit our considerations, unless otherwise required, to the smart city application domain and do not aim to discuss sentiment analysis in its full generality.

2 Tools and Techniques

Sentiment analysis relies on a variety of tools and techniques; their weaknesses and limitations justifies the necessity of a new generation of tools and techniques that either solve present challenges or offer new opportunities in the smart city domain.

2.1 State of the art

Comprehensive surveys like [5] and [3] reviewed various tools and techniques described in an ever growing pool of sentiment analysis publications. The particular focus of our article is on sentiment analysis tools for the smart city domain; therefore, we only briefly review results that we deem applicable to this domain. Thus, [11] used a sentiment detection tool named LIWC2007 which works with a psychometrically validated dictionary. [12] proposed a fusion between a lexicon-based sentiment classifier and a machine-learning based classifier. [13] investigate and evaluate NTLK, a platform for natural language processing in python, and also SentiWordNet3.0, a lexical resource for sentiment analysis. [14] used a Naïve Bayes model on unigram features and Crowdsourcing via Amazon Mechanical Turk (AMT). [15] devised a new sentiment detection mechanism that determines the keywords related to public reaction and descriptions of terrorism. This is done by feeding a list of potential sentiment-related keywords into their sentiment finder module, which will then tag the messages based on category of the keywords detected in the incoming message. The category defines the sentiment of the message. [16] applied logistic regression, Naive Bayes and Decision Tree (ID3) classifiers to perform sentiment classification of Twitter messages regarding Hurricane Irene. They have concluded that Tweets provide real-time insight into public perceptions of a disaster. [17] utilized Saltlux, which is a proprietary crawling and sentiment mining software, in order to convert the social media streams gathered mainly from Twitter into RDF stream and analyze its sentiment. The R programming packages and Support Vector Machines are applied by [18] to extract sentiment scores from newspaper articles and relate these scores to an economic index. [19] argues that the calculation of a metric named “Gross national happiness” serves as a representation of the overall emotional health of the nation. To calculate this metric, he used the sentiment analysis approach via Text Analysis and Word Count (TAWC) program. [20] used OpinionFinder tool that measures positive vs. negative mood and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). Their goal was to investigate whether public mood as measured from large-scale collection of tweets is correlated or even predictive of economic indicators.

2.2 Opportunities

Current research in sentiment analysis focuses on the reducing of human effort needed to analyze content and increasing accuracy. Future research is expected to address such issues as visual representation, multilingual audiovisual opinion mining, building usable, peer to peer opinion mining tools for citizens, real-time sentiment analysis, multilingual reference corpora, recommendation algorithms, and automatic irony detection [21]. Additionally, the use of open source software is also expected to increase, most notably Hadoop ecosystem for batch data analysis (with Hive and Hbase), for interactive data analysis (with Presto and Storm) and for recommender systems in particular and statistical analysis in general (with Mahout).

3 Systems

Various system architectures have been proposed for sentiment analysis in the smart city domain. We describe some of them, discuss principle challenges for such systems, and we propose our own system architecture to address some of the difficulties with earlier systems.

3.1 State of the art

The architecture of a restaurant recommendation system [17] consists of three parts: (i) a client segment that interacts with the user and communicates to the back-end sending SPARQL (Protocol and RDF Query Language) queries, (ii) a data initiated segment that continuously analyses the social media streams, and (iii) a query initiated segment that uses the LarKC platform to answer the SPARQL queries of the client by combining several forms of reasoning. A proposed system for terrorism events detection [15] is implemented in Perl using the Net-Twitter Perl module that connects to Twitter with the Twitter (API); it queries the Twitter 'trends' API for names of places and identifies discussions of a flurry of activities at a specific location. Once the location of a possible threat has been identified, the system harvests all related Twitter messages using Twitter's Search API and Streaming API. A spam and noise filtering phase precedes a sentiment detection phase and a demographic exploration of the message pool. Finally, data mining and reporting phase takes place. In a smart learning context, a sentiment evaluation system by [12] extracts information through the Facebook API and records it in a NOSQL database. The system then models, evaluates, and visualizes the users' sentiment polarity (positive, neutral or negative) and users' significant emotional changes. TweetAlert [22] is a citizen's behavior analysis system which extracts tweets via the Twitter API and stores them it in a data warehouse which runs on top of Apache Lucene. The system uses the Textalytics APIs to perform sentiment analysis on the extracted data. Most notably, its user demographics analysis module extracts some important demographics (type, gender, age). End users can rely on the system's visualization component to exploit the stored annotated data. Several widgets have been developed to present the data, either just for query and reporting or for data analytics purposes. A city sensing system [23] goes through two different phases: the offline training phase and the online phase. In the offline training phase, it collects messages from Twitter that contain emotion-word hashtags. The messages are preprocessed and features are extracted from them. The features, together with the emotion derived from the emotion word hashtags are used to train a neural network. In the online phase, the system collects live geo-tagged tweets from the area of interest, e.g., a city. The trained neural network is used to detect emotion in these new tweets. The geotagged emotion data is then aggregated and visualized on a map. A presidential election analysis system is designed by [14]. The system's real-time data processing infrastructure is built on the IBM's InfoSphere streams platform. All relevant tweets are collected in real-time from the entire Twitter traffic via Gnip Power Track, a commercial Twitter data provider. Christopher Potts' basic Twitter tokenizer is used to preprocess the collected tweets. The system relies on manual sentiment annotation by Amazon Mechanical Turk users. The system outputs the number of positive, negative, neutral and unsure tweets in a sliding five-minute window. An Ajax-based HTML dashboard displays volume and sentiment by candidate as well as trending words and system statistics. The dashboard pulls updated data from a web server and refreshes its display every 30 seconds. This election analysis system can very well be applied to cover smart city elections.

3.2 Opportunities and Proposed System

The info-foundation of a smart city is related to the availability of big data, the interconnection of all city components, e.g., pollution sensors, traffic systems, social media and smartphones generates an increasing amount of data of all types, public or private, structured or unstructured and streamed or static. Collecting, analyzing and visualizing this big data come with the promise of empowering and enhancing the smart city governance. Smart cities data is available in variety of formats. Shared content by users on social media sites is often not only text but images as well and with the recent advances in visual sentiment analysis the now popular selfies can form a rich data source for sentiment analysis on social media. A few recent works attempted to predict visual sentiment using features from images ([24]; [25]; [26]; [27]; [28]). There are also studies that analyze speech-based emotion recognition ([29]; [30]; [31]; [32]; [33]). Additionally, a growing number of studies ([34]; [35] ;[36]; [37]; [38]; [39]) are concerned with multimodal sentiment analysis, integrating visual, acoustic and linguistic modalities. Some scholars ([40]; [41]) have claimed that the integration of visual, audio, and textual features can improve the analysis precision significantly over the individual use of one modality at a time leading to error rate reductions of up to 10.5%. Last, but not least, smartwatches are expected [42] to become capable of providing data on emotions: mood valence (positive vs. negative) and arousal (high vs. low). The location of smartwatches permits easy recording of heart rate variability and galvanic skin response (GSR). These two elements can be used to identify physiological arousal. With the advent of smartwatches and effective multisensor data collection, new algorithms (for sensor data fusion) might be developed that can identify valence without the need for processing a facial image. At this time, only the text data format is sufficiently well-understood for machine learning and sentiment analysis. In this paper, we propose a novel architecture of a sentiment analysis system designed for smart city governance purposes that will support a gradual transition sentiment analysis of text data towards sentiment analysis of multi-modal data. We believe that such architecture should be based on the emerging Apache Hadoop big data ecosystem in general and its Spark engine, and also on recently evolving tools that permit the incorporation within Hadoop of powerful statistical analysis languages and libraries, most notably R and Python's pandas. As a result, our proposed system architecture will amalgamate the big data potential of Hadoop with the existing wealth of R and Python libraries.

R is a popular open source software for statistical computing and data visualization. It was initially announced in 1993 as an extension of the S programming language with Scheme-like scoping rules. R is cross-platform (UNIX platforms, Windows and MacOS). R can be used as an interactive console, where users can try out individual statements and observe the output directly. This is useful in creating custom R scripts and exploring the data, where the output of the first statement can inform which step to take next [43].

R has a rich ecosystem of packages that is continuously enhanced by the active R community. It has over 4800 packages in topics like econometrics, data mining, spatial analysis, and bio-informatics. R excels over other open source machine learning libraries for it implements the majority of ML algorithms, including support vector machines, artificial neural network, naïve bayes, bayesian network, maximum entropy, part-of-speech tagging, named entity recognition, n-gram statistics, regressions, k-nearest-neighbors, hidden markov models, extreme learning machine. The rich R ecosystem has already been used in a number of sentiment analysis projects, such as ([44]; [45]; [18]), to mention a few.

Python is dynamic programming language which was first announced in 1991. Initially conceived as an easy top learn educational language, Python has recently gained popularity as a productive general purpose platform - in contrast to R, which is specialized for statistical computing. The adoption of Python in our architecture will support necessary system functionality beyond statistical analysis, such as web services for example. Yet, Python can be used for statistical processing as well, mainly through the emerging pandas framework - an open-source data manipulation and analysis library, which provides high-performance, easy-to-use data structures and data analysis tools. Python pandas has been created with the goal of becoming the most powerful and flexible open source data analysis tool available in any language [46].

The Apache Hadoop framework allows distributed processing of large data sets across commodity clusters by means of relatively simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the framework itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures [47]. The Hadoop core consists of four modules: Hadoop Commons, Distributed File System, YARN and MapReduce. A number of additional modules complement the Hadoop core by providing specialized services. For instance, Hive is a data warehouse with an SQL interface and Pig is an interpreter of a high-level programming language; both Hive and Pig compile into MapReduce. Hadoop's ability to process large volumes of data in a fault-tolerant batch mode has already attracted the attention of sentiment analysis researchers and its potential to solve various sentiment analysis tasks has been recognized ([18]; [19]). The MapReduce distributed computing engine, for example, can be pared with the HBase distributed storage to handle opinion lexicons and Mahout's machine learning algorithms can be applied to execute sentiment analysis tasks [48].

Apache Spark is notable big data framework, initially developed at the University of California in Berkeley and later transferred to Apache. Spark's capability to fully use the cluster's available fast memory gives it a significant performance edge to Hadoop's original MapReduce engine. Additionally, Spark offers a more flexible – in comparison to MapReduce - programming model that readily supports iterative processing – an area of weakness of the MapReduce model. Hence, Spark is emerging as a viable MapReduce alternative, which has become a most intensively developed big data framework. (According to OpenHub, during 2015 Spark has had 10194 Commits and 656 total contributors). Spark application can run in standalone mode or on Hadoop and be managed by YARN or Mesos. In the sentiment analysis domain, Spark has a clear edge over MapReduce because of its capability to support iterative machine learning algorithms.

While the sentiment analysis capabilities of Spark and, separately, of R and Python have been acknowledged, little is known about the benefits of their combined potential for sentiment analysis in general, and sentiment analysis for smart city governance in particular. This combined potential can now be explored and realized by means of the SparkR, PySpark, and Sparkling pandas tools. SparkR is a tool that provides a light-weight frontend to use the Spark big data engine from R [49], thus enabling native R programs to scale in distributed setting. Such integration of R and Spark brings a number of benefits including scalability to many cores, and machines within large clusters, optimizations in terms of code generation and memory management and the also ability to connect to a variety of data sources, such as Hive tables, JSON files, Parquet files etc. On the other side, the PySpark tool permits the use of the Spark engine from Python. PySpark, in combination with the SparklingPandas tool supports large scale data analysis with pandas on top of Spark.

We propose a system architecture (Fig. 1) which uses emerging tools, such as SparkR, PySpark, and Sparkling pandas to amalgamate the big data potential of the Hadoop ecosystem in general and its Spark engine in particular with powerful statistical analysis languages and libraries, most notably R and Python. We envision a two-stage implementation of such architecture. The core of this architecture comprises well established technologies - such as

R and Hadoop – that are traditionally used to handle text data for machine learning and sentiment analysis. Extension components for multi-modal data analysis, based on Python, PySpark, and Sparkling pandas, can be designed and developed after gathering some initial smart city sentiment analysis experience with the core components.

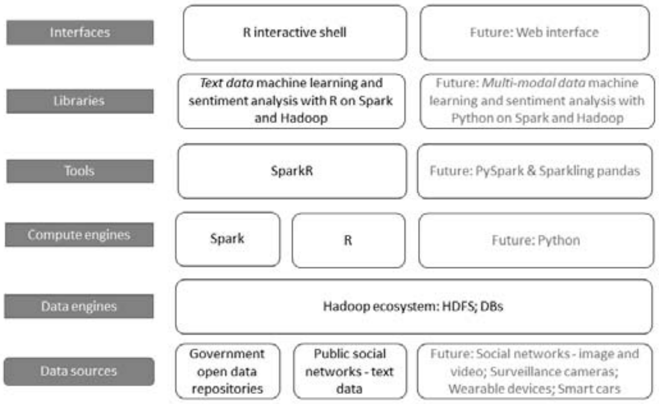

Interfaces
R interactive shell
Future: Web interface
Libraries
Text data machine learning and sentiment analysis with R on Spark and Hadoop
Future: Multi-modal data machine learning and sentiment analysis with Python on Spark and Hadoop
Tools
SparkR
Future: PySpark & Sparkling pandas
Compute engines
Spark
R
Future: Python
Data engines
Hadoop ecosystem: HDFS; DBs
Data sources
Government open data repositories
Public social networks - text data
Future: Social networks - image and video; Surveillance cameras; Wearable devices; Smart cars

Fig.1. System architecture of sentiment analysis applications for smart cities

A central goal of our architecture is to integrate within Apache Hadoop the leading open-source big-data machine learning libraries, SparkR and Sparkling pandas, and support their use as smart city big data technologies, and to also fuse all the advantages and strengths of these tools via an easy-to-use web interface. The proposed architecture can be extended toward a high level, user-friendly decision support system that aims to assist governments and smart cities' managers by the use of civic engagement in policy development and implementation. Sentiment data can be extracted, in the near future, from several sources in modern cities, such as social media, government open data, surveillance cameras, selfie images, wearable devices, humanoid robots, to mention a few. The system core, together with extensions for analysis of non-text data represents a multimodal sentiment analysis system that can be employed as a sensing tool to gauge citizens' sentiments toward public subjects. A user-friendly interface complemented with interactive graphical visualizations capabilities will permit the use of this system by managers and civic leaders who have no technical programming and statistical background.

4 Conclusion

The development of smart city technologies requires joint efforts by the academia, the industry and the government to provide evolving systems and services. In contrast to the past decade, nowadays "the leading urban centers are not placing their technological futures in the hands of a company or a single university research group". "Instead, they are relying on a combination of academics, civic leaders, businesses, and individual citizens working together to create urban information systems that could benefit all these groups" [50]. In recognition of the need for joint efforts, the US Government has recently announced a new "Smart Cities" Initiative that will invest over $160 million in federal research and leverage more than 25 new technology collaborations to improve city services [51]. We believe that sentiment analysis is a key factor in the development of all smart city domains.

5 References

[1] C. Harrison, B. Eckman, R. Hamilton, P. Hartswick, J. Kalagnanam, J. Paraszczak, and P. Williams, "Foundations for Smarter Cities," *IBM J. Res. Dev.*, vol. 54, no. 4, pp. 1–16, Jul. 2010.

[2] F. &amp; Sullivan, "Frost & Sullivan: Global Smart Cities market to reach US$1.56 trillion by 2020." [Online]. Available: http://www.prnewswire.com/news-releases/frost--sullivan-global-smart-cities-market-to-reach-us156-trillion-by-2020-300001531.html. [Accessed: 30-May-2016].

[3] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[4] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the 2nd international conference on Knowledge capture*, 2003, pp. 70–77.

[5] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, 2004, p. 271.

[6] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.

[7] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proceedings of the 20th international conference on Computational Linguistics*, 2004, p. 1367.

[8] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of the conference on human language technology and empirical methods in natural language processing*, 2005, pp. 347–354.

[9] A. Agarwal, F. Biadsy, and K. R. Mckeown, "Contextual Phrase-level Polarity Analysis Using Lexical Affect Scoring and Syntactic N-grams," in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, Stroudsburg, PA, USA, 2009, pp. 24–32.

[10] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," in *Proceedings of the 12th International Conference on World Wide Web*, New York, NY, USA, 2003, pp. 519–528.

[11] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment.," 2010.

[12] A. Ortigosa, J. M. Martín, and R. M. Carro, “Sentiment analysis in Facebook and its application to e-learning,” *Comput. Hum. Behav.*, vol. 31, pp. 527–541, Feb. 2014.

[13] D. Yang, D. Zhang, Z. Yu, and Z. Wang, “A Sentiment-enhanced Personalized Location Recommendation System,” in *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, New York, NY, USA, 2013, pp. 119–128.

[14] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, “A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle,” in *Proceedings of the ACL 2012 System Demonstrations*, Stroudsburg, PA, USA, 2012, pp. 115–120.

[15] M. Cheong and V. C. S. Lee, “A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter,” *Inf. Syst. Front.*, vol. 13, no. 1, pp. 45–59, Sep. 2010.

[16] B. Mandel, A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J. Rodrigue, “A Demographic Analysis of Online Sentiment During Hurricane Irene,” in *Proceedings of the Second Workshop on Language in Social Media*, Stroudsburg, PA, USA, 2012, pp. 27–36.

[17] M. Balduini, I. Celino, D. Dell’Aglio, E. Della Valle, Y. Huang, T. Lee, S.-H. Kim, and V. Tresp, “BOTTARI: An augmented reality mobile application to deliver personalized and location-based recommendations by continuous analysis of social media streams,” *Web Semant. Sci. Serv. Agents World Wide Web*, vol. 16, pp. 33–41, Nov. 2012.

[18] P. Hofmarcher, S. Theu\s sl, and K. Hornik, “Do Media Sentiments Reflect Economic Indices?,” *Chin. Bus. Rev.*, vol. 10, no. 7, 2011.

[19] A. D. I. Kramer, “An Unobtrusive Behavioral Model of ‘Gross National Happiness,’” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2010, pp. 287–290.

[20] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011.

[21] D. Osimo and F. Mureddu, “Research challenge on opinion mining and sentiment analysis,” *Univ. Paris-Sud Lab. LIMSI-CNRS Bâtim.*, vol. 508, 2012.

[22] J. Villena-Román, “TweetAlert: Semantic Analytics in Social Networks for Citizen Opinion Mining in the City of the Future.”

[23] B. Guthier, R. Alharthi, R. Abaalkhail, and A. El Saddik, “Detection and Visualization of Emotions in an Affect-Aware City,” in *Proceedings of the 1st International Workshop on Emerging Multimedia Applications and Services for Smart Cities*, New York, NY, USA, 2014, pp. 23–28.

[24] S. Siersdorfer, E. Minack, F. Deng, and J. Hare, “Analyzing and Predicting Sentiment of Images on the Social Web,” in *Proceedings of the 18th ACM International Conference on Multimedia*, New York, NY, USA, 2010, pp. 715–718.

[25] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, “Large-scale visual sentiment ontology and detectors using adjective noun pairs,” in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 223–232.

[26] D. Borth, T. Chen, R. Ji, and S.-F. Chang, “Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content,” in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 459–460.

[27] J. Yuan, S. Mcdonough, Q. You, and J. Luo, “Sentribute: image sentiment analysis from a mid-level perspective,” in *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, 2013, p. 10.

[28] Q. You, J. Luo, H. Jin, and J. Yang, “Robust image sentiment analysis using progressively trained and domain transferred deep networks,” *ArXiv Prepr. ArXiv150906041*, 2015.

[29] D. Ververidis and C. Kotropoulos, “Emotional speech recognition: Resources, features, and methods,” *Speech Commun.*, vol. 48, no. 9, pp. 1162–1181, 2006.

[30] D. Bitouk, R. Verma, and A. Nenkova, “Class-level spectral features for emotion recognition,” *Speech Commun.*, vol. 52, no. 7, pp. 613–625, 2010.

[31] F. Dellaert, T. Polzin, and A. Waibel, “Recognizing emotion in speech,” in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, 1996, vol. 3, pp. 1970–1973.

[32] R. Tato, R. Santos, R. Kompe, and J. M. Pardo, “Emotional space improves emotion recognition.,” in *INTERSPEECH*, 2002.

[33] M. El Ayadi, M. S. Kamel, and F. Karray, “Survey on speech emotion recognition: Features, classification schemes, and databases,” *Pattern Recognit.*, vol. 44, no. 3, pp. 572–587, 2011.

[34] L. C. De Silva, T. Miyasato, and R. Nakatsu, “Facial emotion recognition using multi-modal information,” in *Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on*, 1997, vol. 1, pp. 397–401.

[35] L. S. Chen, T. S. Huang, T. Miyasato, and R. Nakatsu, “Multimodal human emotion/expression recognition,” in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 366–371.

[36] N. Sebe, I. Cohen, T. Gevers, and T. S. Huang, “Emotion recognition based on joint visual and audio cues,” in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2006, vol. 1, pp. 1136–1139.

[37] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” *Pattern Anal. Mach. Intell. IEEE Trans. On*, vol. 31, no. 1, pp. 39–58, 2009.

[38] M. Wöllmer, B. Schuller, F. Eyben, and G. Rigoll, “Combining long short-term memory and dynamic bayesian networks for incremental emotion-sensitive

artificial listening," *Sel. Top. Signal Process. IEEE J. Of*, vol. 4, no. 5, pp. 867–881, 2010.

[39] I. D. Addo, S. I. Ahamed, and W. C. Chu, "Toward collective intelligence for fighting obesity," in *Computer Software and Applications Conference (COMPSAC), 2013 IEEE 37th Annual*, 2013, pp. 690–695.

[40] L.-P. Morency, R. Mihalcea, and P. Doshi, "Towards multimodal sentiment analysis: Harvesting opinions from the web," in *Proceedings of the 13th international conference on multimodal interfaces*, 2011, pp. 169–176.

[41] V. Pérez-Rosas, R. Mihalcea, and L.-P. Morency, "Utterance-Level Multimodal Sentiment Analysis.," in *ACL (1)*, 2013, pp. 973–982.

[42] R. Rawassizadeh, B. A. Price, and M. Petre, "Wearables: Has the age of smartwatches finally arrived?," *Commun. ACM*, vol. 58, no. 1, pp. 45–47, 2015.

[43] M. A. Pathak, *Beginning Data Science with R*. Springer, 2014.

[44] "Mining Twitter for Airline Consumer Sentiment | inside-R | A Community Site for R." [Online]. Available: http://www.inside-r.org/howto/mining-twitter-airline-consumer-sentiment. [Accessed: 30-May-2016].

[45] "Vik's Blog - Writings on machine learning, data science, and other cool stuff." [Online]. Available: http://www.vikparuchuri.com/blog/tracking-us-sentiments-over-time-in/. [Accessed: 30-May-2016].

[46] "Python Data Analysis Library — pandas: Python Data Analysis Library." [Online]. Available: http://pandas.pydata.org/. [Accessed: 30-May-2016].

[47] "Welcome to Apache^{TM} Hadoop®!" [Online]. Available: https://hadoop.apache.org/. [Accessed: 30-May-2016].

[48] V. N. Khuc, C. Shivade, R. Ramnath, and J. Ramanathan, "Towards building large-scale distributed systems for twitter sentiment analysis," in *Proceedings of the 27th annual ACM symposium on applied computing*, 2012, pp. 459–464.

[49] "SparkR (R on Spark) - Spark 1.6.1 Documentation." [Online]. Available: https://spark.apache.org/docs/latest/sparkr.html. [Accessed: 30-May-2016].

[50] G. Mone, "The new smart cities," *Commun. ACM*, vol. 58, no. 7, pp. 20–21, Jun. 2015.

[51] "FACT SHEET: Administration Announces New 'Smart Cities' Initiative to Help Communities Tackle Local Challenges and Improve City Services," *whitehouse.gov*, 2015. [Online]. Available: https://www.whitehouse.gov/the-press-office/2015/09/14/fact-sheet-administration-announces-new-smart-cities-initiative-help. [Accessed: 27-Sep-2015].

(SHELL) Smart Home for an Extraordinary Long Living

Riccardo Agostini [1] , Elia Brugnoni [1] , Eleonora Paganelli [1] , Alberto Polzonetti [2]

[1]E-Linking On Line System Camerino
[2] University of Camerino UNICAM Computer Science School

Abstract - *The idea behind this research is the identification of an integrated, modular and dry building system (SHELL) that maybe used to restore existing buildings, according to Smart Home and Assisted Living perspective with the aim of improving the quality of life of people with a reduced mobility, and helping them to manage their everyday life autonomously. After reviewing the state of the art has been carefully analyzed the object of research, the impact of the project in the local socio-economic context, and the risk assessment that such a technological innovation could have introduced. Finally, describes the three goals achieved: development of a modular, equipped cell featuring a steel structure that may be adjusted to the building; development and configuration of a mobile platform to access different floors within the home; development of an IT system to be integrated within a household appliance to help elderly or non-autonomous users. A critical analysis of the results concludes.*

Keywords: *Emerging technologies, Smart environments , Smart Cities*

1 Introduction

The idea behind this research is the identification of an integrated, modular and dry building system (SHELL) that may be used to restore existing buildings, according to Smart Home and Assisted Living perspective with the aim of improving the quality of life of people with a reduced mobility, and helping them to manage their everyday life autonomously. The proposed system features an integrated view of available technologies and their incorporation in residential architecture through a package tailored according to the building size and the needs of weaker and less autonomous population groups. Such a system is highly innovative not only in terms of design, but as regards the development process too.

Unlike what usually happens in building renovations nowadays, the SHELL system is designed and produced at a workshop/plant, with the help of workers and experts from different disciplines, so that it can be quickly installed within the building to be renovated. Therefore, long and costly construction processes can be avoided without wasting time and money. The prefabrication process starts from a defined idea based on a technology core which includes standard elements that characterize the domestic space (vertical and horizontal mobility, kitchen and bathroom) and on a customizable and adaptable system that may be changed in relation to the different circumstances and individual needs of each user.

In particular, SHELL is aimed at tackling the various issues that have become real emergencies over the years, which are:

- An increasingly old population;
- An increasing number of women have been entering the world of paid employment and also need to carry out multiple activities simultaneously at home (i.e. taking care of children, working, cooking, etc.) that require more monitoring and management abilities in terms of space and appliances;
- The aging and inadequacy of historical buildings and, more generally, of the urban fabric. The city and its assets (considered as a legacy) are not sustainable in the long term and, in addition to their structural and energy flaws, current buildings and constructions do not fulfill to today’s space habitability requirements, seen as responsiveness to the need for comfort and individual and family well-being;
- Mechanized horizontal and vertical mobility developed through the study of innovative solutions, borrowed from the commercial and industrial sectors (industrial lifts, hoists, moving sidewalks) and transformed through a kind of product design that may be aesthetically suitable for home interiors;
- Saving energy through wall relining based on a modular system of walls that come into contact with the outside environment, including windows that become sensors and help to control temperature inside the home.

1.1 State of the art

Some other experiments were carried out at national and European level based on a multidisciplinary approach focusing on three areas of application, which are:

I. Mobility technology allowing for an easier usage of home space. The very concept of home automation effectively implies the abolition or reduction of complex manual tasks within the living space, to make room for a kind of automation that may increase people’s awareness about the chores and tasks carried out at home in order to maximize the efficiency of available resources. Within such a context, it seems essential that a home designed for assisted living should provide an automated lifting system (when there are two or more floors), which may replace traditional staircases, but also be a conceptually alternative to traditional elevators. To such purpose we studied a solution that features the interaction between rooms on different floors through a "room lift" which moves vertically, that is visually open and provides maximum safety. The device originates from the need to eliminate closed and confined spaces (such as elevators) where people don't feel comfortable, as they may feel to be trapped in an unknown space, and that may trigger a claustrophobia feeling. The idea was inspired by a project for a disabled people home, which was developed in Bordeaux by OMA, a Dutch firm, [6], where the study area moves from the living room to the bedroom, which is located on a different floor. The process of transferring from a kind of technology designed for industrial use to a household application includes a number of steps that falls into two categories: 1) making the structure lighter; 2) designing the object in question so that it is suitable for a home, looks pleasant and, at the same time, is safe and easy to use.

II. Structure and interior design to integrate technologies and functions – devoted to improve mobility, to control devices, lighting, energy consumption, to manage home automation - within a system featuring a strong architectural and emotional connotation. Examples of this development trend from a recent past mainly relate to a devices remotely. One of the first researches on design applied to assisted living that, incidentally, was carried out in the Marche region, was summarized in a publication titled "Design olistico. Progettare secondo i principi del DfA" (by Andrea Lupacchini, from the Architecture &amp; Design School of Ascoli Piceno) [7]. The Design For All concept introduces the topic of inventing interior design devices that, together with home automation, may help people affected by disabilities to enjoy a home dimension that is more suitable to respond to their needs. The research purpose was to offer the tools needed to use architecture and design as social integration instruments. However, such studies have not tackled the issue of recovering existing buildings, since they are based on the assumption of working on newly built ones, where everything is conceived from scratch and different kinds of technology can be freely integrated without any physical constraints. The matter at hand involves reinterpreting the concept of wall units, traditionally seen as furnishing elements, to make them become truly functional structures devoted to a different use of home space, one that better suits the need of people with a reduced mobility, but also of those who are very busy at home and follow a multitasking pattern. The new home structure is shell shaped; it may adjust itself as a glove to the existing building and plays two different roles: 1) it works as an insulation lining; 2) it is also a "smart lining" equipped with tactile and sensory devices allowing for a different and innovative kind of home space management.

III. Interaction design and home automation for a simplified use of technology. There is a great deal of literature available on such an aspect of building development. In 2011, the Ambient Assisted Living Joint Programme (AAL JP) published a catalogue to present the goals and progress made by the project funded through European calls in 2008. [2,3,8,9] The authors have applied their know-how in the field of sensory applications to environment control and energy saving, in order to develop a user-friendly interface that perfectly fits the architecture and technology devices described above.

The ability of the SHELL system to provide an integrated and complex solution has been tested through a prototype, by simulating its operation inside an existing structure.

2 Innovation Feasibility And Risk Analysis

Risks related to technical, financial and regulatory aspects of platform lifts. The first one refers to the need to define a product architecture allowing for the development of a lift without the conventional cabin; or rather a lift where the cabin concept is replaced by a platform which doesn't have any ceiling and can travel without a closed cubicle. Suitable solutions need to be found in order to ensure safety, but also without obstructing the view too much. The second aspect concerns development costs. In fact, the product may be quite expensive due to the high degree of innovation involved, as opposed to the traditional and widely used lifts. The project team shall need to investigate on the most suitable technology to be used in order to reduce costs. A last but very important issue relates to the serious constraints of current European laws and regulations that only allow for very few deviations from traditional layouts. However, the platform lift prototype surely lays down the guide lines for innovative lifting systems. Risks related to the home automation system's ability to interact with its users. From a technology point of view, some of the possible risks that need to be kept into consideration as regards the development of a home automation system are linked to communication amongst different devices. The issues that are still hindering the development of the discipline in question are, primarily, the cabling and communication constraints of integrated systems used for home automation. Secondly, no agreement has been reached about the best solution for the control function architecture – i.e. whether it's better to have a centralized or a distributed one, with microchips placed on every system's node.

2.1 Risk related to the high technology costs.

As far as costs are concerned, the SHELL system was designed according to two different types of intervention:

- A high-technology core that foresees the use of customized elements, according to the user's level of independence. Several solutions will be available on the catalogue and their cost will be proportional to the quality and quantity of the options provided.

- An adjustable, easy to use, modular (cell) system, that is not expensive and may be quickly produced, which works as a link and interface with the technology core, and that should be integrated within the existing in structure to improve home climate and comfort.

3 THE PROJECT'S IMPACT ON ENERGY EFFICIENCY AND SUSTAINABLE DEVELOPMENT

The project's impact is mainly linked to the features of the "container" where the installation is going to take place. However, it is possible to refer to relevant literature mentioning an energy saving of about 30-40% after renovation projects focused on reducing energy consumption. Therefore, it is plausible to envisage that the combination of a building intervention and the automation of devices devoted to energy production or consumption may produce satisfactory results, which may be even higher than general average values.

Reduction of energy consumption after renovations, through an energy efficient cell: 30% Adoption of energy certification for prototype's performance, to be used as a reliable guide for the building renovation market, using the ITACA and EPBD protocol (European Energy Performance of Building Directive) Reduction of energy consumption through device control automation: 15% By interfacing heterogeneous building automation systems, SHELL may be widely used for both new project design (where building designers are not forced to use only single brand automation systems) and existing buildings (where existing systems and new building automation ones (BAS) need to be integrated). Reducing consumption and maximizing the comfort level of buildings are the goals to be pursued.

The SHELL's automation module ensures a centralized control and operation of all the building automation systems, such as access control, automation of home control elements, home climate control, ventilation, lighting, videosurveillance, wireless hot-spot, GPS real-time positioning, integration with ip-pbx, video entryphone, consumption control, especially focusing on energy saving.

Energy saving mainly depends on the building's construction features, as well as on the level of automation. Recent studies show a significant reduction of energy costs ranging from 5% to 15% and, in some cases, even 30-35%. As regards building automation, only in relation to energy saving, the breakeven point is usually reached between 2 and 10 years. Nevertheless, an ROI analysis depends on a number of factors. Besides energy saving, there are many benefits that BASs have to offer. One of the most tangible advantages is an increase in the building's value that may be 10 times higher than the value of a traditional building. The reasonable monthly costs of utilities are one of the benefits that may be immediately appreciated, and one which is not affected by the complex variables of the real estate market. When an efficient automation system is installed, the various home devices (lighting, air conditioning, ventilation, entry system, etc.) are only used when they're actually needed. Therefore, they last longer and maintenance, repair or replacement costs are remarkably reduced. The centralized control of such systems also allows for time savings lowering, therefore, the costs related to human resources devoted to carry out such activities. A further benefit that concerns the building's improved comfort leads to an increase in the productivity of people who live/work there. Moreover, the benefits related to special economic policies for eco-friendly initiatives should also be kept into account. [1,4,5]

From the mobility point of view, some significant reductions in energy consumption are going to take place in terms of lifting kinematics and a 20% reduction can be expected compared to traditional lifts. The platform lift will have to move slowly when travelling through 2/3 floors. Therefore, a maximum travel length of 6 metres and a target speed of 0.5 m/s may be foreseen.

The reduced travel length and speed allow for the installation of small lifting engines. This means that less power is going to be needed and, consequently, also the initial installation and management costs are going to be lower. Moreover, a power range which is compatible with the usual 3Kw of regular connections is certainly in line with the requirements of a senior user.

4 OBJECTIVES reached

As already mentioned, elderly population is progressively increasing and, at the same time, almost 80% of the population owns the home where it lives. Therefore, working on the existing buildings by applying innovative solutions seems to be necessary to give people who have been living in the same home for a long time the opportunity to have some home improvement works done, without being forced to give up their habits and neighbourhood relationships. In principle, the project recipients may be considered those belonging to a section of the elderly population whose quality of life and independence as regards daily chores would be greatly improved. In addition to them, other possible project recipients could be those families who see the organization of the home space as a fundamental support for their professional activities, and also as a guarantee for the safety of family members who have different needs.

Therefore, the initial project objectives included:

- ♦ The identification of innovative intervention methods that could integrate different technology and systems, and would provide an example to be replicated as regards the renovation of existing buildings;

- ♦ A higher performance level required by laws and regulations in terms of accessibility categories, living comfort, visitability, by showing an integrated vision that would fulfill some other requirements, in addition to those related to size or functionality, such as the interaction between space, objects and technology within the home environment;

- ♦ Combining the need for a different organization of the living space with energy saving, so that energy consumption could be reduced by 30% (by adopting the ITACA protocol's indicators), even though the level of technology and device performance is going to be higher;

- ♦ The development of a pilot product that would provide relevant companies the opportunity to interact in order to reach quality-cost- time ratio goals which may be competitive within their respective markets.

To such purpose, the following product has been developed

4.1 OR1: Development of a modular, equipped cell featuring a steel structure that may be adjusted to the building

The first objective concerns the design and development of a prototype for an equipped cell that may be adjusted to any kind of building, in order to optimize energy comfort and technology equipment and, consequently, to enhance the quality of living within the home environment. The cell has been conceived as a complex, basic element that may be standardized. It is composed of a modular structure and equipped with an interface system for both horizontal and vertical applications, in order to ensure that it may fully adjust to the existing building. The technology developed for wall and other system components (that is going to provide the cell's final look and remain visible inside the home) will be perfectly integrated within the home functions and will make the home interiors look warm and welcoming. The new structure, that allows for intervention on existing buildings, is going to fulfill several objectives, such as: a) combining lightness, modularity and adaptability so that it can be installed inside existing buildings; b) optimizing energy saving by acting on external components (window frames and walls), while operating them from inside; c) integrating technology systems and sensors that should manage thermal comfort, ventilation and natural lighting, and work as collectors for IT and mobility systems; d) ensuring acoustic and visual comfort without changing the property's homely character or interfering with the integration of furnishing accessories, such as curtains, pictures and/or paintings, electronic appliances and wall lighting items.

4.2 OR2: Development and configuration of a mobile platform to access different floors within the home

The second objective aimed at studying, designing and developing a prototype of a platform that may help elderly or generally "weak" users to go up and down two/three floors inside their home. So, the project is about a platform lift to be installed together with the equipped cell. It looks pleasant and is as big as a small room whose walls may be pulled out when it's being used, so that they won't obstruct the overall view of the home. The platform can be installed inside the home, so that it may easily be accessed and used, without looking oppressive and bulky. The platform's mechanical and engineering components are housed within the equipped cells that are installed on different floors. Such a device could fully replace existing staircases that may be used just in case of emergency. The equipped cell shall house the control and management modules for the platform.

The main technical solutions that were conceived and developed for the platform lift are:

- ♦ Development of a concept for a platform lift equipped with retractable walls whose size and location may be customized. The platform is going to have parapet walls made of an opaque or transparent material, which come out when the platform goes up or down. The platform's size may vary and its wall and/or floor anchoring systems may be built in different ways.

- ♦ Development of a moving platform featuring stable and smooth operation. The mechanical components include rails, pulleys, ropes, hoists and hydraulic cylinders that allow for a smooth and trouble-free movement. The operating speed needs to be low, and the base positioning precise, so that getting in and out of the platformshould be easy as there are no dangerous steps to prevent that. The platform's acceleration shall not produceany speed surges, so that users won't feel uncomfortable or worried when using it.

- ♦ Development of a highly interactive and easily accessible control dashboard. The platform shall be equipped with a dashboard to be used to select floors and operate the platform. The dashboard shall be very easy to read and highly interactive. Touch operation and voice activation systems may also be evaluated.

- Development of a CAD-based software to design and set up both platform and cell was also used to draw the platform's geometry within the cell and helped to select and place the modules to be installed (lighting, air conditioning, "Active Diary" module, etc.). In particular, it was very useful to define the building's architectural features, the platform's size and positioning and its interconnection with the cell's

4.3 OR3: Development of an IT system to be integrated within a household appliance to help elderly or non-autonomous users.

An IT system has been developed that may be integrated within a household appliance in order to help elderly or non-autonomous users. Such a system helps users by guiding them through the daily house chores and their communication with the outside environment, in order to prevent their isolation. The software is composed of two sub-systems. The first one deals with home automation and is devoted to ensure home living comfort, the smooth running of appliances, safety and, consequently, energy saving. The second sub-system, which may be called "Active Diary", was conceived to manage interpersonal communication and to follow the activities and needs of users, as proactively as possible. The application's main features consist of the following modules: booking medical consultations and requesting prescriptions, asking for assistance, online shopping, surveillance, SOS alerts, diary, home automation, condominium social network.

The main technical solutions that were conceived and developed for the IT system are:

- ♦ Analysis and mapping of home parameters to be controlled;
- ♦ Setting up the parameters for environment monitoring;
- ♦ Alarm system management;
- ♦ Energy saving management;
- ♦ Development of an "Active Diary" system enabling users to book medical consultations and request prescriptions, ask for assistance, shop online, check home safety on a tablet/smart phone, sending SOS alerts, take part in social activities – e.g. condominium social network.-

5 Conclusions

The SHELL research project, which was aimed at introducing the development of an integrated system for the renovation of existing buildings focusing on assisted living, has enabled the participating companies to widen their business scope. The product in question, which includes a group of products that may also be taken out or added according to non-standard options, allowed the companies involved to enter a new market where they are going to meet new competitors.

The results to be obtained are twofold since each company will benefit from:

- ♦ The development of innovative products made of iron and aluminum leading to more opportunities to take advantage of the available economies of scale. A market share increase and a geographical expansion can be expected for such companies;
- ♦ The development of innovative products in the field of lifting equipment leading to the discovery of new market slices, which may not necessarily be limited to people with disabilities;
- ♦ The definition of new smart home model equipped with suitable technology customized for different users.

On the other hand, a highly innovative product may become well known, subject to the ability of the companies involved to interact and use their expertise in order to grab the renovation market slice, with a special focus on historical buildings located in city centers, where interventions are often exceedingly invasive since new technologies are directly placed inside existing buildings and walls.

The platform lift integrated within the SHELL cell enables companies to diversify their product offering. In fact, the platform lift provides a solution that enhances both actual and perceived safety, as well as size and energy performances. It took approximately one year to manufacture the new products, including new technical solutions, component engineering for scale production and sourcing of new components. In particular, the following tasks and issues had to be tackled:

- ♦ Contacting suppliers to get information on costs and quantities for the new components according to the expected sale volume;
- ♦ Testing the trial components sent by suppliers, i.e. ropes, pulleys, electrical boards, data transmission systems, batteries, video cameras, plastic and metal components;
- ♦ Carrying out additional testing using a larger number of users, also in order to check regulatory and safety aspects;
- ♦ Optimizing the production cycle and defining the production layout for the new product line;

- ♦ Detailed calculations of all the costs (materials, labour, machinery depreciation, production consumption, etc.) and production timing in order to set an appropriate selling price and delivery times for the orders received.).

6 References

[1] Agarwal, Y., Balaji, B., Gupta, R., Lyles, J., Wei, M., &amp; Weng, T. (2010, November). Occupancy-driven energy management for smart building automation. In Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building (pp. 1-6). ACM.

[2] Cavallo, F., Pujol, L., Garcia, A., &amp; Dario, P. (2010). The AALIANCE research agenda on ICT for ageing well. Gerontechnology, 9(2), 181-182.

[3] Gorman, J., Mikalsen, M., Stav, E., &amp; Walderhaug, S. (2010). universAAL–European Commission collaborative research and development to develop an open architecture and platform for Ambient Assisted Living (AAL). Gerontechnology, 9(2), 183-184.

[4] Ippolito, M. G., Sanseverino, E. R., &amp; Zizzo, G. (2014). Impact of building automation control systems and technical building management systems on the energy performance class of residential buildings: An Italian case study.Energy and Buildings, 69, 33-40.

[5] Yang, R., &amp; Wang, L. (2012). Multi-objective optimization for decision-making of energy and comfort management in building automation and control. Sustainable Cities and Society, 2(1), 1-7.

[6] Lampariello, B. (2011). Villa a Floirac. Rem Koolhaas/OMA 1994-98 (Quaderni di laurea. DIPSA). Aracne.

[7] Lupacchini, A. (2010). Design olistico. Progettare secondo i principi del DfA. Alinea Editrice.

[8] Olsson, S. (2010). Ambient Assisted Living Joint Program: A European wide initiative. Gerontechnology, 9(2), 182.

[9] Wintlev-Jensen, P. (2010). Key European research and innovation initiatives: Addressing new technologies and services for ageing well. Gerontechnology,9(2), 180

SESSION

BIG DATA ANALYTICS AND APPLICATIONS

Chair(s)

TBA

PolyEHR: A Framework for Polyglot Persistence of the Electronic Health Record

André Magno Costa de Araújo[1], **Valéria Cesário Times**[1], **Marcus Urbano da Silva**[1]
[1]Center for Informatics, Federal University of Pernambuco, Recife, Brazil

Abstract - *Building data schemas for storage in the Electronic Health Record (EHR) has traditionally been done using a single data model. This practice increases the complexity in application development due to the heterogeneity of data in the health care sector, which consequently makes the data schema rigid. In addition, an approach that stores EHR from applications built from heterogeneous database archetypes is lacking or unknown. This article presents a framework that builds application templates for the health case sector using archetypes and storing EHR data in heterogeneous databases through polyglot persistence. To generate templates and data schemas, we extract three elements from archetypes: data attributes that define the EHR, terminologies and vocabulary which give the clinical data a semantic meaning as well as constraints specified on data attributes. Finally, we demonstrate the creation of health applications templates using archetypes and EHR storage in heterogeneous databases.*

Keywords: Applications in healthcare, frameworks for Big Data, Archetypes, Health Information Systems.

1 Introduction

The creation of data schemas for storage in the Electronic Health Record (EHR) is a relevant theme when considering the life cycle of a Health Information System (HIS) and is the object of study in several research papers [1,2]. According to ISO/TS 18308 [3], an EHR data schema must store all relevant clinical events that shall be used in patient's care, including textual descriptions, numeric values, logical values, date and time expressions and hierarchical data structures. Furthermore, it must store data in tables in a way that the relationships between columns and rows are preserved and allow data storage by pairs of attributes (e.g. key-value).

As indicated in [4], the concept of archetypes and templates proposed in openEHR represent an important standard in health care. It minimizes the problems of modeling heterogeneity of EHR data and facilitates the standardization of terminologies and constraints for a given health care sector. An archetype may be defined as a computer expression represented by domain-specific constraints that model and add semantic meaning to the EHR, while templates are graphical user interfaces (GUI) automatically generated and based on archetype specifications [5].

Traditionally, a single data model has been used to represent the different types of EHR data. This practice is present in data schemas designed with or without the use of archetypes [6, 7]. Due to the variety of data types (i.e. structured and non-structured) found in health care sectors, the use of a single storage model might increase the complexity in HIS development, resulting in rigidity in the data schema (i.e. an alteration in the data schema depends on a software development team) and compromises the application performance in data processing.

In the health sector, storage in heterogeneous databases is one alternative to represent the data diversity of several HIS applications (e.g. diagnosis and therapy, clinical and assistance, supplies and revenue management). In that sense, the concept of polyglot persistence proposes a mixed approach for combined storage – the consistent characteristic of relational data models, alongside the flexibility of NoSQL data models (i.e. key-value, document and graph). The core idea is to store structured data using a relational approach, while semi-structured or non-structured data are stored in NoSQL data models.

Polyglot persistence has been applied in a variety of applications, such as IBM's auto scaling PaaS architecture [8], source-code-based data schemas identifications tools [9] and the re-engineering of legacy systems for heterogeneous databases [10]. As pointed out in [11], polyglot persistence is new, promising and evolving. Thus, one may notice that some research projects are dedicated to the creation of new storage architectures for HIS [11, 12], adapting to a new reality in heterogeneous database storage.

Although polyglot persistence has been debated and applied in the health care sector, an approach that stores EHR from applications built from heterogeneous database archetypes is lacking or unknown. There are three main advantages in the use of polyglot persistence for EHR storage created from archetypes. First of all, the data schema is created from a health standard that makes data attributes, terminologies and constraints uniform. Secondly, heterogeneous data can be stored in different data models, for example, the hierarchical data of a lab exam result can be organized as a document set in a NoSQL database, while the patient's demographic data can be stored in a relational data schema. Finally, it minimizes problems caused by the constant changes in data schema. Here, the idea is to store the clinical EHR data that undergo the most changes in flexible data models (e.g. key-value, document and graph), while other data that suffer less alterations are stored in relational data schemas.

This paper proposes a framework named PolyEHR, which builds health application templates using archetypes and store EHR data in heterogeneous databases by means of polyglot persistence. To generate templates and data schema, PolyEHR extracts the following elements from archetypes; i) EHR data attributes; ii) terminologies and vocabulary that give semantic meaning to clinical data, and iii) constraints specified over data attributes. As the main contributions of the present work, we highlight; i) the specification of an architecture that demonstrates how PolyEHR creates templates for health applications and persists data in heterogeneous databases; ii) the creation of a data schema from archetypes attributes, terminologies and constraints; iii) the development of a mechanism for automatic generation of templates using archetypes (i.e. Graphical user interfaces); iv) implementation of a REST API to make the data manipulation process transparent in the multi-model storage architecture; and finally, v) the demonstration of template generation based on public domain archetypes from the openEHR repository.

The remaining sections of this work are organized as follows: Section 2 contextualizes the concept of archetypes and polyglot persistence used in this paper, and surveys related works. Section 3 describes the characteristics and functionalities of the framework hereby proposed and demonstrates the construction of templates from extracted archetypes. Finally, the conclusion is found in Section 4.

2 Background and related work

This section contextualizes the basic concepts of archetypes (Section 2.1) and polyglot persistence (Section 2.2) and presents a comparative analysis of related works.

2.1 Archetypes and templates

The specification of archetype-based EHR consists in organizing its components through a two-level modeling approach [5]. The first, called information, represents the stable level of the model from which systems and software components can be built (e.g. XML schema, UML or OWL). At this level, data have no semantic characteristics and only the data types that are chosen to represent them are known. The second level called knowledge consists of domain-driven definitions represented as archetypes and templates.

An archetype consists of a computational expression based on a reference model and represented by domain constraints and terminologies [4] (e.g. data attributes of a blood test). A template is a structure used to group archetypes and allow their use in a particular context of application. It is often associated with a graphical user interface (e.g. a GUI used by a professional for defining the elements of a leukogram list, such as leukocytes and neutrophils). Dual modeling is the separation between information and knowledge of health care system architectures. In this approach, the components responsible for modeling the clinical and demographic data of EHR are specified through generic data structures, which are composed of data types, constraints and terminologies.

In an archetype, the specification of attributes is achieved through data entry builders named generic data structures. Such structures allow the representation of EHR data heterogeneity through the following types: ITEM_SINGLE, ITEM_LIST, ITEM_TREE and ITEM_TABLE.

ITEM_SINGLE models a single data attribute such as a patient's weight, height and age. ITEM_LIST groups a set of attributes in a list. A patient's address containing number, street and zip code for example. ITEM_TREE specifies a hierarchical data structure that is logically represented as a tree. It can be used, for instance, to model a patient's physical or neurological evaluations. Finally, ITEM_TABLE models data elements by using columns for field definition and rows for field value respectively. Each attribute of a data structure is characterized by a type of data and can have a related set of associated domain restrictions and terminologies. The terminologies give semantic meaning to clinical data and can be represented as a set of health terms defined by a professional.

2.2 Polyglot persistence

Polyglot Persistence defines the use of different data storage technologies to deal with different storage needs [10]. Different types of data are better represented by different storage approaches. The core idea in using polyglot persistence is the storage of structured data through a relational approach, while semi-structured or non-structured data is stored in NoSQL data models. Figure 1 shows how the different types of data from a health care sector can be stored using polyglot persistence.

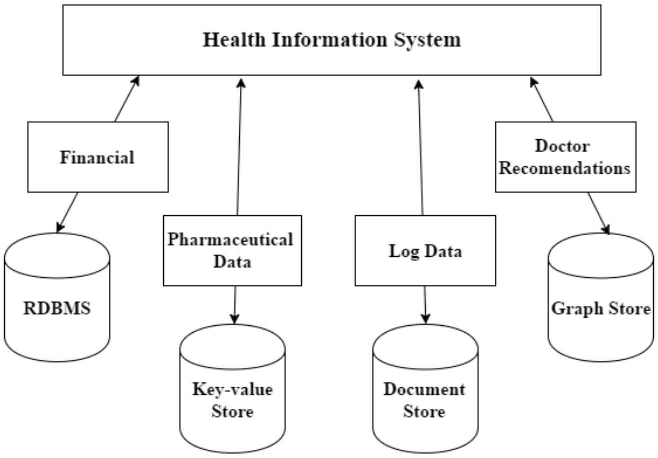


Figure 1. Example of multi-model storage in the health care sector

As shown in Figure 1, the NoSQL approach offers different data models. Among the most common are: Key-value store, Document store e Graph Store.

Key-value store: Stores information in a table, with rows, keys and value. The key field displays the information description while the value field denotes the information itself.

Document store: Stores document sets. A set consists of a collection of fields that can be displayed as a single element, a list, or nested documents.

Graph store: It makes use of nodes, relationships and properties to store information. The node represents the vertices in a graph, the relationships, its edges and the properties, the attributes.

The NoSQL data models have characteristics that differentiate them from more traditional approaches. Such characteristics offer a more adequate support for non-conventional data storage and more flexibility when creating or altering a data schema.

2.3 Related work

Based on the state-of-the-art works reviewed, we present an analysis of the main related works in the fields of i) archetype mapping and persistence in databases, ii) generation of templates from archetypes, and iii) the use of polyglot persistence in health applications.

One of the pioneering works in the field of archetype mapping was developed by Späth and Grimson (2010) [13]. They mapped a set of archetypes for a legacy database and exposed the lack of tools and methodologies that would have helped in modeling archetypes in a database. Using an automatic approach, Georg et al. [14] specified a set of rules to map archetype data attributes in a relational database, generating templates in a specific problem domain. However, the proposed solution could not map archetypes with hierarchical data structures (i.e. ITEM_TREE).

The openEHR foundation offers a free solution in which the data attributes of an XML archetype are serialized in a Database Management System (DBMS). Node+Path [15] uses Entity- attribute-value (EAV) approach to store the path (i.e. address) of an attribute in the first column, while the value of such attribute is stored in the second column. Yet, this solution requires the creation of complex logical sentences for data manipulation that may compromise the performance of the application. The persistence solution proposed by Wang et al. (2015) [16], specifies a set of mapping rules, extracts the archetypes attributes and stores them in tables of a relational data schema. However, the terminologies and constraints specified in those archetypes are not considered.

The EHRScape[1] and EHRGen[2] frameworks support the health application development process by using specifications from openEHR and generate templates and data schema from archetype mapping. However, they use a single data model to organize EHR data.

To minimize the rigidity caused by relational data schema and provide support to the continuous data requirement changes which commonly occur in legacy HIS, Prasad and Sha (2013) [12] specify an architecture and a HIS prototype that allows polyglot persistence and improves health data management in a legacy application. Similarly, Kaur and Rani (2015) [10] specify a polyglot storage architecture to store structured data in a relational database (i.e. PostgreSQL), while two NoSQL databases (i.e. MongoDB e Neo4j) store semi-structured data, such as laboratorial exams and medicine prescriptions. Nevertheless, neither solutions of polyglot persistence use archetypes to standardize EHR data attributes and terminologies.

Table 1 depicts the main characteristics of related works and the framework we are proposing. In this assessment, we have compared the following aspects A1) Reads an archetype and outputs an open standard; A2) generation of data schema from archetypes; A3) polyglot persistence of archetypes; A4) template generation; and A5) HIS architecture for heterogeneous databases.

Table 1. Comparative analysis of related work

Work/Criterion	A1	A2	A3	A4	A5
Späth and Grimson (2010)	✗	✗	✗	✗	✗
Georg et al.(2013)	✗	✗	✗	✓	✗
openEHR Node+Path	✓	✗	✗	✗	✗
Wang et al.(2015)	✗	✗	✗	✗	✗
EHRScape Framework	✓	✓	✗	✓	✗
EHRGen Framework	✗	✓	✗	✓	✗
Prasad and Sha (2013)	✗	✗	✗	✗	✓
Kaur and Rani (2015)	✗	✗	✗	✗	✓
PolyEHR Framework	✗	✓	✓	✓	✓

Only the Node+Path solution [15] and the EHRScape framework offer resources that allow mapped archetype attributes to persist in other solutions, whereas data schema creation from archetypes is only present in EHRScape, EHRGen and PolyEHR frameworks. The EHR polyglot persistence resource built from archetypes is possible only in PolyEHR framework, while template generation is possible in EHRScape, EHRGen, PolyEHR and in the solution proposed by Georg et al. Finally, PolyEHR and the works of Prasad and Sha (2013) [12] and Kaur and Rani (2015) [10] specify HIS architectures that exemplify polyglot persistence.

As shown in Table 1, the main motivation for the proposed solution is to develop a framework capable of building health application templates, generating data schemas in heterogeneous databases using archetypes and storing EHR data through polyglot persistence.

3 The PolyEHR framework

This section describes the proposed archetype-based framework and is organized as follows: Section 3.1 details the storage architecture in heterogeneous databases, while the main features of PolyEHR are presented in Section 3.2. Section 3.3 describes how the archetypes are displayed in heterogeneous data models. Finally, Section 3.4 shows the generation of templates and exemplifies the multi-model storage through polyglot persistence.

3.1 Architecture and overview

The framework proposed in this paper is built for health care applications and consists in a computing environment dedicated to the construction of templates from the specifications existent in EHR archetypes. The core feature consists in extracting the elements from archetypes for persistence in different data approaches, as well as the automatic generation of GUIs. Figure 2 illustrates the components and architecture relationships designed for PolyEHR.

[1] https://www.ehrscape.com/

[2] https://code.google.com/p/open-ehr-gen-framework/

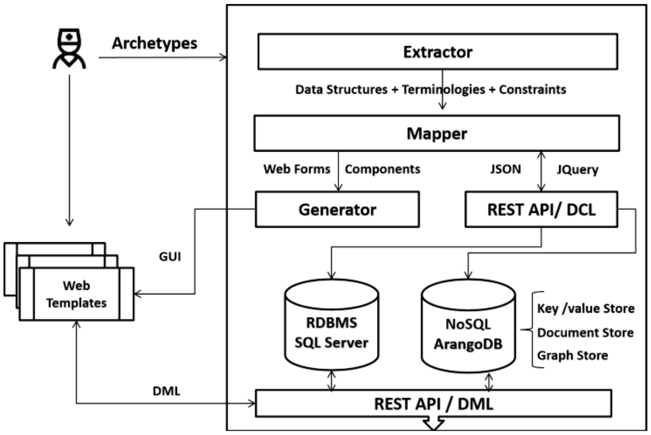


Figure 2. Framework architecture for multi-model storage.

One of the main features of the framework is to help health professionals build templates specific to their sector. As shown in Figure 2, the framework generates data schemas and GUI by reading an archetype imported by the user.

The *Extractor* component shown in Figure 2 extracts from the informed archetype the attributes that define the EHR, the health care terminologies and vocabulary that give a semantic meaning to the clinical data as well as constraints specified in the attributes. With the extracted elements, the *Mapper* component performs the following operations: i) transforms the attributes into data entry fields in the GUI; ii) uses the constraints extracted from archetypes as data entry validation mechanisms (e.g. range of values, data type restrictions); and iii) provides the terminologies extracted from archetypes to give a semantic meaning to their respective GUI fields. Afterwards, the *Generator* component groups and organizes the created elements, subsequently providing a usable template.

After the *Extractor* component obtains the attributes, terminologies and constraints, the REST API / DCL dynamically creates data schemas in the databases, both relational and NoSQL. Every GUI created by the framework has data manipulation capabilities (i.e. insert, update, query, and delete). Since the framework storage is based on heterogeneous data models, we specify a REST API that makes the process of data manipulation transparent to the user. Thus, all data persistence processes carried out in the template generated by the framework are managed by the REST API / DML, as shown in Figure 2.

3.2 Main functionalities

In addition to the data schemas generation feature in heterogeneous databases, we provide a host of features to help a health care professional use the templates generated by our framework. Such features are shown in Figure 3 and detailed below:

Demographic information management:

In order to reduce the demographic information redundancy in the generated data schema, we provide the management of health care organizations, patients, doctors, nursing staff and system users. Based on the type of organization (e.g. hospital, clinic, clinical laboratory, basic health unit), the framework generates instances of applications, i.e. an instance of a dedicated application for the admission of patients at a given hospital, or an instance of the application focused on outpatient and emergency care. In addition, information from patients, doctors and nursing staff can be integrated with features generated from the archetypes.

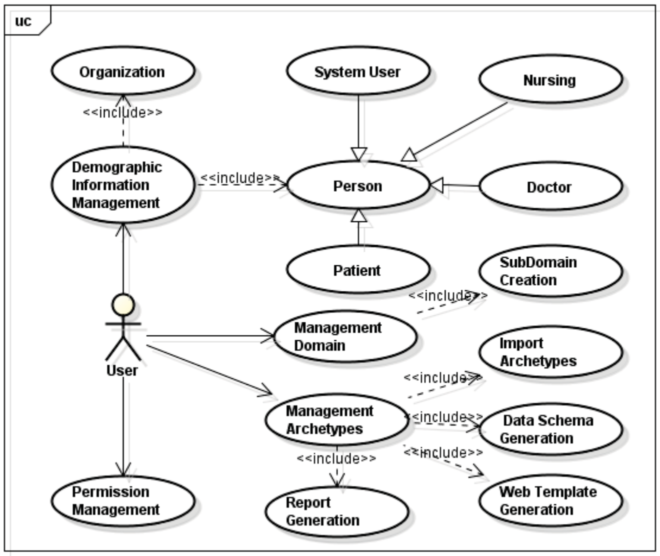


Figure 3. Main functionalities of the PolyEHR framework

Health care sector management: An organization can offer different types of health services. For example, a hospital may perform laboratory testing services, diagnostic imaging, emergency care, hospitalization, etc. Considering that, we have made it possible to create and configure domains and sub-domains that represent the services offered by each organization. In order to facilitate its use, the features generated by the framework are grouped into themes. Additionally, these features can also create a sequence of activities that shall be performed by health professionals; for example, an application may have a domain called *Nursing*, where professionals in this area have a determined drill to carry out, including assessment of patients' vital signs, water balance and physical examination.

Archetype management: This functionality allows one to import and map the archetypes that will be used to generate templates for a health application. When importing an archetype, the framework allows the professional to choose which elements will be part of the data schema and manage the GUI elements by adding, removing or disabling fields. Such elements can be modified at a later time. In this case, the framework will automatically extend the database schema created. After this activity, the user can then associate the generated GUI to a domain or sub-domain application and manage user and organization access permissions for each functionality. In addition to the GUI automatic generation, a mechanism allows users to create their own reports from the information stored in the data schema.

3.3 Persistence of archetypes in heterogeneous data models

The polyglote persistence proposed is this framework is performed as follows; data of a structured nature such as demographic information is stored in a relational database (i.e. SQL Server), while non-structured data such as a laboratory exam is stored in a NoSQL database. The generation of NoSQL data schemas is based on the type of data structures found in archetypes, such as ITEM_SINGLE, ITEM_LIST, ITEM_TREE and ITEM_TABLE.

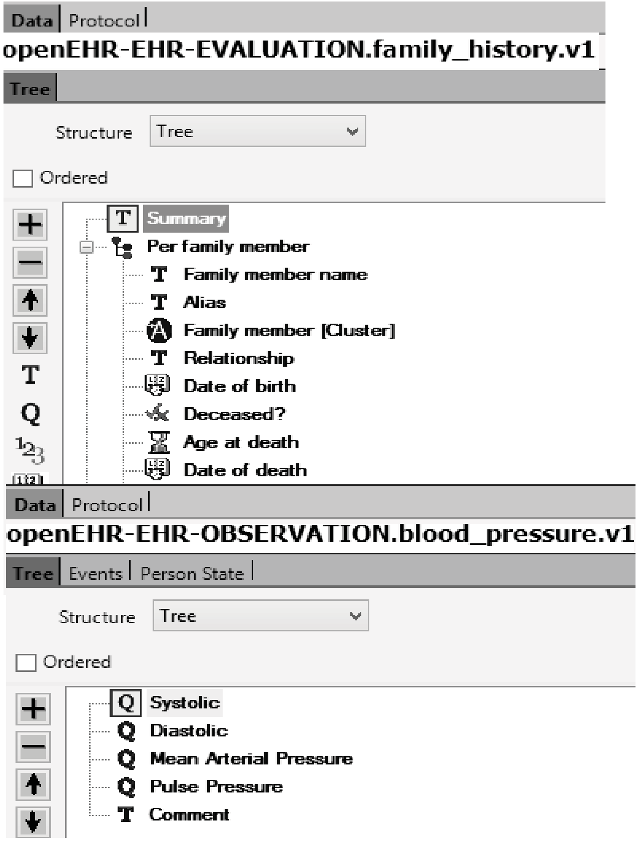
As the name suggest, ITEM_SINGLE displays a single attribute, while ITEM_LIST groups a set of attributes into a vertical list. Due to the nature of these items, we propose the key-value data model to generate data schemas. ITEM_TREE displays data attributes in a hierarchical structure containing the several levels of an archetype. In order to maintain such data organization, the framework uses document data model (i.e. collections). Finally, ITEM_TABLE organizes the data attributes in a table made of rows and columns. For this type of storage, we propose the graph data model, where data found in rows are represented as properties, data from columns as vertices, and attribute description as the relationship between the two.

3.4 Generating templates for health applications

In this section, we demonstrate how the framework generates templates and exemplify how the data is persisted in heterogeneous databases through polyglot persistence. Here, we will use the family history and blood pressure archetypes to generate templates and data schemas. The blood pressure archetype was chosen to demonstrate that we obtained the same results using EHRScape and EHRGen, as well as the approach proposed by Georg et al. [14]. The family history archetype was in turn chosen because it contains a significant amount of attributes represented by a multilevel hierarchical structure. Both these archetypes are shown in Figure 4 and are available in the openEHR repository [17].

We initially registered a fictitious health facility named *Model Hospital* along with some patients, doctors, nursing staff and system users to manipulate data in the templates. Before importing the family history and blood pressure archetypes, we created a domain named *Ambulatory* and two sub-domains named *Family Background* and *Patient assessment*, which will then be linked to their respective GUIs.

Afterwards, the user can choose to import the archetypes that will be used for the generation of data schemas and GUIs. To customize importation, we have developed a feature that allows the user to choose archetypal elements, and visualize how they were mapped by the framework with their respective data types. Furthermore, data input for each field can be made mandatory or not.



Data | Protocol

openEHR-EHR-EVALUATION.family_history.v1

Tree

Structure: Tree

□ Ordered

- Summary
- Per family member
  - Family member name
  - Alias
  - Family member [Cluster]
  - Relationship
  - Date of birth
  - Deceased?
  - Age at death
  - Date of death

Data | Protocol

openEHR-EHR-OBSERVATION.blood_pressure.v1

Tree | Events | Person State

Structure: Tree

□ Ordered

- Systolic
- Diastolic
- Mean Arterial Pressure
- Pulse Pressure
- Comment

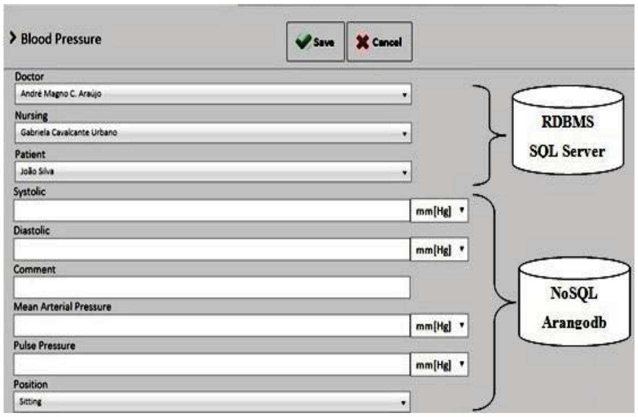
Figure 4. Data attributes of family history and blood pressure archetypes [17]

Once the user inputs all information, the framework dynamically generates the data schemas from the elements extracted from the archetypes (i.e. attributes, terminologies and constraints) following the criteria described in Section 3.3. To enable the functionality which was just generated, the user simply needs to associate the GUI to a domain and sub-domain previously configured in the framework.

To the GUI generated from the blood pressure archetype, we added demographic information managed by the framework. Therefore, beyond the chosen archetypal elements, the GUI will have the following additional fields: the name of the doctor responsible for clinical care, the nurse responsible for the exams and the actual patient.

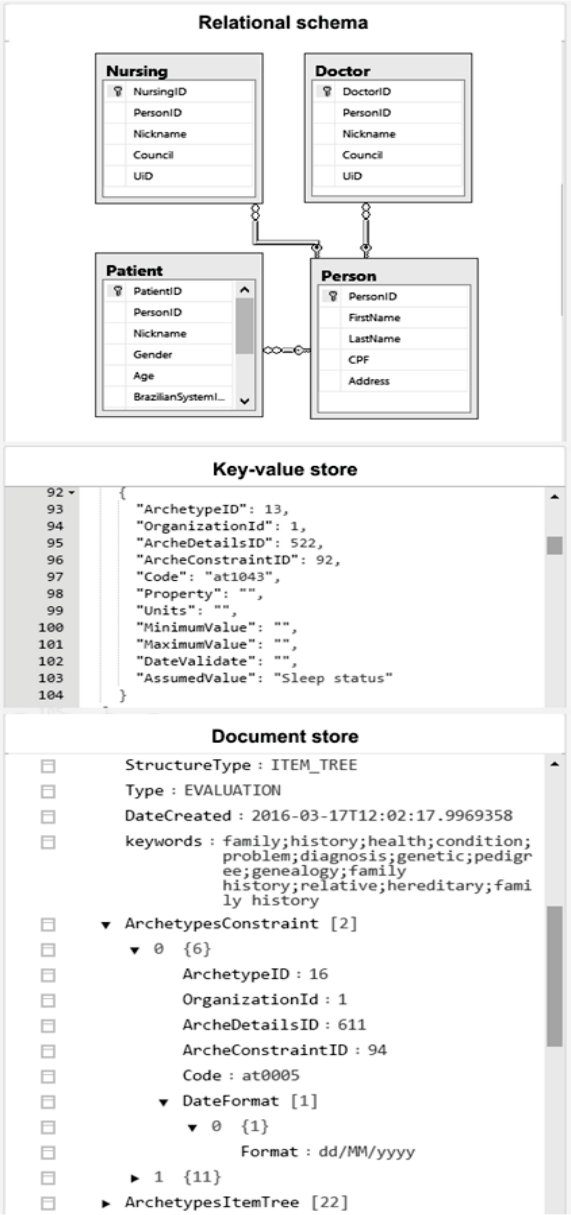
Figure 5 depicts the GUI created from the blood pressure archetype. One may observe that the demographic information (i.e. Doctor, Nursing and Patient) is stored in a relational database, while the other archetype fields are stored in a NoSQL database. Since the structure that defines the attributes of the blood pressure archetype is an ITEM_LIST type, the framework persists the data in a key-value database.

Figure 6 shows parts of the relational and NoSQL data schema (key-value and document) used to store the data templates generated from the blood pressure and family history archetypes.



Doctor: André Magno C. Araújo
Nursing: Gabriela Cavalcante Urbano
Patient: João Silva
Systolic: mm[Hg]
Diastolic: mm[Hg]
Comment:
Mean Arterial Pressure: mm[Hg]
Pulse Pressure: mm[Hg]
Position: Sitting

Figure 5. Example of a template generated by our framework based on the blood pressure archetype



```
{
  "ArchetypeID": 13,
  "OrganizationId": 1,
  "ArcheDetailsID": 522,
  "ArcheConstraintID": 92,
  "Code": "at1043",
  "Property": "",
  "Units": "",
  "MinimumValue": "",
  "MaximumValue": "",
  "DateValidate": "",
  "AssumedValue": "Sleep status"
}
```

```
      StructureType : ITEM_TREE
      Type : EVALUATION
      DateCreated : 2016-03-17T12:02:17.9969358
      keywords : family;history;health;condition;
                 problem;diagnosis;genetic;pedigr
                 ee;genealogy;family
                 history;relative;hereditary;fami
                 ly history
    ▾ ArchetypesConstraint [2]
        ▾ 0  {6}
            ArchetypeID : 16
            OrganizationId : 1
            ArcheDetailsID : 611
            ArcheConstraintID : 94
            Code : at0005
          ▾ DateFormat [1]
              ▾ 0  {1}
                  Format : dd/MM/yyyy
        ▸ 1  {11}
    ▸ ArchetypesItemTree [22]
```

Figure 6. Heterogeneous database storage

As the framework generates templates that store data in a heterogeneous database, we had to develop a mechanism to dynamically create the data manipulation resources for each GUI generated. Such resources are input, edit, print, delete and retrieve information from the databases. Here, the REST API encapsulates the heterogeneity of manipulation languages used by the adopted data models (i.e. SQL and AQL.) The template shown in Figure 5 and all other features described in this article are available at archeweb.dotdesign.com.br.

4 Conclusion

In this paper we presented a framework capable of building application templates from archetypes to be used in the health care sector, generating archetype-based data schemas in heterogeneous databases and storing EHR data through polyglot persistence.

We developed an approach to generate graphical user interfaces with data manipulation capabilities and specified a mechanism to extract attributes, terminologies and constraints from archetypes and allow the dynamic persistence of data in heterogeneous databases. To help a health care professional use the templates generated, PolyEHR provides a host of features such as the management of health care sector, archetypes and demographic information. In addition, all data persistence processes carried out in the template generated by PolyEHR are managed by a REST API. Finally, we demonstrated the generation of health application templates based on public domain archetypes from the openEHR repository.

Acknowledgment

This work was partially supported by Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), under the grants APQ-0173-1.03/15 and IBPG-0809-1.03/13.

5 References

[1] K. k. Lee, W. Tangb, K. Choia. “Alternatives to relational database: Comparison of NoSQL and XML approaches for clinical data storage”; Computer Methods and Programs in Biomedicine, pp. 99-109, 2013.

[2] V. Dinu, P. Nadkarni. “Guidelines for the Effective Use of Entity-Attribute-Value Modeling for Biomedical Databases”; International Journal of Medical Informatics. pp. 769-779, 2007.

[3] International Organization for Standardization: ISO/TS 18308 health informatics - requirements for an electronic health record architecture, available at http:// www.iso.org/iso/iso_catalogue.htm.

[4] Marco E., Thomas A., Jorg. R, Asuman D., Gokce L. “A Survey and Analysis of Electronic Healthcare Record Standards”; ACM Computing Surveys, pp. 277–315, 2005.

[5] D. Lloyd, T. Beale, S. Heard. “openEHR Architecture: Architecture Overview”; available at http://www.openehr.org/releases/1.0.2/architecture/overview.pdf, last accessed September 2015.

[6] Jumaa H, Rubel P, Fayn J. “An XML-based framework for automating data exchange in healthcare”; IEEE International Conference on e-Health Networking Applications and Services (Healthcom), pp.264 – 269, 2010.

[7] Bernstein K, Bruun R. M, Vingtoft S, Andersen S. K, Nøhr C. “Modelling and implementing electronic health records in Denmark”; International Journal of Medical Informatic, pp. 213-220, 2005.

[8] Seetharami R. Seelam, Paolo Dettori, Peter Westerink, Ben Bo Yang. “Polyglot Application Auto Scaling Service for Platform As A Service Cloud”; IEEE International Conference on Cloud Engineering, 2015.

[9] Martyn Ellison, Radu Calinescu. “Richard Paige, Re-engineering the Database Layer of Legacy Applications for Scalable Cloud Deployment”; IEEE/ACM 7th International Conference on Utility and Cloud Computing, 2014.

[10] Karamjit Kaur, Rinkle Rani. “Managing Data in Healthcare Information Systems: Many Models, One Solution”; IEEE Computer Society, 2015.

[11] Rami Sellami, Sami Bhiri, Bruno Defude. “Supporting multi data stores applications in cloud environments”; IEEE Transactions on Services Computing, 2015.

[12] Srikrishna Prasad, Nunifar Sha, “NextGen Data Persistence Pattern in Healthcare: Polyglot Persistence”; Fourth International Conference on Computing, Communications and Networking Technologies, 2013.

[13] Späth M. B., Grimson J. “Applying the archetype approach to the database of a biobank information management system”; International Journal of Medical Informatics, pp. 1-22, 2010.

[14] D. Georg, C. Judith, R. Christoph. “Towards plug-and-play integration of archetypes into legacy electronic health record systems: the ArchiMed experience”; BMC Medical Informatics and Decision Making, pp. 1-12, 2013.

[15] Node + Path Persistence, available at https://openehr.atlassian.net/wiki/pages/viewpage.action?pageId=6553626, last accessed April 2016.

[16] Wang L, Min L, Lu X, Duan H. “Archetype relational mapping - a practical openEHR persistence solution”; BMC Medical Informatics and Decision Making, pp. 1-18, 2015.

[17] Archetype Editor, available at www.openehr.org, last accessed April 2016.

Towards a Reference Model for Advanced Visual Interfaces Supporting Big Data Analysis

Marco X. Bornschlegl*, Kevin Berwind*, Michael Kaufmann†, Matthias L. Hemmje*
*University of Hagen, Faculty of Mathematics and Computer Science, 58097 Hagen, Germany
{marco-xaver.bornschlegl, kevin.berwind, matthias.hemmje}@fernuni-hagen.de
†Lucerne University of Applied Sciences and Arts, Engineering & Architecture, 6048 Horw, Switzerland
m.kaufmann@hslu.ch

Abstract—**This paper introduces an approach to develop an up-to-date reference model that can support advanced visual user interfaces for distributed Big Data analysis in virtual labs to be used in e-Science, industrial research, and data science education. The paper introduces and motivates the current situation in this application area as a basis for a corresponding problem statement that is utilized to derive goals and objectives of the approach. Furthermore, the relevant state-of-the-art is revisited and remaining challenges are identified. An exemplary set of use cases, corresponding user stereotypes as well as a conceptual design model to address these challenges are introduced. Conclusions and an outlook on future work complete the paper.**

Keywords: Advanced Visual User Interfaces, Distributed Big Data Analysis, Information Visualization, User Empowerment, Virtual Research Environments

I. Introduction and Motivation

The availability of data has changed dramatically over the past ten years. The wide distribution of web-enabled mobile devices and the evolution of web 2.0 technologies are contributing to a large amount of data (so-called Big Data) [1]. Usable access to complex and large amounts of data poses, e.g., an immense challenge for current solutions in, e.g., business analytics. Handling the complexity of relevant data (generated through information deluge and being targeted with Big Data technologies) requires new techniques about data access, visualization, perception, and interaction for innovative and successful strategies. As a consequence research communities as well as industry, but especially research teams at small universities and **Small and Medium-sized Enterprises (SMEs)**, will be faced with enormous challenges. Furthermore, current e-Science research resources and infrastructures (i.e., data, tools, and related **Information and Communication Technology (ICT)** services) are often confined to computer science expert usage only and fail to leverage the abundant opportunities that distributed, dynamic, and eventually interdisciplinary **Virtual Research Environments (VREs)** can provide to scientists, industrial research users as well as to learners in computer science, data science and related educational environments.

The overall goal of this research is to develop a reference model that can support advanced visual user interfaces for distributed Big Data analysis in virtual labs to be used in e-Science, industrial research, and data science education. The surrounding infrastructure will support the life cycle of VREs by enabling the dynamic ad-hoc definition of new interdisciplinary research projects within advanced visual user interfaces supporting cognitive efficiency as well as user empowerment.

II. State of the Art

Information Visualization (IVIS) has emerged *"from research in human-computer interaction, computer science, graphics, visual design, psychology, and business methods"* [2]. Nevertheless, IVIS can also be seen as a result of the question for interchanging ideas and information between human, keeping with Rainer Kuhlen [3], because of the missing direct way. The most precise and common definition of IVIS as *"the use of computer-supported, interactive, visual representations of abstract data to amplify cognition"* stems from Card et al. [4]. To simplify the discussion about information visualization systems and to compare and contrast them, Card et al. [4] defined a reference model, which is illustrated in Figure 1, for mapping data to visual forms for human perception.

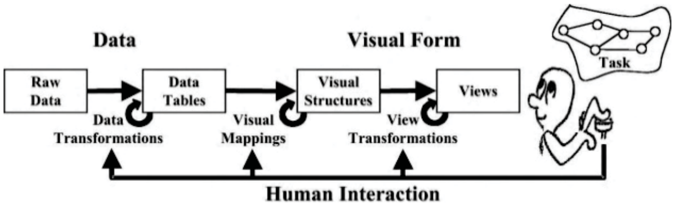


Fig. 1. IVIS Reference Model [4]

In this model, arrows lead from **Raw Data** to visual data presentation of the raw data within a cognitive efficient IVIS based on a **Visual Structure** and its rendering of a view that is easy to perceive and interact with for humans. The arrows in this model indicate a series of data transformations whereas each arrow might indicate multiple chained transformations. Moreover, additional arrows from the human at the right into the transformations themselves indicate the adjustment of these transformations by user-operated controls supporting human-computer interaction [4]. **Data Transformations** map raw data such as text data, processing information, (database-) tables, e-mails, feeds and sensor data into **Data Tables** which define the data with relational descriptions and extended meta data [1], [5]. **Visual Mappings** transform data tables into visual structures, that combine spatial substrates, marks and graphical properties. Finally, **View Transformations** create views of the visual structures by specifying graphical parameters such as position, scaling and clipping [4]. *"Although raw data can be visualized directly, data tables are an important step when the data are abstract, without a direct spatial component"* [4]. Therefore, Card et al. define the mapping of a data table to a visual structure, i.e. a visual mapping, as the core of reference model, this operation translates the mathematical relations within data tables to **Graphical Properties** within visual structures.

Current development leads to a continuous growth of both computer systems and end-user population [6]. Fischer [7] emphasizes that *"people and tasks are different."* Moreover, he explains that humans start from a partial specification of a task, and refine it incrementally, on the basis of the feedback that they get from their environment. Thus, *"users must be able to articulate incrementally the task at hand. The information provided in response to these problem-solving activities based on partial specifications and constructions must assist users to refine the definition of their problem"* [7]. Moreover, all stakeholders of an interactive system, including end users, are "owners" of a part of the problem: Software engineers know the technology, end users know the application domain, human-computer interaction experts know human factors, etc.; *"they must all contribute to system design by bringing their own expertise"* [6].

Kaufmann [8] describes **Big Data Management (BDM)** as *"the process of optimally controlling flows of large-volume, high-velocity, heterogeneous and uncertain data."* Usually, Big Data is approached with a technological focus, but a key question for businesses is how Big Data can be effectively used to create value. Therefore in this model, as illustrated in Figure 2, the classical technological aspects of Big Data (cf., e.g., Singh and Reddy [9]), namely **Data Integration** (i.e. Hadoop [10] clusters NoSQL databases and other database management systems) and **Data Analytics** (i.e. statistical, machine learning and data mining tools and techniques) are complemented by three additional layers, called **Data Interaction**, **Data Effectuation** and **Data Intelligence**, to improve the effective benefit of Big Data technologies.

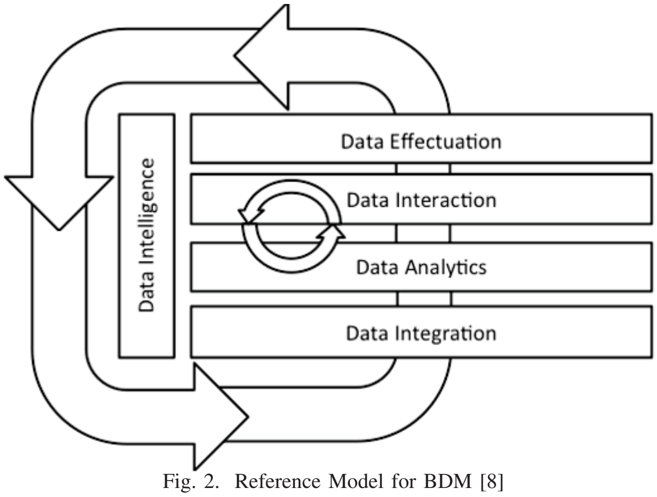


Fig. 2. Reference Model for BDM [8]

On top of the two basic layers, which are focusing on data technology, there is a layer called **Data Interaction** to illustrate the importance of human-computer interaction in the process for BDM, which is defined as a key aspect in this model. In this step, human decision makers are getting in touch with analysis results to view, manipulate, correct and communicate them. Conversely, data are getting in touch with the users of the information system, creating a bi-directional interaction between technology and its users, as symbolized by the small feedback loop. Nevertheless, interacting with analysis results is not sufficient to create actual value from data. For the optimization of business objectives, according to Davenport [11], it is necessary that the effects of data analysis are integrated with products and services of an organization. Therefore, another layer called **Data Effectuation** is defined, in which value creation from data is addressed. Finally, the model describes a cross-sectional function of knowledge-based processes and technologies which support the big data management life cycle. Value creation from data depends highly on emergent knowledge processes in the organization (Markus, Majchrzak, and Gasser [12]; Patel and Ghoneim [13]). The layer called **Data Intelligence** ought to ensure that knowledge and skills can be acquired and properly managed in the context of Big Data management. Kaufmann [8] proposes to apply KM and **Knowledge Engineering (KE)** techniques to all layers of the BDM from integration to effectuation, for example, by archiving and semantically annotating analysis results, by optimally communicating data-based insights, or by actively managing data science know-how.

III. Conceptual Modeling

As illustrated in Figure 3, Big Data analysis is based on different perspectives and intentions. To support management functions in their ability of making sustainable decisions, Big Data analysis specialists are filling the gap between Big Data analysis result consumers and Big Data technologies. Thus, these specialists need to understand their consumers/customers intentions as well as a strong technology watch, but are not the same like developers, because they care about having impact on the business [14].

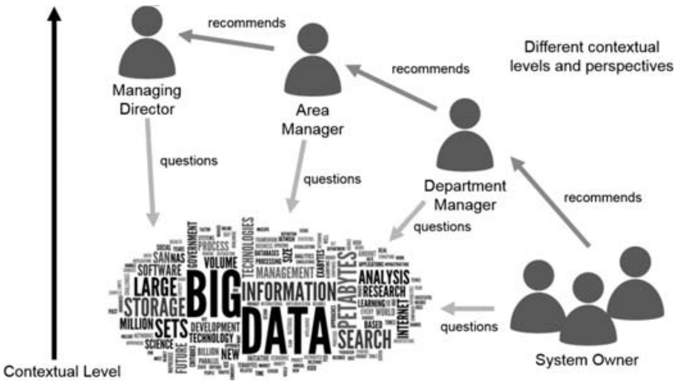


Fig. 3. Big Data Perspectives in Industry

Deduced from this perspectives and intentions, there are different use cases and related user stereotypes that can be identified for performing Big Data analysis collaboratively within an organization. Users with the highest contextual level, like e.g. managers of different hierarchy levels of such organizations, need to interact with visual analysis results for their decision making processes. On the other hand, users with low contextual levels, like system owners or administrators, need to interact directly with data sources, data streams or data tables for operating, customizing or manipulating their systems. Nevertheless, user stereotypes with lower contextual levels are interested in visualization techniques as well, in case these techniques are focusing on their lower contextual levels. Finally, there are user stereotype perspectives in the middle of those excesses, representing the connection between user stereotypes with low and high contextual levels. As a consequence from these various perspectives and contextual levels, it is important to provide the different user stereotypes a context aware system for their individual use cases. *”The ’right’ information, at the ’right’ time, in the ’right’ place, in the ’right’ way to the ’right’ person“* [15]. One could add to this citation *”with the right competences“* and/or *”with the right user empowerment“*.

As a response to increased graphics performance in computing technologies and information visualization, Card et al. [4] developed the IVIS reference model. Due to further developments in information systems as well as knowledge management systems in recent years, this model has to be adapted for covering the recent advancements. Modern cloud technologies and distributed computing are leading to almost unlimited storage and computing performance. Moreover, usable access to complex and large amounts of data over several data sources requires new techniques for accessing data and visualizing data with innovative and successful strategies, at the border between automated data analysis and enterprise decision making [16]. Not in alignment to these new required techniques, the original information visualization reference model transforms data from a single data source on the left directly to a visual representation for the end user on the right, without a direct view and interaction possibility on the single process stages. Thus, our hybridly refined and extended **IVIS4BigData** reference model (cf. Figure 4), an adaptation of the IVIS reference model, in combination with Kaufmann’s BDM reference model [8] is achieved to cover the new conditions of the present situation with advanced visual interface opportunities for perceiving, managing, and interpreting Big Data analysis results to support insight. Integrated into the underlying reference model for BDM, which illustrates different stages of BDM, the adaptation of the IVIS reference model represents the interactive part of the BDM life cycle.

According to Card et al. arrows which indicate a series of (multiple) data transformations lead from raw data to data presentation for humans. However, instead of collecting raw data from a single data source, multiple data sources can be connected, integrated by means of mediator architectures, and in this way globally managed in **Data Collections** inside the **Data Collection, Management & Curation** layer. The first transformation, which is located in the **Analytics** layer of the underlying BDM model, maps the data from the connected data sources into **Data Structures**, which represent the first stage in the **Interaction & Perception** layer. The generic term **Data Structures** also includes the use of modern **Big Data Storage Technologies** (like, e.g., NoSQL, RDBMS, HDFS), instead of using only data tables with relational schemata. The following steps **Visual Mappings**, which transforms data tables into **Visual Structures**, and **View Transformations**, which creates **Views** of the **Visual Structures**, by specifying graphical parameters such as position, scaling, and clipping, do not differ from the original IVIS reference model. As a consequence, only interacting with analysis results leads

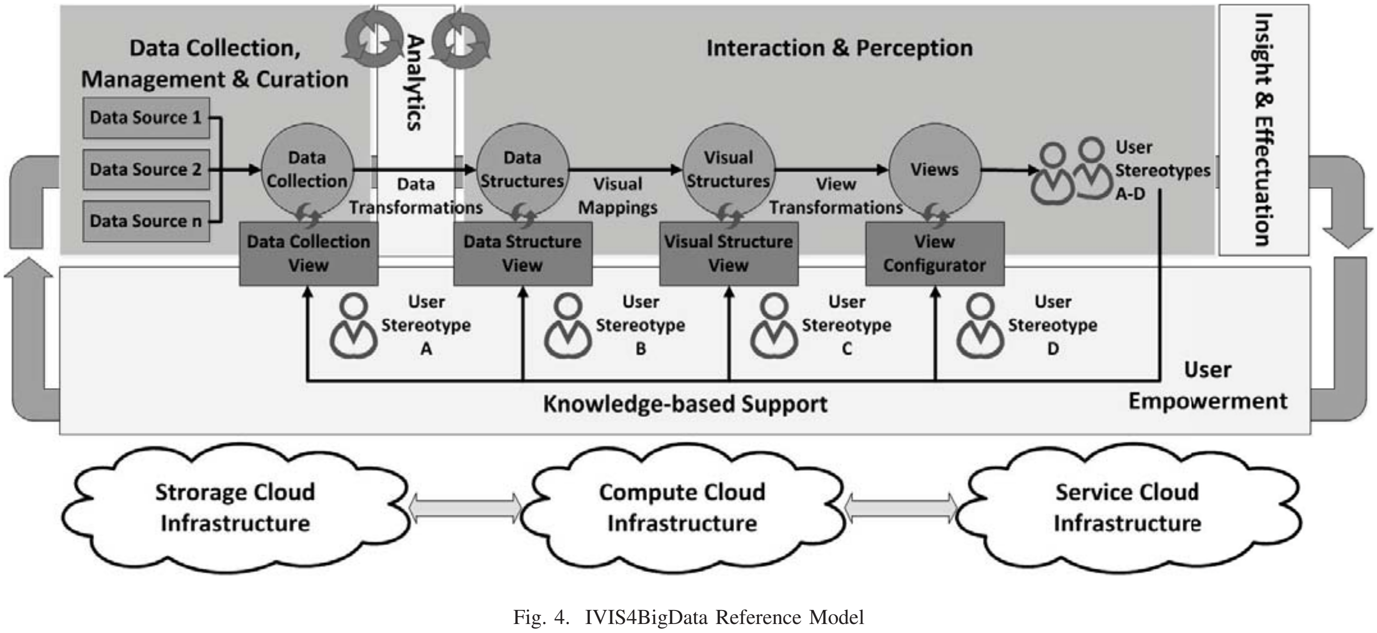


Fig. 4. IVIS4BigData Reference Model

not to *”added value“* for the optimization of, e.g., research results or business objectives. Furthermore, no process steps are currently located within the **Insight & Effectuation** layer because such *”added value“* is rather generated from knowledge, which is a *”function of a particular perspective“* [17] and will be generated within this layer by combining the analysis results with existing knowledge.

The major adaptations are located between the cross-functional **Knowledge-Based Support** layer and the corresponding layers above. As a consequence from the various perspectives and contextual levels of Big Data analysis and management user stereotypes, additional arrows lead from the human users on the right into multiple **Views**. These arrows are illustrating the interaction between user stereotypes with single process stages and the adjustments of the respective transformations by user-operated controls to provide *”the ’right’ information, at the ’right’ time, in the ’right’ place, in the ’right’ way to the ’right’ person“* [15], within a context aware and user-empowering system for individual use cases. Finally, the circulation around the whole layers clarifies that IVIS4BigData is not solely an one time process, because the results can be used as the input for a new process circulation.

IV. Discussion and Outlook

The use case scenarios, user stereotypes, as well as the conceptual model were already validated during presentations and discussions in an expert round table with experts from European e-Science and e-Infrastructure research institutions during the EGI community forum conference 2015 [18]. The evaluation approach of the developed model will be conducted with two different strategies. The first strategy will use case study research to answer the needs of Big Data user stereotypes. In addition to the case study evaluation, the second evaluation strategy will use the outcomes of existing research literature of the underlying models. Based on their scientific popularity, the advantages and disadvantages will be compared in conjunction to the characteristics of the designed model to demonstrate its usability. Furthermore, the model will be implemented and integrated in a proof-of-concept implementation paving the way towards user experiments with the EGI technologies and infrastructures, based on the EGI Federated Cloud infrastructure (Fed-Cloud [19]).

However, what is still missing and can be considered as a ***remaining challenge*** for our work, is an integration of the BDM model with semantic integration and IVIS reference infrastructures as well as an integration of visual analytics support into such a hybrid and extended model for supporting advanced visual interfaces for Big Data analysis in VREs. Furthermore, from the point of view of user empowerment, an identification of use cases and corresponding user stereotypes for utilizing such advanced visual interfaces for Big Data analysis in VREs remains as an additional challenge. In the remainder of this paper we will therefore work on the establishment of an identification of initial use cases, an initial identification of user stereotypes for user empowerment related to these use cases, as well as the design of a hybrid model integrating the BDM reference model with the IVIS reference model and visual analysis support.

Acknowledgment and Disclaimer

This publication has been produced in the context of the EDISON project. The project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 675419. However, this paper reflects only the author’s view and the European Commission is not responsible for any use that may be made of the information it contains.

References

[1] J. Freiknecht, *Big Data in der Praxis*. München, Deutschland: Carl Hanser Verlag GmbH & Co. KG, 2014.

[2] J. J. Thomas, K. Cook *et al.*, “A visual analytics agenda,” *Computer Graphics and Applications, IEEE*, vol. 26, no. 1, pp. 10–13, 2006.

[3] R. Kuhlen, *Informationsethik: Umgang mit Wissen und Information in elektronischen Räumen*, ser. UTB / UTB. UVK-Verlag-Ges., 2004.

[4] S. K. Card, J. D. Mackinlay, and B. Shneiderman, Eds., *Readings in Information Visualization: Using Vision to Think*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999.

[5] C. Beath, I. Becerra-Fernandez, J. Ross, and J. Short, “Finding value in the information explosion,” *MIT Sloan Management Review*, vol. 53, no. 4, p. 18, 2012.

[6] M. Costabile, P. Mussio, L. Parasiliti Provenza, and A. Piccinno, “Supporting end users to be co-designers of their tools,” in *End-User Development*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009, vol. 5435, pp. 70–85.

[7] G. Fischer and K. Nakakoji, “Beyond the macho approach of artificial intelligence: empower human designers - do not replace them,” *Knowledge-Based Systems*, vol. 5, no. 1, pp. 15 – 30, 1992.

[8] M. Kaufmann, “Towards a reference model for big data management,” 2016, research Report, forthcoming.

[9] D. Singh and C. K. Reddy, “A survey on platforms for big data analytics,” *Journal of Big Data*, vol. 2, no. 1, pp. 1–20, 2014. [Online]. Available: http://dx.doi.org/10.1186/s40537-014-0008-6

[10] Apache Software Foundation, “Hadoop,” accessed: 2016-01-10. [Online]. Available: https://hadoop.apache.org

[11] T. H. Davenport, “Analytics 3.0,” https://hbr.org/2013/12/analytics-30, Dec 2013, accessed: 2016-01-05.

[12] M. L. Markus, A. Majchrzak, and L. Gasser, “A design theory for systems that support emergent knowledge processes,” *MIS Q.*, vol. 26, no. 3, pp. 179–212, Sep. 2002. [Online]. Available: http://dl.acm.org/citation.cfm?id=2017167.2017170

[13] N. V. Patel and A. Ghoneim, “Managing emergent knowledge through deferred action design principles: The case of ecommerce virtual teams,” Bingley, pp. 424–440, 2011.

[14] S. Upadhyay and R. Grant, “5 data scientists who became ceos and are leading thriving companies,” http://venturebeat.com/2013/12/03/5-data-scientists-who-became-ceos-and-are-leading-thriving-companies, Oct 2013, accessed: 2015-10-30.

[15] G. Fischer, “Context-aware systems: The ’right’ information, at the ’right’ time, in the ’right’ place, in the ’right’ way, to the ’right’ person,” in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ser. AVI ’12. New York, NY, USA: ACM, 2012, pp. 287–294.

[16] Fraunhofer Institute for Computer Graphics Research IGD, “Visual business analytics,” 2015, accessed: 2015-12-02. [Online]. Available: http://www.igd.fraunhofer.de/en/Institut/Abteilungen/Informationsvisualisierung-und-Visual-Analytics/Visual-Business-Analytics

[17] I. Nonaka and H. Takeuchi, *The knowledge-creating company: how Japanese companies create the dynamics of innovation*. Oxford University Press, 1995.

[18] M. L. Hemmje, H. Brocks, and J. Becker, “Demand of data science skills & competences (expert roundtable),” *EGI Community Forum*, 2015.

[19] EGI Foundation - EGI.eu, “Egi federated cloud,” accessed: 2016-01-10. [Online]. Available: https://www.egi.eu/infrastructure/cloud

Cyber Computing Culture – Smart Cyber University

V. Hahanov[1], **Wajeb Gharibi**[2], **S. Chumachenko**[1], **and E. Litvinova**[1]
[1]Computer Engineering Faculty, National University of Radioelectronics, Kharkov, Ukraine
hahanov@icloud.com, litvinova_eugenia@mail.ru, ch_s_v@mail.ru
[2] Computer Science Department, Jazan University, Jazan, KSA
gharibiw2002@yahoo.com

***Abstract** – Cyber physical system Smart Cyber University (CyUni) is proposed. It is characterized by the presence of a metric space of regulatory rules, digital monitoring and active cyber management of addressable components of scientific and educational processes, automatic generation of operational and regulatory actions, human free cyber decision making focused on management of financial and human resources. The main components of CyUni are smart infrastructure, personnel, relationship, management, roadmap, and resources, represented in a cyberspace to perform scientific and educational processes. CyUni allows improving the quality of educational services and scientific achievements of high school through the use of the metric system of relationship defining the rules of digital monitoring and active cloud cyber management of scientific and educational processes.*

Keywords: research and education cyberspace; cyber-physical system; smart cyber university, big-data driven SIMD-multiprocessor.

1 Introduction

The goal of the CyUni-project is improving the quality of educational services and scientific achievements of higher education through the development of the metric relationship defining the rules of human-free digital monitoring and active cloud cyber management of scientific and educational processes. The global objectives are: 1) development of a standard for metric evaluating science and education of universities; 2) development of cloud services for human-free digital monitoring and active cyber management of scientific and educational processes; 3) creation of secured e-document that eliminates paper carriers through the use of digital signatures, ID-card, E-mail and mobile phones; 4) development of mobile e-voting for monitoring public opinion, election of experts, heads and professors for vacancies; 5) assessment of the quality of educational processes and departments, and also on-line knowledge testing that eliminates an illegal relationship between teacher and student; 6) metric management of scientific processes based on digital assessment of research activities, departments and scientific results for human-free distribution of the financial resources between departments and employees; 7) implementation of the cloud services Smart Cyber University in scientific and educational space of the EU and new independent states (NIS); 8) destruction of corruption in NIS by creation of human-free cloud cyber management of resources and personnel. The project objectives will be achieved by: 1) writing the metric evaluation standard of research and education based on the integration of rating rules of the world leading universities; 2) programming cyber services for human-free mobile cloud monitoring and management; 3) digitization of relationship between departments, students and staff, and also metric measurement of processes and phenomena at the university; 4) the implementation of a digital signature in the e-document by using an ID-card, E-mail and mobile phone.

CyUni – Smart Cyber University (SCU) is proposed as the metric culture of cyber-social relations, creating a network of personnel and smart infrastructure for digital monitoring and cloud management of scientific, educational processes and resources in order to achieve high quality of science, education and the life of employees (Fig. 1).

2 State of the art

There exist smart things, houses and cities, which use the Internet of Things for monitoring and managing phenomena, but there are no smart universities, research and educational processes [1-6]. The cyberspace of the planet has been created and passively used for monitoring and representation of all the physical, biological and social processes and phenomena in the clouds, where there is no active management of the society [7-11].

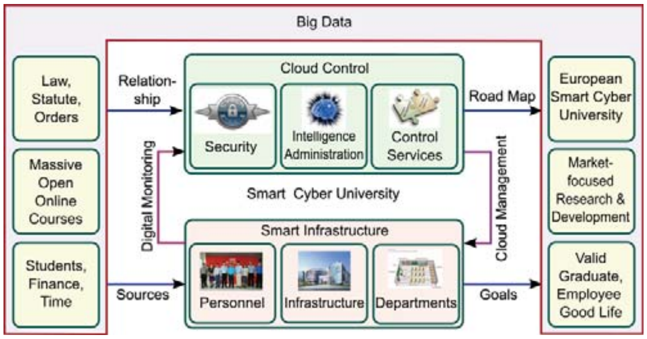


Figure 1. Smart Cyber University

Electronic document flow has been developed in the form of transactions of electronic documents to inform citizens and does not provide digital monitoring, measurement and actuation of social processes available for human-free cyber management [12-14]. There exist driver-free vehicles, aircrafts and factories, automatically controlled by computers and clouds, but there are not cloud human-free monitoring and management of social groups and state structures closed in the loop [15-20]. E-voting in decision making and E-election of officials by usual citizens have just been appeared, but they are not supported by digital metric for measuring objects, available for the corresponding cyber services, that causes of social disasters [21-24]. There are metric for evaluation of research, educational processes and academic professors to generate ratings, but don't exist adequate cyber services for cloud distribution of moral and material rewards that causes corruption [25-29]. Reliable and secure cloud services for storage and data transactions, digital signatures and electronic keys have been created, but the universities spend tons of paper, destroying the forest [15-18, 30]. Scientists at the universities have created a cyber ecosystem of the planet, smart cities, global cloud services, but there doesn't exist the cloud human-free cyber service for closed in the loop digital monitoring and metric management of science, education, personnel, students, resources and infrastructure [31-35].

3 The feasibility of the SCU

The main definitions are represented below [36-40]. Smart is the definition of the kind of process or phenomenon associated with the network interaction of addressable system components in time and space between themselves and the environment based on SCU artificial intelligence to achieve their goals. Cyberspace is a network of addressable and metrically interacting digitized processes and phenomena in the global telecommunication and computer infrastructure with distinct functions of monitoring, computing, storage, transaction and management to achieve defined goals. E-document management is legitimate and intelligent transactions of digitized document flows (sensor and regulation signals) in smart logically distributed data network designed to implement paperless relationship with the outside environment, digital monitoring and cloud management of the scientific and educational processes and university departments. Competence is a metric assessment of the spiritual, physical, emotional, intellectual and professional culture of the individual, which determines its value for possible leveraging the knowledge, skills and abilities in the performance of the social role aimed at improving the life quality and preserve the ecosystems of the planet. Relationship (Legislation) is a network of social and technological relations between officials, staff and departments, forming system structure of the university for monitoring and management of scientific and educational processes, personnel, infrastructure, financial and time resources based on existing laws, statutes, regulations, orders, traditions. The metric of relationship is digitized set of information and actuation documents (orders, regulations, statutes, traditions and laws) that defines the basis of interaction, operational and strategic digital monitoring and cloud management of processes, personnel, infrastructure, financial and timing resources aimed at creating an external image and internal moral and ethical climate. The relationship in high school is the main system-creating component, providing the successful of the university in the education market. Everything else: goals, personnel, infrastructure, management, science and education are directly depending on the relationship. There are three axioms of legitimate cyber relationship: 1) the head looks at the employee through the metric of his creative activity, valuable for the university, and stimulates according his achievements; 2) the scientist receives services from the manager that provide career growth, moral and material compensation; 3) a scientist looks at his colleague in the light of achievements, which cause admiration and an example to follow.

4 The ground-breaking nature of the research

1) Instead of passive IT-monitoring it is proposed human-free active IoT-management in the digitized cyber physical space, based on fog-networks of sensor-actuators, big data analytics, Google cloud service and quantum computing [41, 42]. 2) The fact – measurement – evaluation – action is a cyclical format of the SCU cyber service, related to digital monitoring and management, which is based on a postulate: "No measure – no control." 3) Creation of competency metrics for the rating processes and phenomena, based on expert digital evaluation metrics of employees, departments, scientific and educational processes, is proposed. 4) Use of accumulative competence matrices of processes and phenomena in the transparent cyber distribution of moral and material rewards for departments and employees according to their ratings is offered. Cloud cyber services of decision-making are focused on high performance analysis of big data by using multiprocessor-driven intelligent filters of metric relationship [36], eliminating the direct participation of officials and leaving them only decorative representative function. The digitization of the physical and virtual components of research and education processes, personnel and structural departments is a requirement of digital monitoring and cloud management of the university. To do this, the experts create competency metrics to measure quality of system components of the university: 1) relationship; 2) roadmap; 3) management; 4) infrastructure; 5) personnel; 6) resources; 7) graduates and research results; 8) science; 9) education. Smart cyber university integrates modern technologies: big data analytics and quantum data structures, cloud and quantum computing, mobile services and cyber physical systems, real and virtual multiprocessors [36, 41, 42] within the IoT-culture through the use of service-oriented computing platforms offered by world leading companies IBM, Google, Microsoft, NASA, Amazon, Facebook.

The goal of the SCU is improving the quality of educational services and scientific achievements of high school through the creation of the metric relationship, defining the rules of digital monitoring and cloud management of scientific and educational processes that allows eliminating corruption, attracting foreign investment, increasing productivity and the life quality of scientists and professors, producing market oriented graduates and scientific achievements.

The effectiveness of the university as a social system is defined by three parameters: consumption, exports and investment (Fig. 2). However, the relationship in social structure is primary (not people and economy); it is formed by the leader, legislations, history, culture. Relationship, consumption and exports form the market attraction of the SCU and image for foreign investments, closing the loop of SCU processes. Some time later a cause-effect chain (the leader – relationship – consumption – export – investment) becomes independent of the leader, the main role of which is the creation and activation of cyber social relationship at the university. Investments are the effect, not the cause of effectively functioning social system – money likes silence and stability of creative social relationship. The reason of corruption is not personnel that form the tree of the social hierarchy of officials, but the incorrect relationship, generated by the legislation, permitting illegal distribution of public funds and positions by officials. The problem solution is to give the distribution function of state resources to cyber management, which uses digitized metric relationships, approved by the society.

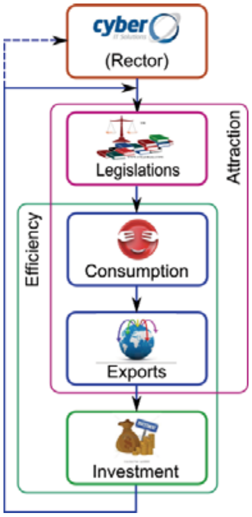


Figure 2. Social system

5 Innovative services of Smart Cyber University

Innovative cloud-mobile services, which form the SCU as a prototype of a global scientific and educational cyberspace, are represented below: 1) A cloud cyber service of secured E-document flow for digital monitoring and intelligent cyber management of research and education processes (creation, realization and utilization of E-document) in the format of closed loop "fact – measurement – evaluation – action", completely eliminating papers through the use of Cloud-Mobile Service Computing, cloud data storages, digital signatures, ID-cards, E-mail and mobile phone. 2) A cloud cyber service of mobile e-voting for monitoring public opinion, democratic decision-making at meetings, academic council, conferences, elections of experts, scientific and teaching staff for vacancies. 3) A cloud cyber service of personnel management based on online monitoring, measurement, rating and accumulation of digital competency metrics for evaluating the activity of students and staff in order to define transparent regulatory moral and material rewards, to select the candidates for the vacant positions of heads, researchers, and teachers. 4) A cloud cyber service for managing structural departments based on online monitoring, measurement and storage of digital competency metrics of the departments for defining regulatory management actions and generating the documents necessary for functioning the research and education processes of the university. 5) A cloud cyber service for quality assessment of educational processes and components, online testing of knowledge and skills, which excludes illegitimate relationship between teachers and students in the exams and tests. 6) A cloud cyber service for management of science based on the digital evaluation of research proposals, projects and results, competences of scientists, departments by using the expert metrics in order to transparent distribution of financial, human and time resources between departments and employees. 7) A cloud cyber service for education in the form of MOOC online and on-site courses, and also management of educational process based on transparent distribution of financial and time (credit) resources between departments and employees in strict accordance with the metric contribution assessment of each university department. 8) A cloud cyber service for monitoring and management of scientific and educational processes of the student in real time, and also for the creation and storage of electronic documents in order to support a student in time and space through the personal virtual cabinet, mobile device and e-mail. 9) A cloud cyber service for measurement and support undergraduate, master's and PhD theses, as well as submitted projects based on the international metrics and local quality criteria for evaluating scientific novelty and practical value of the results. 10) A cloud cyber service for licensing and accreditation of the university departments based on the measurement of scientific and educational achievements and automatic creation of corresponding documents required for the external evaluation of the education quality. 11) A cloud cyber service of electronic access and monitoring (24/7) the presence of staff and students in the infrastructure rooms of university through the use of mobile devices and the ID-cards. 12) A cloud service – cyber security for protection of informational and physical space of the university, and also the authorization of electronic access to all cyber physical components and processes associated with the SCU.

6 Rating formulae for assessment of employees and departments

The following components of the university: relationships, personnel, management, infrastructure, resources, road map, enrollees, students, alumni, research and educational processes, the life of employees should be digitally measured. Each entity has primitive measures, to which should be reduced any phenomenon or complex process of the university. The basic metric primitive for measuring the activity of the scientist is a printed article page in the journal, the size of which is 3000 symbols without spaces, which is estimated as 1 point. All other assessments of scientist's activity can be reduced to this primitive by scaling the complexity of solving problems by expert coefficients. The most important instrument for the creation of new corporate and moral-ethical relationship is the evaluation metric of scientific and educational results, having a cash equivalent. Achievements of scientists and departments are defined by the expert-driven coefficients of scientific, educational and socio-economic value in the matrix of competencies. The coefficients reflect the level of innovative solutions; they can be changed over time according to the new market trends. The weighted and normalized in the interval (0-1) criterion quality Q of integral department activity in the current year Y is proposed; it takes into account the average activity of the team in recent m years, where S is a number of staff, n is a quantity of metric parameters Pi. Each of them is reduced to the maximum value Pi (max) in the structure of the departments or employees. The essentiality of the parameter is determined through its multiplying by the coefficient of scientific-educational and socioeconomic value that is approved by experts and defined in one of the two closed intervals, $k_i = \{[0,1], [1,q]\}$:

$$Q_Y = \frac{1}{m+1}\left[\frac{1}{S\times n}\times\sum_{i=1}^{n}\frac{k_i\times P_i}{P_{i(max)}}+\sum_{j=1}^{m}Q_{Y-j}\right].$$

Quality criterion should have a corresponding level of moral and material rewards for each department and employee, which depends on scientific and educational achievements. The integrated metric for assessing the effectiveness of scientific and educational activities of the scientist-professor is determined by the previous formula with excluded symbol S. In fact, the metric values averaged performance of a scientist at the scale of the department, faculty or university. The numerator of the fraction determines the personal achievements, and the denominator - the best values of the achievements among all scientists of the university in each of the n nominations. According to the values of quality metric parameters of each scientist the cloud service assigns rewards within the university or department. It is important that this information was available to all employees in order to avoid the spread of rumors about the unfair distribution of rewards.

7 Leveraging cyber democracy at the university

Cyber democracy is a metric culture of social-technological relationship, formed by experts, that combines social groups and intelligent infrastructure in cyber physical space for comprehensive digital monitoring public opinion and cloud management of digitized social processes and phenomena in order to save the planet and achieve a high quality of life. More popular, cyber democracy is a structural combination of a comprehensive democratic discussion and digital exact monitoring social problems with the exact cyber management of society by intelligent cloud regulations.

The process model of cyber-democratic interaction of university staff and cloud management is shown in Fig. 3: 1) the social group of the university elects experts democratically by using their transparent metrics and delegates them powers to prepare the competency metric of digital measurement of the process or phenomenon; 2) the experts create the digital metric for assessment of the process or phenomenon in the form of a specification that is approved democratically by academic council or administration and after that it is inserted into the cloud service as the rules of the game when choosing decision; 3) the decision having the highest rating among all the proposals is generated by the cloud cyber service of management based on the digital metric, approved by the social group, for measurement of input data of the process or phenomenon.

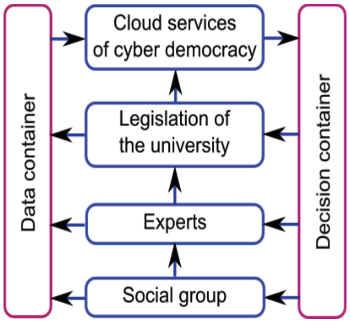

Cloud services of cyber democracy
Legislation of the university
Experts
Social group
Data container
Decision container

Figure 3. Cyber democracy

It is important to adopt new rules of the game: a metric of decision-making is under voting, not the decision itself, which is generated by the cloud service! A winner is defined by the cloud service as a manager, scientist, professor, which has the best metric value. Each member of the team obtains equal rights of moral and material rewards according to its activity, which are independent of subjective leader opinion. The new relationship between the cloud service and the team in a short time lead to the structure of the university, where the law 20/80 (active/passive employees) becomes the ratio of 80/20 that characterizes private companies. The new relationship makes the university attractive to the outside world that attracts a flow of investment and enrollees from highly developed countries. The image of the university is gradually transformed into a significant ranking of the university in the world market of scientific and educational services, which increases the level of employees' salaries at least in 2 times.

Principles of cyber democracy consistent with the structure of the proposed system: 1) Initiation of cyber democracy is fulfilled by a leader or social group. 2) The choice of experts is performed through metric evaluation of the competence matrices of candidates. It is possible to withdraw expert, laws, regulations, orders and decisions in the case of their negative impact on the functioning of the social system. 3) The digitization of all physical and virtual social processes, phenomena and relationships between them in cyber physical (cyber social) space is a necessary condition. Creation of the competence metrics in the cyber physical space for measurement of all social processes, events and relationships between them is necessary. 4) A cloud cyber service based on digitized cyber physical space of social-technological relationship provides comprehensive digital monitoring public opinion, and also metric evaluation of processes and phenomena in order to generate managing regulations. 5) Data storage (processes, phenomena, relationships) and regulatory actions (decisions, orders, regulations, statutes and laws) are performed in the intellectual containers of big data in cyberspace. 6) The transparency of monitoring and management of all processes and phenomena for all members of the social group is required. 7) Digital modeling, simulation and forecasting [43] of possible consequences in the processes and phenomena of the social system is necessary to define a response of cyber service to the generated decisions.

The innovations of cyber democracy are the following: decision making is performed not by the head, experts or society, but corruption-free cloud service, according to the rules, generated by experts delegated from the society. Cyber democracy operates in real time that allows monitoring all processes and phenomena for generating regulatory actions on time. The digitization of all members in social groups is possible if each individual has a mobile phone, e-mail, digital signatures (for adults). In this case, there is a legitimate interactive relationship between state institutions and citizens that is necessary for monitoring, measurement and management.

A cloud service for monitoring and management of student career is described below. Innovation is the creation of a mathematical model of career growth in the form of the equation describing the difference-driven interaction [36, 43] of three competence matrices: 1) the future social role P (Purpose); 2) the current achievements C (Current); 3) the current activity A (Activity). The equation (PCA) is universal and determines the distance between three components-matrices: goal – competence – activity by using xor-operation: $P \oplus C \oplus A = 0.$ PCA-equation allows to specify three problem for student activity: 1) P-problem: who do you want to be – potential reachability of desired social role $P = C \oplus A.$ 2) C-problem: how are you clever today – assessment of the current level of competence $C = P \oplus A.$ 3) A-problem: what do you do to achieve the desired future – roadmap in the educational space during the time interval $A = P \oplus C.$ Thus, the proposed mathematical model describes and assesses all the processes of formation of the individual as a socially significant personality during the time interval. In accordance with the PCA-equation, it is simple to create a cloud service for monitoring and management the process student education in the form of computing model that answers the most difficult question – how to achieve the desired future (universities, courses, professors, companies). The model includes cloud service, which gives the student recommendations on the gadget: roadmap – what, where and when to study in response to the conditions (syllabus) in the form of the matrices of future social role and current competence. The competence PCA-model is scaled to all social, cyber and technological processes and objects, where exists a need to address any of three above mentioned problems, if two of three components (goal, roadmap, status) are known.

8 Big Data driven multiprocessor

Vector-logical SIMD-multiprocessor is proposed. It is characterized by the absence of arithmetic operations, parallel computing the distances between the query and information components, as well as the simultaneous determination of the best possible solution by minimum value of the membership function, which makes it possible to significantly increase the speed of the most accurate data retrieval in big data. The multiprocessor structure is shown in Fig. 4, which includes only logical primitives for performing Boolean and vector (bit) operations.

The processor operates as follows: vector query m, consisting of k-bits interacts by xor-function with matrix M having n lines or vectors. As a result of xor-operation n membership functions are formed, which define the value of Hamming distance between the query and each vector-row of the matrix M.

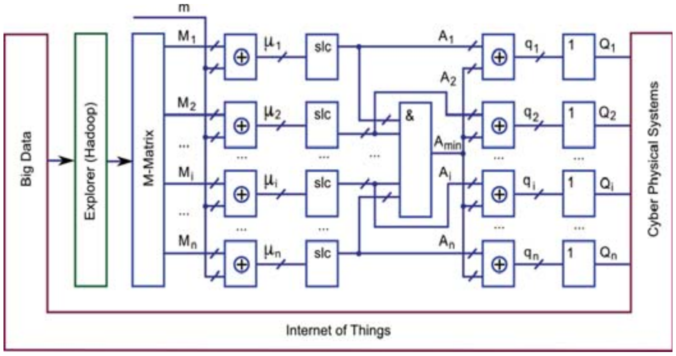


Figure 4. Interaction between multiprocessor and cyberspace

To estimate distances and to select the best (minimum) interaction a register slc-operation (shift-left-crowding) is performed, which crowds all 1-units to the left for one cycle, and makes it possible to mark the minimum distance $m \oplus M_i$ by the last right 1-unit bit number. To determine the number of vector-row forming the minimum of the membership function, parallel bitwise logical multiplication is performed over all the vectors containing crowded to the left 1-values, which allows calculating A_{min} vector with the minimum number of 1-units. This vector is used to determine the number or index of the vector-row of the matrix M, which has the best value of the membership function by performing vector xor-operation between A_{min} and all left shift crowded membership functions $A_i \ (i = \overline{1,n})$. As a result the vectors $q_i \ (i = \overline{1,n})$ are generated, bits of which specify the input values for each of the n logic or-elements. The output of each or-element is equal to one, if there is at least one 1-unit value in the results of the comparison $A_i \oplus A_{min}$. If no such 1-

units, the minimum distance between $m \oplus M_i$ is identified by 0-state of one or possibly several outputs $Q_i \ (i = \overline{1,n})$. An analytical model for finding the optimal solution in cyberspace by the query-vector is based on five parallel logical operations performed sequentially:

$$Q^i = \underset{j=\overline{1,k}}{\vee}\{[\underset{p=\overline{1,n}}{\wedge}\ \underset{s=\overline{1,n}}{slc}(m \underset{r=\overline{1,n}}{\oplus} M_i)] \underset{i=\overline{1,n}}{\oplus} M_i\}.$$

For efficient operation of the logic multiprocessor it is necessary to generate M-matrix of possible solutions of the problem, which, in particular, may be the result of usage of the retrieval engine Google (Hadoop) to the Cyberspace Internet (big data), used for rude and large extracts, when the number of found information components is the hundreds or thousands of variants. Then comes the turn of multiprocessor operation, forming an exact solution on the query m, which should be stored in a structured, specialized part of cyberspace for subsequent reuse. Therefore, the input and output of the logic multiprocessor should be forms of cyberspace: Internet of Things, Big Data, Cyber-Physical Systems. Market feasibility of the proposed multi-processor is the ability to use it for improving the quality and speed of retrieval procedures in big data, creating built-in systems for diagnosis and repair, tools for targeting and pattern recognition. Typical functionality for cyber-physical systems using information space is the generation of multi-alternative variants of responses to a query in vector-logical form of cyberspace components (subjects, processes or phenomena), which are necessary for human-free management of social, biological and non-natural production process.

9 Conclusions

1) A cyber physical system Smart Cyber University is characterized by using the digitized space of regulatory rules, exact monitoring and active cloud management of addressable components of scientific and educational processes, automatic generation of regulatory signals, independent of the heads cyber decision-making to manage financial and human resources, and also eliminating paper documents from scientific and educational processes.

2) The main components of the Smart Cyber University are defined: infrastructure, personnel, relationships, management, road map and resources, which represented in a cyberspace to perform scientific and educational processes based on digital monitoring and cloud management.

3) Innovative services, focused on realization of the Smart Cyber University as a prototype of a global virtual scientific and educational cyberspace as a picture of the future of higher education, are offered.

4) To implement cloud services of the Smart Cyber University a model for metric assessment of students, scientists, teachers, departments, science and education of a university is proposed in order to perform digital monitoring and management of resources and processes and achieve high quality of market-oriented researches and specialists.

5) Cyber management services are scaled to the high school for significant reduction in public expenditure on administrative staff, improve the efficiency of scientific and educational processes through the complete elimination of corruption, cyber rewarding scientists and professors, creating a market-demanded researches and specialists.

6) New services for monitoring and managing digitized scientific and educational processes, departments and service units of the university are offered; they are based on IoT technological culture, having the hierarchy structure Cloud - Fog Networks - Mobile, which exclude paper documents and dependence from the subjectivity of academic officials.

7) A model of tolerant interaction between democratic voting rules and cyber management in decision-making processes for public universities and organizations is developed.

8) The market attractiveness of cloud services for state scientific and educational institutions regulating constructive activity of the teaching staff is defined that could increase at least twice the productivity of scientists.

9) A model of human-free cyber management of social significant processes and resources (personnel and finances) is proposed; a model is characterized by using cloud services for allocating government procurement and finance between the structural units based on the competence metric, and also the distribution of staff positions.

10 References

[1] P. Barnaghi, A. Sheth, V. Singh, M. Hauswirth, “Physical-Cyber-Social Computing: Looking Back, Looking Forward,” *IEEE Internet Computing*, pp. 7-11, May-June 2015.

[2] Proceedings of IEEE SERVICES / BigData Congress CLOUD/ICWS/SCC/MS, New York City, 2015.

[3] Smart City: How to Create Public and Economic Value with High Technology in Urban Space, Editors: R. P. Dameri, C. Rosenthal-Sabroux, Springer, 2014, ISSN: 2196-8705.

[4] T. Clohessy, T. Acton, L. Morgan, “Smart City as a Service (ScaaS): A Future Roadmap for E-Government Smart City Cloud Computing Initiatives,” in *IEEE/ACM 7th International Conference on Utility and Cloud Computing (UCC)*, pp. 836-841, 8-11 Dec. 2014.

[5] A. Copie, T. Fortis, V.I. Munteanu, V. Negru, “From Cloud Governance to IoT Governance,” *in 27th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 1229-1234, 25-28 March 2013.

[6] Hai-Ning Liang, Ka Lok Man. “Building a smart laboratory environment at a university via a cyber-physical system,” in *IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE),* pp. 239-247, 2013.

[7] M.V. Bueno-Delgado, P. Pavon-Marino, A. De-Gea-Garcia, A. Dolon-Garcia, “The Smart University Experience: An NFC-Based Ubiquitous Environment,” in *Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, pp. 799-804, 4-6 July 2012.

[8] Tao Li; Wei Mao, “Intelligent document technology in university educational administration management system,” in *IEEE International Symposium on IT in Medicine and Education, ITME* 2008, pp. 103-107, 12-14 Dec. 2008.

[9] M. Owoc, K. Marciniak, “Knowledge management as foundation of smart university,” in *Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp.1267-1272, 8-11 Sept. 2013.

[10] T. Pardo, “Guest Editors’ Introduction: Research in the Digital Government Realm,” in *Computer,* vol. 38, no. 12, pp. 26-32, Dec. 2005.

[11] Jongbae Moon, Chongam Kim, Kum Won Cho, “CFD Cyber Education Service Using Cyber infrastructure for e-Science,“ in *Fourth International Conference on Networked Computing and Advanced Information Management*, pp. 306-313, 2008.

[12] S.M.M.Gilani, J. Ahmed, M.A. Abbas, “Electronic document management: A paperless university model,” in *2nd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2009*, pp. 440-444, 8-11 Aug. 2009.

[13] K.G. Alberto, C.M. Abella, M.G.C.E. Sicat, J.D. Niguidula, J.M. Caballero, “Compiling Remote Files: Redefining Electronic Document Management System Infrastructure (CreED),” in *International Conference on Information and Multimedia Technology,* pp.347-350, 16-18 Dec. 2009.

[14] J. Paradis, M. Zimmerman, The MIT Guide to Science and Engineering Communication. Electronic Documents. MIT Press, 2002.

[15] Xiao Xi Liu, Jian Qiu, Jian Ming Zhang, “High Availability Benchmarking for Cloud Management Infrastructure,” in *International Conference on Service Sciences (ICSS)*, pp. 163-168, 22-23 May 2014.

[16] Li Xu, Guozhen Tan, Xia Zhang, Jingang Zhou, “Aclome: Agile Cloud Environment Management Platform,” in *Fourth International Conference on Digital Manufacturing and Automation (ICDMA)*, pp. 101-105, 29-30 June 2013.

[17] T. Forell, D. Milojicic, V. Talwar, “Cloud Management: Challenges and Opportunities,” in *IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW)*, pp. 881-889, 16-20 May 2011.

[18] P. Bellini, D. Cenni, P. Nesi, “A Knowledge Base Driven Solution for Smart Cloud Management,” in *IEEE 8th International Conference on Cloud Computing (CLOUD)*, pp. 1069-1072, June 27-July 2, 2015.

[19] N. da Fonseca, R. Boutaba, Cloud Services, Networking, and Management. Wiley-IEEE Press, 2015,

[20] V. Hahanov, W. Gharibi, A. Zhalilo, E. Litvinova, “Cloud-driven traffic control: Formal modeling and technical realization,” in *4th Mediterranean Conference on Embedded Computing (MECO)*, pp. 21-24, 14-18 June 2015.

[21] T. Kalvet, “Management of Technology: The Case of e-Voting in Estonia,” in *International Conference on Computer Technology and Development*, vol. 2, pp. 512-515, 13-15 Nov. 2009.

[22] A. Villafiorita, K. Weldemariam, R. Tiella, “Development, Formal Verification, and Evaluation of an E-Voting System With VVPAT,” in *IEEE Transactions on Information Forensics and Security*, vol.4, no. 4, pp. 651-661, Dec. 2009.

[23] M. Bishop, D.A. Frincke, “Achieving Learning Objectives through E-Voting Case Studies,” in *IEEE Security & Privacy*, vol. 5, no. 1, pp. 53-56, Jan.-Feb. 2007.

[24] R. Buckland, R. Wen, “The Future of E-voting in Australia,” in *IEEE Security & Privacy*, vol. 10, no. 5, pp. 25-32, Sept.-Oct. 2012.

[25] B. Priyogi, B.A. Nan Cenka, A.A.G.Y. Paramartha, A. Rubhasy, “Work in progress – Open Education Metric (OEM) developing metric to measure open education service quality,” in *1st International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE),* pp. 319-323, Nov. 2014.

[26] D. Tsay, E. P. Matthews, “Metrics for comparative analysis of operations competency,” in *Bell Labs Technical Journal*, vol. 10, no. 1, pp. 175-179, Spring 2005.

[27] A.M. Fairchild, “Knowledge management metrics via a balanced scorecard methodology,” in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, pp. 3173-3180, 7-10 Jan. 2002.

[28] Jin Zhu, Hu Xujie, “Evaluation of the Teacher Quality in University Based on the Unascertained Measurement Model,” in *Second International Symposium on Electronic Commerce and Security*, vol. 2, pp. 222-225, 22-24 May 2009.

[29] H.M. Jani, “Intellectual capacity building in higher education: Quality assurance and management,” in *5th International Conference on New Trends in Information Science and Service Science (NISS)*, vol. 2, pp. 361-366, 24-26 Oct. 2011.

[30] Linquan Zhang, Chuan Wu, Zongpeng Li, Chuanxiong Guo, Minghua Chen, F.C.M. Lau, “Moving Big Data to The Cloud: An Online Cost-Minimizing Approach,” in *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 12, pp. 2710-2721, December 2013.

[31] P. Tantatsanawong, A. Kawtrakul, W. Lertwipatrakul, “Enabling Future Education with Smart Services,” in *Annual SRII Global Conference (SRII)*, pp. 550-556, March 29 2011-April 2 2011.

[32] Liu Bing, Su Yuan, Li Yuan, “Research on Customer Relationship Management Intelligent Decision-Making Mechanism of University Based on System Dynamics,” in *International Workshop on Intelligent Systems and Applications*, pp. 1-5, 23-24 May 2009.

[33] H. Nugroho, K. Surendro, “Proposed model of Vocational University Governance and measurement model by utilizing the ISO 38500 framework and COBIT 5 enabler,” in *International Conference on ICT for Smart Society (ICISS)*, pp. 1-5, 13-14 June 2013.

[34] D. Braun, “Governance of universities and scientific innovation,” in *Atlanta Conference on Science and Innovation Policy*, pp. 1-37, 15-17 Sept. 2011.

[35] ISO/IEC 38500:2008. Corporate governance of information technology.

[36] V.I. Hahanov, M.F. Bondarenko, E.I. Litvinova, “Structure of logic associative multiprocessor,” in *Automation and Remote Control,* no 10, pp. 73-94, 2012.

[37] V. Hahanov, W. Gharibi, A.P. Kudin, I. Hahanov, C. Ngene, Y. Tiekura, D. Krulevska, A. Yerchenko, A. Mishchenko, D. Shcherbin, A. Priymak, “Cyber Physical Social Systems – Future of Ukraine,” in *Proceedings of IEEE East-West Design & Test Symposium (EWDTS’2014)*, Kiev, Ukraine, pp. 67 – 81, 2014.

[38] V. Hahanov, E. Litvinova, W. Gharibi, S. Chumachenko, “Big Data Driven Cyber Analytic System,” in *IEEE International Congress on Big Data, New York City*, pp. 615-622, 2015.

[39] V. Hahanov, S. Chumachenko, A. Mishchenko, Mazen Abdelrahman Abdelaziz Hussein, A. Hahanova, I. Filippenko, “CyUni Service – Smart Cyber University,” in *Proceedings of IEEE East-West Design & Test Symposium (EWDTS-2015), Batumi, Georgia*, pp. 129-136, 2015.

[40] V. Abdullayev, V. Hahanov, E. Litvinova, F. Dahiri, A. Arefiev, Y. Hahanova, “Cloud Service – Cyber Social Democracy and Smart University,” in *Proceedings of IEEE East-West Design & Test Symposium (EWDTS-2015), Batumi, Georgia*, pp. 176-180, 2015.

[41] V.I. Hahanov, W. Gharibi, E.I. Litvinova, A.S. Shkil, “Qubit data structure of computing devices,” in *Electronic modeling*, vol. 37, no 1, pp. 76-99, 2015.

[42] V.I. Hahanov, Tamer Bani Amer, S.V. Chumachenko, E.I. Litvinova, “Qubit technology analysis and diagnosis of digital devices,” in *Electronic modeling*, vol. 37, no 3, pp. 17-40, 2015.

[43] V. Hahanov, A. Barkalov, M. Adamsky, Design of Digital Systems and Devices. Infrastructure intellectual property for SoC simulation and diagnosis service, Springer, pp. 289-330, 2011.

Disease Surveillance Big Data Platform for Large Scale Event Processing

Silvino Neto[1] **and Felipe Silva Ferraz**[1]
[1]Recife Center for Advanced Studies and Systems, Recife, Brazil
silvino.neto@gmail.com, fsf@cesar.org.br

Abstract - *In a globalized world, disease outbreaks are likely to spread rapidly across the countries and territories. Therefore, early reports are extremely important in order to predict, identify, confirm, and respond to these occurrences, reducing the risks and consequences of large epidemics. This paper discusses the issues related to this topic and presents a Predictive Analytical Decision Support System (PADSS) integrated into a cloud-based Message-oriented Middleware (MOM) platform, developed to connect healthcare organizations in order to share electronic health records and statistical reports. This platform uses a customized version of the Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR) specification to support system-level data exchange, enabling the prediction of disease outbreaks using a combination of techniques to perform real-time data analysis.*

Keywords: Disease Surveillance; Big Data; Message-oriented Middleware; FHIR; Predictive Analysis.

1 Introduction

In our previous work [1], we have presented the Platform for Real-Time Verification of Epidemic Notification (PREVENT), a cloud-based message-oriented middleware (MOM) platform for real-time disease surveillance. PREVENT was developed in order to process information streams originated from numerous sources allowing for the early identification of threats and prompt response. Modern communication infrastructure improves our capacity to report disease outbreaks worldwide in a timely manner. Institutions such as the World Health Organization (WHO) and the Centers for Disease Control (CDC) have been engaged in the development of an intelligence network in order to gather and process information from a wide range of sources, promoting a systematic event detection mechanism to support alert and response operations.

In order to support system-level exchange of clinical data in a large scale, PREVENT uses Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR) specification [2] [3]. FHIR is HL7's new specification that comprises a set of international standards to exchange clinical and administrative data between healthcare applications. It vastly improves previous standards and technologies used in terms of simplicity, scalability, and extensibility. When compared with its predecessors, HL7 FHIR offers a whole new set of features, such as: support for multiple data formats like Extensible Markup Language (XML) and JavaScript Object Notation (JSON), an extensible data model, and a RESTful API.

In this paper, we introduce our new Predictive Analytical Decision Support System (PADSS) built on top of the PREVENT middleware platform. Aiming to quantify reported case numbers, PREVENT was designed to use a customized instance of the FHIR specification to carry statistical reports related to disease occurrences in order to monitor and notify disease outbreaks in real-time fashion. PADSS loads and analyzes data extracted from messages received from healthcare organizations, in order to identify patterns found in historical and transactional data enabling the anticipation of an outbreak occurrence.

This paper is further structured as follows: In section 2, we examine the background for this paper. In section 3, we present an overview of this platform system architecture and introduce PADSS design. In Section 4, we discuss our evaluation approach and present the results obtained. At last, section 5 closes the paper by presenting our conclusions.

2 Background

In this section, we introduce the concepts that served as basis for the development of this research.

2.1 Disease Surveillance

Disease surveillance is the ongoing systematic collection, analysis, and interpretation of outcome-specific data for use in planning, implementing and evaluating public health policies and practices [4].

In order to monitor potential threats, several initiatives have been coordinated by the WHO in collaboration with a wide network of institutions, such as the CDC, national public health institutes, and international healthcare agencies. These initiatives resulted in the development of a global surveillance network established to control and prevent disease outbreaks. This network shares information provided by a broad range of formal and informal sources, therefore every piece of information gathered needs to be verified, in order to assess its worth. In a joint effort, Health Canada and the WHO developed the Global Public Health Intelligence Network (GPHIN) which is a secure Internet-based early-warning tool that continuously searches global media sources to identify information about disease outbreaks and other incidents of potential international public health concern [5].

As a significant portion of initial outbreak reports comes from unofficial non-electronic sources, a large effort is required for the validation and verification of the information received. As a consequence, many researchers have been working on the development of disease surveillance platforms that use a structured approach, based on electronic information provided from reliable sources. The CDC in the United States developed the National Electronic Disease Surveillance System (NEDSS), which is a standards-based approach to facilitate electronic transfer of public health surveillance data from healthcare systems to public health departments [6]. NEDSS uses a system-level message exchange strategy, based on HL7 messaging standards to ensure interoperability between healthcare institutions.

2.2 Healthcare Interoperability

Over the last three decades, there have been several attempts to improve interoperability between healthcare systems. Application vendors and scholars have collaborated on the development of a set of international standards that establish a framework for clinical and administrative data exchange between healthcare systems. As a result of this collaboration, in 1987 the HL7 was created. HL7 is a nonprofit international organization that supports and promotes the development of international interoperability standards and specifications for healthcare software applications [7].

In 1989, HL7 introduced the HL7 v2 messaging standard. HL7 v2 is an ad hoc messaging approach used to transfer several sorts of health-related information. HL7 v2 has become a widely used standard, being adopted and supported by most healthcare software application vendors in North America [2]. Despite HL7 v2 significant acceptance, the limitations imposed by its non-XML encoding syntax based on segments and delimiters have not allowed significant high scale use in larger multiplatform environments. The lack of a formal data model is considered a major drawback in HL7 v2 messaging approach. Several limitations were observed such as no common data dictionary or message transmission interfaces available.

HL7 v3 messaging standard emerged as a response to all the limitations and issues observed on the previous version. Despite its XML-based syntax and object-oriented approach, HL7 v3 was heavily criticized by healthcare software vendors for being inconsistent, overly complex and infeasible to implement or migrate to in production environments [2]. Given that HL7 had spent a considerable amount of time working on the development of the now unpopular HL7 v3 messaging standard, it seemed as if interoperability efforts for healthcare were stalled.

Hence, FHIR was created with the objective of being a simple, extensible, and scalable healthcare messaging standard. There have been several discussions towards a new messaging approach for data exchange in healthcare systems. As result, FHIR provides a Representational State Transfer (REST) interface, which is a simple, efficient and lightweight interoperable strategy for system integration.

FHIR provides a simple and modular object-oriented data model for exchanging electronic health records, but it also supports the data models introduced by HL7 previous specifications, in order to facilitate interoperation with legacy platforms. FHIR data model is extensible, allowing applications to define and use a set of customized resources and data structures. However, in order to declare and use new custom resource types, a set of requirements must be met, in order to guarantee the security and consistency of the model. FHIR supports both XML and JSON based syntax, it simplifies system-level communication by using a common set of interfaces. It also presents a resource interoperable design that allows information to be promptly distributed, providing an alternative to document-centric approaches by directly exposing data elements as services.

2.3 Big Data

Big Data refers to data sets so large, complex and dynamic that conventional data processing tools are insufficient to capture, store, manage and analyze. These data sets hold large volumes of many kinds of information that may be useful for several purposes, ranging from modeling customer behavior to disease outbreak tracking. Predictive analytics is a collection of methods used to extract value from stored data, in form of predictions about future events. The data analysis process can lead to improved decision making, which reflects in greater operational efficiency and risk reduction. Predictive analytics allows researchers to map correlations and identify trends that can help prevent disease outbreaks.

In the healthcare domain, Big Data is a relatively recent research topic. The work of Andreu-Perez et al. has presented an accurate assessment of recent developments in big data in the context of health informatics [15]. As observed in [15], there are several big data related topics concerning medical and healthcare computer-based solutions. Considering the scope of this paper, we focus mostly on the issues related to social, environmental and public health research fields.

Moreover, according to the work of Hay et al., one major limitation observed on most disease surveillance systems is related to their ability to combine static spatially continuous maps of infectious disease risk and continually updated reports of infectious disease occurrences [16]. Companies such as Epidemico [17] have been dedicated to working on commercial products that provide continuous monitoring of disease outbreaks, aiming to address this limitation.

Promising projects have emerged from researches related to the use of Big Data processing platforms. In China, Anying et al. have introduced a disease trend analysis model developed to predict regional outbreaks applying big data processing and data mining techniques to process data collected from drug sales reports, electronic medical records and geospatial data [18]. Recently, social media has surfaced as an important source of information to support real-time monitoring of disease occurrences, using geographical, temporal and text analysis. For instance, Lee et al. developed a real-time disease surveillance system that uses data captured from twitter messages to automatically track flu and cancer related activities [19].

2.4 Complex Event Processing (CEP)

CEP is the leading paradigm for the development of reactive monitoring systems, using an event-driven approach. CEP gathers and analyzes data from multiple distinct sources, allowing users of a system to receive chunks of information on the occurrence of certain circumstances. CEP processes streams of data in order to identify their significance within a cloud of information [8]. In order to accomplish that, CEP applies a set of rules to aggregate, filter and match low-level events, coupled with actions to generate new, higher-level ones derived from those events [9]. If combined with the appropriate technology, CEP enables the development of event-based information systems that are capable of performing real-time data analysis.

There are many forms of implementation of CEP based technologies both in the academy and industry. Most solutions are classified amongst two main categories: Aggregation-oriented CEP or Detection-oriented CEP. The first approach uses real-time processing of data captured from inbound events. Thus, on-line algorithms are executed in response to each event data unit entering the system. The second approach focuses on the examination of event data searching for patterns or recurring behaviors. Several aplicattions rely on a combination of both approaches [12].

Once data extraction has completed, the information obtained usually goes through a second-step analysis that aims to identify whether the events triggered at that time indicate a threat or opportunity. For instance, a significant amount of reports of an infectious disease coming from a geographically limited area could imply on an alert situation.

Therefore, if any threats have been identified, CEP solutions will immediately act on them, in order to initiate contingency measures. An event-driven processing strategy is an optimal choice for applications concerned with regular delivery of situational awareness and response [14]. For sensitive information that requires a high level of confidentiality and integrity, it is extremely important that CEP solutions provide security mechanisms that guarantee information safety. Migliavacca's work introduces a robust middleware platform that enforces security policies for distributed event-driven applications [13].

3 Middleware

PREVENT is a MOM platform, designed to gather and process data received from healthcare organizations, ranging from hospitals to public health institutions. Data providers must go through an electronic subscription process in order to send and receive notifications. Data is exchanged in the form of electronic medical records that conform to the HL7 FHIR messaging standard. The data received goes through a real-time analysis process, in order to identify possible occurrences of disease outbreaks, using pre-calculated risk profiles. A CEP unit is used to identify recurring patterns and to infer trends over the analyzed information. Given that an alert situation has been detected, PREVENT will asynchronously notify subscribed healthcare organizations systems using an HTTPS push request that holds an extended HL7 FHIR message, improved to support statistical reports. The deployment diagram exhibited in Figure 1 illustrates the middleware system architecture.

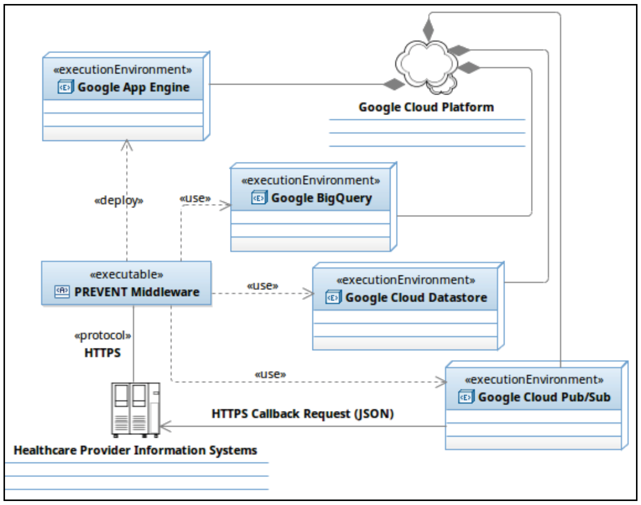


Fig. 1. Middleware System Architecture

In our initial work [1], we have established a small set of features and improvements that would further this research. In the remainder of this section, we introduce our most recent developments.

3.1 Decision Support System

The PREVENT platform was designed to acquire and process streaming data in real-time fashion, using stream analysis within a CEP engine. Given that PREVENT handles large data sets containing collections of data types, we concluded that the results obtained would be vastly improved by using Big Data analytics in a joint approach with CEP. As a consequence, PREVENT platform has been upgraded with a new decision support system, capable of performing large scale data analysis using a design strategy that resembles an Online Analytical Processing (OLAP) system. Prior to the big data hype, similar approaches have been attempted using data warehousing techniques, as the one introduced by Santos [11].

Considering that PADSS must be capable of processing large analytical datasets really quickly, PREVENT platform architecture has evolved to include the Google BigQuery framework, which is a cloud-based analytical database that is able to query massive datasets in few seconds [10]. Hence, it is capable of processing huge amounts of data in close to real-time. Google BigQuery implements a read-only dialect based on the Structured Query Language (SQL) standard, which is a simple and well-known domain-specific language used to manipulate data held in relational databases.

As discussed in our previous work [1], PREVENT stores the notifications received from healthcare organizations in a NoSQL database. Each notification holds a HL7 FHIR message in JSON format, containing relevant information

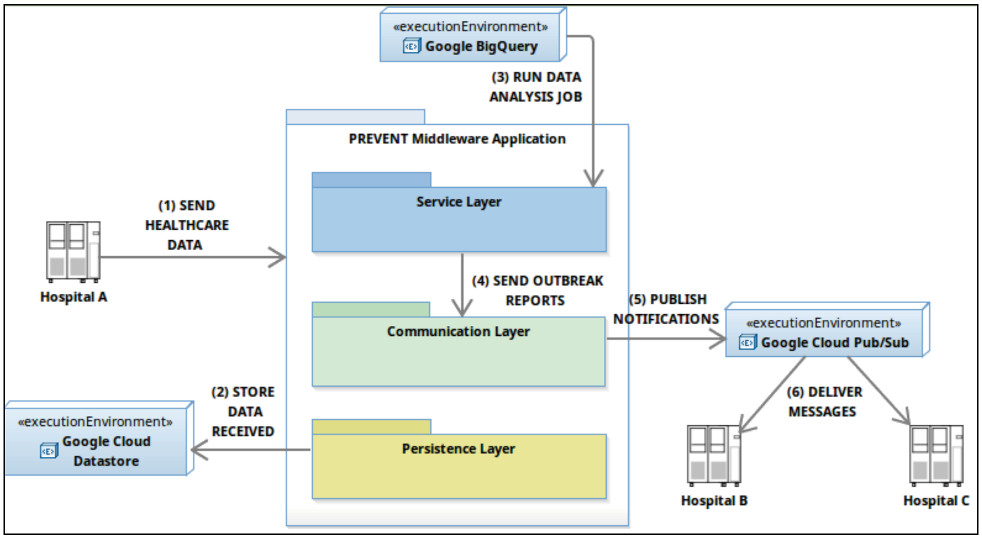


Fig. 2. Middleware Data Flow

about patient diagnostics, location, incidence rates, etc. The information provided is useful to identify recurring patterns that could be matched to historical data, favoring the prediction or early identification of outbreak occurrences.

In the following subsections, we explain how PADSS loads data from PREVENT NoSQL database into BigQuery and how it queries analytical data to be further processed by its CEP unit.

3.2 Data Analysis Process

The PADSS data analysis is performed as an asynchronous task, similarly to a batch process. Data selection operations are carried out in Query Jobs that are periodically executed. These jobs are designed to run a set of predefined queries that use aggregate and window functions in order to extract valuable information related to a specific time window.

The information collected is analyzed using statistical methods such as standard deviation and Z-scores. A threshold value is used to set the boundaries for an outbreak classification. For this particular case study, an arbitrary epidemic threshold of 1.96 times the standard deviation of the mean, for a two-week time window, is used for the characterization and detection of an outbreak [23]. A logistic regression model is used to estimate the probabilty of an outbreak occurrence. For However, given the adaptive nature of our rule-based processing engine, algorithms and rules used for outbreak classification may be easily modified to better adjust to seasonal and geographical variations. The collected data is used as input to PREVENT CEP engine, in order to decide whether an alert message should be dispatched to the subscribed healthcare applications. Traditional approaches used for mining historical data, in order to perform pattern recognition and trend prediction, relied mostly on statistical analysis. However, CEP-based strategies have proven to be much faster and scalable for high-frequency data, producing results of equivalent accuracy in time-constrained scenarios [14].

The strategy we used to integrate our CEP engine with an analytical database is partially inspired by the patterns presented on Maier's reference architecture proposal for big data technologies [20]. In his work, Maier performs an assessment of current state of art tools and technologies for big data, and introduces some patterns used by both industry and academy to implement predictive data analysis. One of the patterns discussed, presents the strategies implemented by commercial products to streamline analytical data into a CEP engine. As result, PREVENT platform is now capable of processing streaming data (real-time) in combination with historical datasets (batch), leveraging its confidence levels and accuracy. In order to systematically monitor the data received, PREVENT uses a set of SQL queries that retrieves historical information to be compared against real-time data. The objective of this process is to identify abnormal patterns or data spikes amongst the statistics collected from the messages received. The middleware data processing chain is exhibited in Figure 2.

BigQuery enhanced SQL dialect accelerates statistical analysis by providing a large collection of built-in aggregate and window functions. Additionally, it provides several utility functions that vastly simplify operations such as: Retrieving scalar values persisted on JSON data structures, using JSONPath expressions, regular expressions, date time operations, mathematical and trigonometric operations.

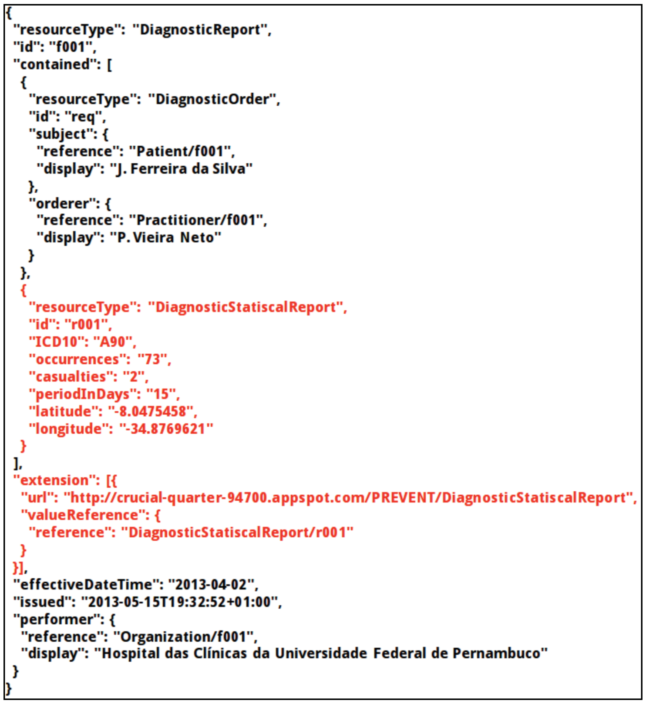
4 Evaluation

Our evaluation approach aims to portray real-life usage scenarios, employing a set of simulations that will be further described. This strategy attempts to illustrate effective use of this platform in order to anticipate and respond to disease outbreaks. The criteria established for this validation is based on the timely delivery of outbreak reports according to the results obtained by analytical data processing.

In our experiments, we evaluate the middleware by using a set of simulation tools. The test environment prepared for this evaluation is composed by a cloud-based instance of PREVENT/PADSS middleware application deployed into the Google Cloud Platform. A set of 50 HTTPS endpoints have been deployed in order to simulate individual health care units that were previously registered on the PREVENT middleware platform, to send and receive notifications. Furthermore, to emulate enlisted healthcare systems, each HTTPS endpoint is either a Java Servlet class or a PHP file that logs the messages received and returns an HTTP status code 200 (OK) to acknowledge successful reception of notifications delivered. This evaluation process is comprised by a single test scenario that affects all the components that are connected to perform real-time predictive analysis, using a combined approach that includes big data analytics and large scale event processing.

4.1 Simulation Description

The simulation experiment performed in the scope of this evaluation is based on the ongoing dengue fever outbreak reported in the Recife metropolitan area. Dengue fever is a mosquito-borne tropical disease, which is commonly reported in developing countries. Given the unavailability or restricted access to actual electronic medical records, in order to perform this experiment, we had to prepare a simulated dataset, comprised of hypothetical and fictional information, but representative of the occurrences we expect to observe. Thus, we created a dataset composed by a total of 100,000 HL7 FHIR messages. Each message holds information related to: healthcare unit location (latitude and longitude) and patient diagnostics. The message samples are comprised of several distinct ICD-10 codes, representing numerous diseases or health conditions. A FHIR message sample is illustrated in Figure 3.



```
{
 "resourceType": "DiagnosticReport",
 "id": "f001",
 "contained": [
  {
   "resourceType": "DiagnosticOrder",
   "id": "req",
   "subject": {
    "reference": "Patient/f001",
    "display": "J. Ferreira da Silva"
   },
   "orderer": {
    "reference": "Practitioner/f001",
    "display": "P. Vieira Neto"
   }
  },
  {
   "resourceType": "DiagnosticStatiscalReport",
   "id": "r001",
   "ICD10": "A90",
   "occurrences": "73",
   "casualties": "2",
   "periodInDays": "15",
   "latitude": "-8.0475458",
   "longitude": "-34.8769621"
  }
 ],
 "extension": [{
  "url": "http://crucial-quarter-94700.appspot.com/PREVENT/DiagnosticStatiscalReport",
  "valueReference": {
   "reference": "DiagnosticStatiscalReport/r001"
  }
 }],
 "effectiveDateTime": "2013-04-02",
 "issued": "2013-05-15T19:32:52+01:00",
 "performer": {
  "reference": "Organization/f001",
  "display": "Hospital das Clínicas da Universidade Federal de Pernambuco"
 }
}
```

Fig. 3. FHIR Message Sample

A total of 10,000 messages have been customized to carry statistical data reports as an extension of the HL7 FHIR messaging specification. The FHIR message samples used in this experiment are represented in JSON format, including the extensions of the FHIR model introduced by this research, which are highlighted in red as depicted in Figure 3. The customization of the FHIR standard was necessary given the current limitations observed in the FHIR messaging data model. Up to now, there is no available support for aggregated data reporting, including public health information. A complete list of pending improvements and additions that are expected to be included in the FHIR specification can be found at [21].

In addition to the information previously described, each message contained in the dataset is associated with one specific sender. The sender is one of the 50 subscribed HTTPS endpoints that have been set up in order to simulate health care unit applications. Each health care unit represents a facility that could be a hospital, health care agency, or any other health-related government organization. Finally, each sender is linked with a single location, using traditional geographical coordinates. Chosen locations are restricted to the cities located in the State of Pernambuco, Brazil. If the data analysis outcome indicates that an outbreak occurrence has been identified, the PREVENT platform will start dispatching alert notifications to the registered HTTPS endpoints that are located within a limited risk zone. The alert messages dispatched are customized HL7 FHIR messages improved to report alert and emergency events, given that there is currently no available support for outbreak reporting in the FHIR specification. To confirm that the HTTPS endpoints, which are eligible to receive notifications, have been properly notified, we developed a shell script that is able to extract information from log files using a set of regular expressions, allowing us to identify that an outbreak report has been received.

4.2 Results

Our previous experiments [1] have successfully demonstrated the effectiveness of the PREVENT platform concerning QoS requirements. In this paper, our evaluation approach performs a case study based on a Dengue fever outbreak, using a simulated dataset. The results obtained by this evaluation show that a mixed approach, that uses both CEP and Big Data technologies, enables the timely delivery of outbreak reports through the implementation of pattern recognition and statistical methods, allowing a more agile decision making.

The use of large analytical datasets improves our ability to identify emerging patterns, by matching historical data with current information. As depicted on Figure 5, in the following SQL query, we match historical aggregated and geospatial data, in order to identify the locations with higher incidence rates of Dengue fever.

The information obtained by the query exhibited in Figure 5 is used to map the historical data retrieved with newly reported cases. This query obtains the total and average case numbers for each quarter of the year, followed by the standard deviation. The results are restricted to Dengue fever cases

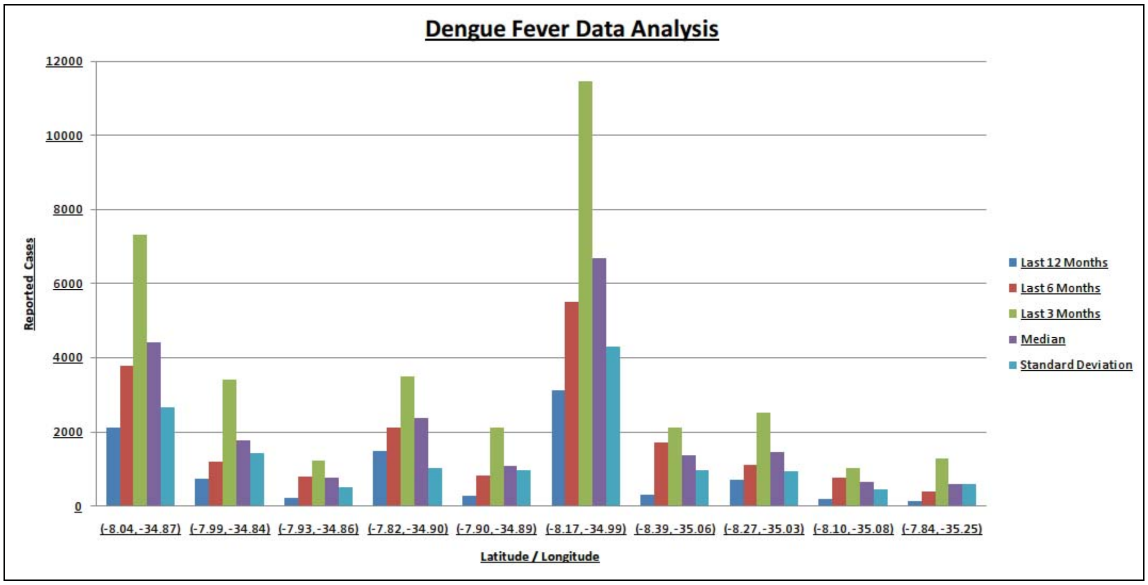
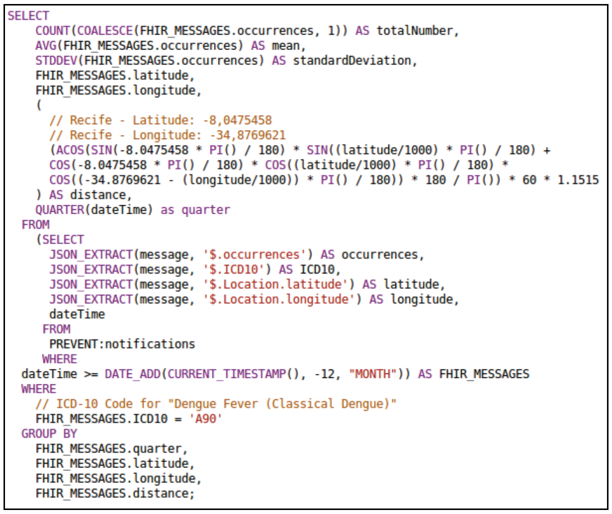


Fig. 4. Case Study Data Analysis

only, occurred in the past 12 months, and grouped by latitude and longitude. Considering the results obtained, we are able to build a column chart that compares each location statistics, and analyzes deviation patterns occurred in the processed datasets, as demonstrated in Figure 4, located on the next page.



```
SELECT
    COUNT(COALESCE(FHIR_MESSAGES.occurrences, 1)) AS totalNumber,
    AVG(FHIR_MESSAGES.occurrences) AS mean,
    STDDEV(FHIR_MESSAGES.occurrences) AS standardDeviation,
    FHIR_MESSAGES.latitude,
    FHIR_MESSAGES.longitude,
    (
      // Recife - Latitude: -8,0475458
      // Recife - Longitude: -34,8769621
      (ACOS(SIN(-8.0475458 * PI() / 180) * SIN((latitude/1000) * PI() / 180) +
      COS(-8.0475458 * PI() / 180) * COS((latitude/1000) * PI() / 180) *
      COS((-34.8769621 - (longitude/1000)) * PI() / 180)) * 180 / PI()) * 60 * 1.1515
    ) AS distance,
    QUARTER(dateTime) as quarter
  FROM
    (SELECT
      JSON_EXTRACT(message, '$.occurrences') AS occurrences,
      JSON_EXTRACT(message, '$.ICD10') AS ICD10,
      JSON_EXTRACT(message, '$.Location.latitude') AS latitude,
      JSON_EXTRACT(message, '$.Location.longitude') AS longitude,
      dateTime
     FROM
      PREVENT:notifications
     WHERE
  dateTime >= DATE_ADD(CURRENT_TIMESTAMP(), -12, "MONTH")) AS FHIR_MESSAGES
  WHERE
    // ICD-10 Code for "Dengue Fever (Classical Dengue)"
    FHIR_MESSAGES.ICD10 = 'A90'
  GROUP BY
    FHIR_MESSAGES.quarter,
    FHIR_MESSAGES.latitude,
    FHIR_MESSAGES.longitude,
    FHIR_MESSAGES.distance;
```

Fig. 5. PADSS Query Used For Evaluation Experiment

Also, using the geographical coordinates retrieved, the incidence rates calculated, and the spherical law of cosines [22], we have drawn a circle on the map representing the locations that are at increased risk of exposure. This is significantly valuable information, considering that we may be able to assess the spreading patterns of an outbreak, improving our response and control mechanisms. Figure 6 exhibits the outbreak incidence map for our Dengue fever study case, based on the data processed.

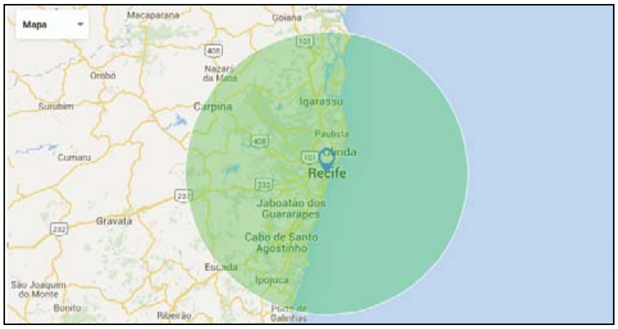


Fig. 6. PADSS Limited Risk Zone Circle

Finally, based on the information extracted from the log files generated by each HTTPS endpoint, the following metrics have been collected:

TABLE I. EVALUATION METRICS

Registered HTTPS endpoints	*Notified HTTPS endpoints*	*Processed Data Amount*	*Data Analysis Execution Time(s)*	*Message Delivery Execution Time(s)*	*Total Execution Time(s)*
50	37	300MB	6.321	6.862	13.183

A few observations must be made based on the results gathered. First, the SQL queries used in this experiment needs to be tuned in order to process real large datasets. Regardless of the capacity offered by Google Cloud infrastructure, when processing large datasets in such time constrained scenarios, it is advisable to improve your queries to optimal levels. Second, this present research is not addressing statistical errors due to the computation of false positive results, considering that a wrong diagnosis may have been given by a physician. Last, for accurate results, profile settings must be enabled in order to use adaptive rules for decision making.

5 Conclusions

This paper has introduced an analytical decision support system, designed to perform predictive analysis using big data technologies combined with large scale event processing. There is an increasing need for real-time processing of analytical data that improves pattern recognition and trends prediction using very large historical datasets. In the context of this work, merging big data predictive analysis with real-time event processing improves our ability to predict or identify outbreak occurrences at early stages. Therefore, healthcare organizations are better equipped to respond to these occurrences in a swift and appropriate manner.

This platform uses the FHIR specification in order to exchange system-level notifications, enabling the massive dissemination of real-time outbreak reports using a standards-based messaging model. Since the beginning of this research, the FHIR standard has evolved significantly. There is much work left to be done as demonstrated by [21], but we are optimistic about the outcome.

The results presented demonstrate the benefits of using an approach that integrates big data analysis and streaming data processing, achieving a higher level of accuracy by matching transactional and historical data. The possibility of performing real-time predictive analysis even for very large datasets is a major advantage when compared with traditional data mining strategies. Coupling the outcomes of analytical data processing with a rule-based CEP engine improves our ability to perform pattern recognition or trends prediction, in a much faster and scalable manner, especially in scenarios that involves high-frequency data such as electronic medical records. As future works, we intend to perform field experiments using real heterogeneous data, given that in the scope of this paper we were unable to perform experiments with larger and richer datasets.

6 References

[1] Silvino Neto, Márcia Valéria, Plínio Manoel, and Felipe Ferraz, "Publish/Subscribe Cloud Middleware for Real-Time Disease Surveillance," 10th International Conference on Software Engineering Advances, 2015, pp. 131-138.

[2] D. Bender and K. Sartipi, "HL7 FHIR: An Agile and RESTful approach to healthcare information Exchange," Computer-Based Medical Systems (CBMS), IEEE 26th International Symposium, 2013, pp. 326-331.

[3] "FHIR Specification v. 1.0.2", 2015. [Online].Available: http://www.hl7.org/FHIR/index.html. [Accessed: 02-Nov-2015].

[4] "Communicable Disease Surveillance and Response Systems", 2006. [Online]. Available: http://www.who.int/csr/resources/publications/surveillance/WHO_CDS_EPR_LYO_2006_2.pdf?ua=1 [Accessed: 04-Oct-2015]

[5] Eric Mykhalovskiy, Lorna Weir, "The Global Public Health Intelligence Network and Early Warning Outbreak Detection," Canadian Journal of Public Health, 2006, v. 97, n.1, pp. 42-44.

[6] National Electronic Disease Surveillance System Working Group, "National Electronic Disease Surveillance System (NEDSS): a standards-based approach to connect public health and clinical medicine," Journal of Public Health Management and Practice, 2001, v. 7, n. 6, pp. 43-50.

[7] "Health Level Seven International". [Online]. Available: http://www.hl7.org/. [Accessed: 02-Nov-2015]

[8] V. Vaidehi, R. Bhargavi, K. Ganapathy, and C.S. Hemalatha, "Multi-sensor based in-home health monitoring using Complex Event Processing," International Conference on Recent Trends In Information Technology (ICRTIT), 2012, pp. 570-575.

[9] D. Robins, "Complex event processing," Second International Workshop on Education Technology and Computer Science, Wuhan, 2010. [Online]. Available: http://goo.gl/jLROpc . [Accessed: 30-May-2015].

[10] "Google Big Query". [Online]. Available: https://cloud.google.com/bigquery/. [Accessed: 30-Oct-2015].

[11] Ricardo Jorge Santos and Jorge Bernardino, "Global Epidemiological Outbreak Surveillance System Architecture," 10th International Database Engineering and Applications Symposium, 2006, pp. 281-284.

[12] Yingmei Qi, Lei Cao, Medhabi Ray, and Elke A. Rundensteiner, "Complex event analytics: online aggregation of stream sequence patterns," ACM SIGMOD International Conference on Management of Data, 2014, pp. 229-240.

[13] Matteo Migliavacca et al., "Distributed Middleware Enforcement of Event Flow Security Policy," ACM/IFIP/USENIX 11th International Middleware Conference, 2010, pp. 334-354.

[14] Eva Zámečníková and Jitka Kreslíková, "Design of Adaptive Business Rules Model for High Frequency Data Processing," in Information Systems Architecture and Technology: System Analysis Approach to the Design, Control and Decision Support, Wrocław, Poland: Oficyna Wydawnicza Politechniki Wrocławskiej, 2014, pp. 75-84.

[15] Javier Andreu-Perez et al., "Big Data for Health," IEEE Journal of Biomedical and Health Informatics, 2015, v. 19, n. 4, pp. 1193-1208.

[16] S. I. Hay, D. B. George, C. L. Moyes, and J. S. Brownstein, "Big data opportunities for global infectious disease surveillance," PLoS Medicine, 2013, v. 10, n. 4, e1001413.

[17] "Epidemico – Website", 2015. [Online]. Available: http://www.epidemico.com/. [Accessed: 28-Oct-2015].

[18] Li Anying, Chen Ke, Song He, and Lei Yu, "The Industry Data Analysis Processing Model Design: The Regional Health Disease Trend Analysis Model," International Conference on Cloud Computing and Big Data, 2014, pp. 130-133, doi: 10.1109/CCBD.2014.11.

[19] Kathy Lee, Ankit Agrawal, and Alok Choudhary, "Real-Time Disease Surveillance Using Twitter Data: Demonstration on Flu and Cancer," 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 1474-1477, doi: 10.1145/2487575.2487709.

[20] M. Maier, "Towards a Big Data Reference Architecture," Master's Thesis, Eindhoven University of Technoly, Eindhoven, The Netherlands, 2013.

[21] "FHIR v.1.0.2 – Outstanding Issues", 2015. [Online]. Available: http://hl7.org/fhir/todo.html. [Accessed: 02-Nov-2015].

[22] Weisstein, Eric W. "Spherical Trigonometry." From MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/SphericalTrigonometry.html [Accessed: 02-Nov-2015]

[23] R. Snacken, J. Lion, W. Aelvoet, "Five years of sentinel surveillance of acute respiratory infections (1985–1990): The benefits of an influenza early warning system", European Journal of Epidemiology, 1992, v. 8, n. 4, pp. 485-490.

SESSION

MOBILE DEVICES, CLOUD COMPUTING, COMMUNICATION SYSTEMS AND NETWORKS

Chair(s)

TBA

An Efficient Algorithm for a Self-Organizing Mobile Sensor Network Based on RSSI

Mohammad Nurul Afsar Shaon, Ken Ferens, Marcia Friesen, and Robert Mcleod
Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba, Canada
{shaonmna@myumanitoba.ca, ken.ferens@umanitoba.ca, marcia.friesen@umanitoba.ca, robert.mcleod@umanitoba.ca}

Abstract— One significant research aim in mobile wireless sensor networks is to resolve coverage issues in order to use computation and communication resources efficiently [1] . *This paper presents a self-governing mobile sensor node distribution model based on RSSI. The proposed model does not require any special message exchange protocol and it is able to distribute mobile sensor nodes uniformly within the target area. The performance of the proposed algorithm is evaluated by taking account of coverage and uniformity within targeted area. The results are compared with the performance of uniform mobile sensor distribution. The simulation results validate the model performance in distributing mobile sensor nodes uniformly within the area of interest.*

Keywords—RSSI; mobile sensor node; uniform sensor distribution.

I. INTRODUCTION

‘Distributed systems’ is one of the most significant and important terms of today’s modern communication and computing systems [2]. A distributed system refers to a system that has individual autonomous units to perform a task collectively. Those autonomous bodies are able to communicate with each other by exchanging information through wired or radio channels. In a distributed system, those self-governing entities may perform the same or diverse jobs from one another, assigned by the system. Furthermore, they are given some degree of freedom to execute their assigned tasks. However, some nodes are equipped with special hardware with respect to the other nodes to perform a specific job. Furthermore, the main feature of the distributed system is coordination among the system elements in order to use computation and communication resources of the self-directed node more efficiently. In a large distributed system, a global view of the network or system is not known to a node. A node has only the local view of the system to enhance the performance of the system (e.g. monitoring the temperature of the field) [2]. There are numerous reasons behind the popularity of distributed approaches. Firstly, a distributed approach enhances the efficient use of the computation and communication resources. Secondly, it also facilitates security and the privacy issues of the system. More importantly, a distributed approach makes the network robust during sensing and communication failures of an individual node(s) [3].

A wireless sensor network (WSN) is simply a pool of self-directed devices which sense a given parameter and which are organized into a mutually connected network. Sensors are usually autonomous and spatially distributed within a certain area to monitor targeted physical and environmental phenomena, such as temperature, sound and pressure. They have a small degree of computational resources and are capable of communicating with each other through a radio channel. In a WSN, sensor nodes collect data from the region of interest and transfer those collected data to the base station. However, a sensor node has limited communication and battery resources. To save battery life, a sensor node transfers the collected data to a base station through multiple hops, instead of single hop [4]. Therefore, a sensor node has two major tasks. It not only collects and transmits data from the deployed area, but, also, it collects and forwards data packets on behalf of the other sensor nodes to the base station. A distributed approach not only optimizes its energy resources but also it helps prevent memory overload issue (storing too much previous data) of sensor nodes. Therefore, from a mobile sensor network’s point of view, a distributed approach is an efficient solution for utilizing the maximum potential of this kind of network.

One of the most significant issues of the mobile sensor network is how to optimally distribute the mobile sensor nodes in the region of interest where conditions are to be monitored [3]. As such, mobile sensor node distribution within the target area is an active research topic for researchers in this field for the last few decades. This can be challenging due to the nature of the mobile sensor nodes and the uncertainty about the environment in which they are deployed. Mobile sensor nodes can be deployed randomly from aircraft, for example, or transported by ground to the region of interest as the topography dictates. In an aerial drop, for example, exact positioning is not possible [1]. Though random sensor distribution does not provide uniformity within the region of interest, uniform sensor distribution remains the most desirable situation for any kind of sensor deployment due to the establishment of the routing structure. Ideally, sensors have a mechanism to distribute themselves in the target zone uniformly. The objective of this research is to design an algorithm by which mobile sensor nodes are able to self-organize, that is, to distribute themselves uniformly within the target area considering computation and energy constraints.

In this paper, we propose a new mobile sensor distribution scheme based on received signal strength indicator (RSSI) to enhance the network lifetime and energy consumption. The proposed mobile sensor distribution scheme is able to distribute mobile sensor nodes uniformly within a target area, where sensors are initially randomly deployed at the center of the region of interest. Here, a mobile node can change its initial position based on the RSSI signal received from its neighbor nodes. Simultaneously, other mobile sensors can do the same. All sensor nodes keep on changing their position until they are informed to stop by the base station.

The performance metrics of the proposed mobile sensor distribution algorithm are analyzed and compared with a uniform sensor distribution through detailed simulations. The simulation results confirm that an RSSI based scheme is able to distribute the mobile nodes effectively, with performance characteristics similar to a known uniform distribution of nodes.

The rest of this paper is organized as follows: Section 2 presents a literature survey on existing sensor distribution models. We discuss RSSI in Section 3 accordingly. The proposed RSSI based sensor distribution scheme (i.e. algorithm) is described in detail in Section 4. The simulation results are discussed in Section 5. Section 6 concludes the paper and provide a scope of future work.

II. Related work

As an active research area, there is related literature on the problem of optimizing node placement within a distributed sensor network. For example, in [5], a mobile sensor distribution model is described with the objective to achieve maximum coverage within the region of interest, assuming >>>>. Each of the sensor nodes is equipped with a GPS device. A dynamic algorithm finds the optimal position for each sensor node so that they can communicate efficiently. This scheme is not cost effective due to the addition of the GPS device on an inexpensive sensor node.

In [6], three distinct algorithms have been proposed to improve the coverage performance within the target area. A combination of static sensor nodes and mobile sensor nodes is used in this scheme to distribute nodes. Three distinct algorithms are used to select the number of sensor nodes required to cover the whole area. In this proposed scheme, mobile nodes are not taken into account to determine the targeted area boundary.

A proxy based protocol has been proposed in [7] to minimize the energy used by mobile sensor nodes to reach their optimal position in a target area. Logical movement patterns of a node are computed before a mobile sensor moves from its initial position, in order to minimize the steps taken to reach its optimal position. However, the positions of the mobile sensor nodes are noteworthy information for the execution of this kind of algorithm.

In [8], the authors propose a simple sensor distribution scheme based on Voronoi algorithm to move mobile sensor nodes to the location of events within targeted region for monitoring. At the same time, this algorithm updates the nodes about the boundary condition of the target area for maximal coverage. Similar to others ([5]-[8]),sensor nodes are assumed to know their position by other means, such as GPS.

In [9], a sensor distribution scheme has been introduced to obtain maximum coverage in minimum time to control the movements of the mobile sensor nodes. The idea is to distribute the mobile sensor nodes uniformly over the target area within the minimum amount of time by implementing the minimum possible change of a sensor node's position from their initial location. In this case, all the sensor nodes are aware of their location by GPS.

Overall, most of the proposed schemes rely on the special hardware like GPS, which makes the system expensive. On the other hand, the applying multiple algorithms for the optimal placement of the mobile nodes is computationally expensive and time-consuming. Sometimes sensor nodes are deployed in the hostile area to monitor environmental conditions. In that case, Deployment of both static and mobile sensor nodes as a combination is not a practical solution for the placement of the sensor nodes (static nodes are used as the reference node). Therefore, an algorithm is required to optimize the node placement, which is computationally inexpensive, efficient, and less dependent on special hardware.

III. RECEIVED SIGNAL STRENGTH INDICATOR

Received signal strength indicator (RSSI) is used to measure the strength of a radio signal coming from a neighboring sensor node or nodes [10]. RSSI is tightly related to the distance between two sensor nodes. As the distance (d) between the sender and the receiver increases, the RSSI magnitude at the receiver sensor end declines accordingly [10] . In other words, it can be stated that the propagation loss is correlated to the distance (d) between nodes. The mathematical expression of this relationship can be stated as follows.

$$P_r(d)(dbm) = P_r(d_0)(dbm) - 10nlog\left(\frac{d}{d_0}\right) + X(dbm) \quad (1)$$

Where d_0 is the reference distance from the receiver end, $P_r(d_0)$ represents the received radio signal energy at the reference distance (d_0), n presents the path loss coefficient which is related to the environment, and X represents a variable whose value is taken from the normal distribution. A simplified version of the above equation can be expressed as follows

$$P_r(d)(dbm) = P_r(d_0)(dbm) - 10nlog\left(\frac{d}{d_0}\right) \quad (2)$$

If it can be assumed that A is the received radio signal energy at the reference distance $(d_0 = 1m)$, then the equation can be rewritten as follows

$$P_r(d)(dbm) = A - 10nlog(d) \quad (3)$$

IV. PROPOSED ALGORITHM

In this work, a unique algorithm is proposed which will distribute mobile sensor nodes within the target region using RSSI, when the nodes are initially randomly deployed at the center of the target zone. Each sensor has a mechanism to measure the received radio signal strength generated from the neighboring sensor node. In this algorithm, this feature of the mobile sensor node has been exploited. If the sum of the received signal at the newly selected position is higher with respect to the previous position, it indicates that sensor density around the selected sensor is higher at the new location relative to the previous location. Therefore, to achieve uniform sensor distribution, it is important to place the mobile sensor node in less densely populated areas. This principle drives a mobile sensor to find an optimum location within the target area. The advantage of the proposed algorithm is that it doesn't need any special message exchanging protocol that causes significant overhead in the network operation. We also assume that each sensor node is aware of its initial position, the boundary of the target area, and any obstacles in the area that may restrain movements. All of the sensors must update their position according to the algorithm and inform the base station of their new positions. This "spreading out" process from their initial position continues until the base station instructs them to stop.

The details of the algorithm are given below.

1. Set the input parameters of the algorithm
   Number of sensors (N)
   Radio range (r)
   Initial sensor deployment zone $(Km \times Mm)$
   Total coverage area $(Rm \times Pm)$

2. N Number of sensors which are deployed initially at the center of the targeted area.

3. Select the ith sensor node among N sensor nodes. It is assumed that ith sensor node knows its coordinate, (x, y) ,within the target region area

4. The ith sensor stores the RSSI signal level coming from the other neighboring sensor nodes at its initial position and sums them up.

$$P_r(d)(dbm) = A - 10nlog(d)$$
$$P_{rtotal\,(initial)} = \sum_{i=1}^{N} RSSI_i$$

5. Randomly select any position $[(x_1, y_1)coordinate]$ at the edge of its communication zone from the initial position; store the RSSI level coming from other sensor nodes around it and takes the summation of the received RSSI.

$$P_{rtotal\,(new)} = \sum_{i=1}^{N} RSSI_i$$

   If the summation of the RSSI at newly selected position is less than the summed RSSI values at previous position $\left(P_{rtotal\,(new)} < P_{rtotal\,(intial)}\right)$, then change the position of the ith sensor node at the newly selected coordinate.

$$(x, y) = (x_1, y_1)$$

6. Repeat the procedure from 3 to 5 for all sensor nodes.

7. Repeat the whole procedure from 3 to 6 until all the sensor nodes are instructed to stop.

V. Simulation and Results

This section presents the simulation setup and performance of the proposed self-organizing mobile sensor distribution algorithm within the region of interest. The first phase of the simulation was conducted to see if the sensor nodes are distributed with a certain area efficiently after applying the proposed model. In the second phase, the performance of the proposed algorithm has been evaluated and compared with the performance of an initial uniform sensor distribution. In addition, all simulation works are conducted in MATLAB.

In the first simulation setup, 30 mobile sensor nodes were deployed at the center of the given area, simulating an aerial drop in which they are distributed randomly within a 400 square meter area around the center of the targeted zone. In fact, sensor nodes, which are deployed at the center of the target area, are required to spread out within 100m by 100m area in order to cover the whole region of interest. Each of the sensor nodes has a radio range of 50 meters. Moreover, all of the sensor nodes are aware of the obstacles and the boundary of the given area.

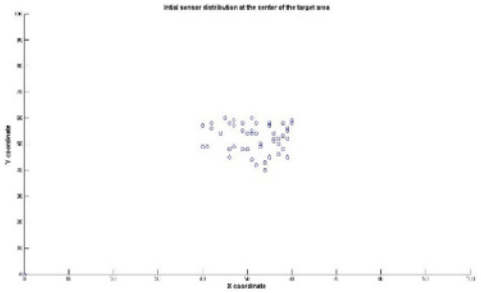


Fig. 1 Initial mobile sensor deployment

Fig. 1 presents the initial sensor distribution at the center of the target area. After applying the proposed algorithm based on RSSI, sensors are successfully distributed all around the target area efficiently and they are settled within the boundary of the target area (Fig. 2). Each of the sensor nodes has changed only 10 new positions from its initial position before reaching their optimal positions.

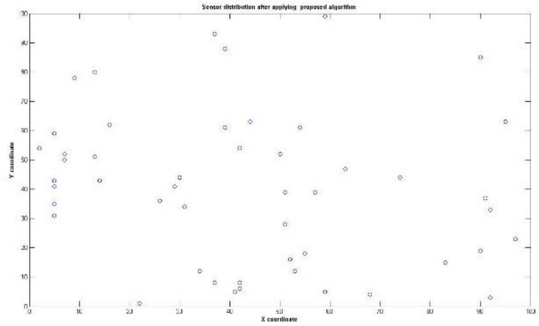


Fig. 2 Position of mobile sensors after applying proposed algorithm

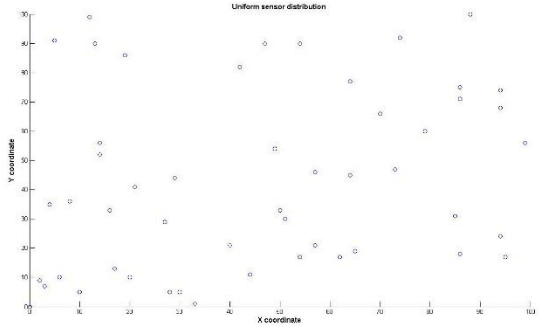


Fig. 3 Uniform sensor distribution within 100 meters by 100 meters area

At the same time, Fig. 3 presents the location of the sensor nodes within 100 meters by 100 meters area when they are distributed uniformly. If Fig. 2 and Fig. 3 are compared, then it can be stated that sensor nodes are distributed similarly to a uniform sensor distribution after applying proposed RSSI based sensor distribution algorithm.

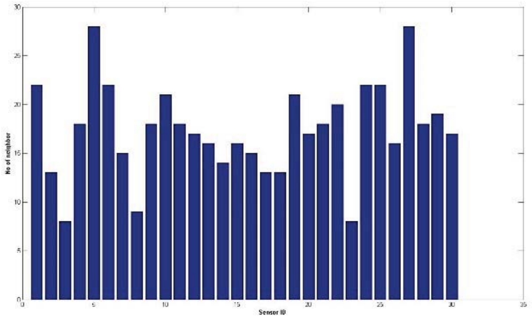


Fig. 4 Neighborhood counts per sensor node (applying proposed algorithm)

Subsequently, another performance metric was evaluated for the proposed algorithm based on RSSI, namely is the average number of neighbors per sensor node. Fig. 4 presents the neighborhood counts of each sensor node when they have reached their optimal position after applying the proposed algorithm, with the average number of neighbors per sensor node being 17.44.

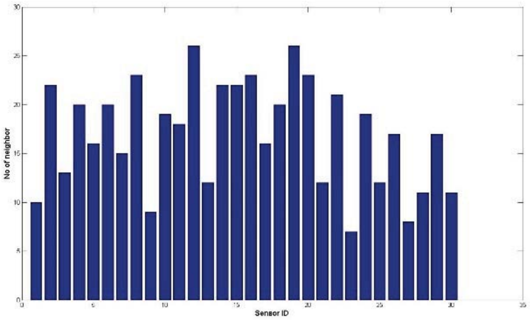


Fig. 5 Neighborhood counts per sensor node (uniform sensor distribution)

Similarly, Fig. 5 shows the neighborhood count of each sensor nodes within target region when they are uniformly distributed (like Fig. 3). In comparison to the proposed RSSI-based sensor distribution algorithm, the average neighborhood counts per sensor node are 17.23 where sensors are distributed uniformly. Therefore, it can be said that proposed RSSI based algorithm successfully distributes the mobile sensor node uniformly within the targeted area.

VI. Conclusion

This paper has presented a self-organizing mobile sensor node distribution scheme based on RSSI to place sensors within a target area. This proposed scheme is able to distribute mobile sensor nodes uniformly within the region of interest. The application is helpful for establishing seamless coverage and routing protocol among sensor nodes. This scheme doesn't require any special message exchange protocol and uses a minimum number of steps to distribute nodes from their initial deployment in the center of the target area. Simulation results validate the performance of the newly proposed algorithm. In the future, this algorithm will be tested for different numbers of mobile sensor nodes and with varying radio ranges.

References

[1] C. W. Yu, C.-H. Wang, L. C. Hsu, and K. J. Cheng, "Coverage algorithms in GPS-less wireless mobile sensor networks," *Proc. Int. Conf. Mob. Technol. Appl. Syst. - Mobil. '08*, vol. 2008, p. 1, 2008.

[2] C. Science, E. S. Campus, and G. Ram, "Distributed algorithms for sensor networks," pp. 11–26, 2012.

[3] N. Heo and P. K. Varshney, "Energy-efficient deployment of Intelligent Mobile sensor networks," *IEEE Trans. Syst. Man Cybern. Part A Syst. Humans*, vol. 35, no. 1, pp. 78–92, 2005.

[4] A. A. Taleb, T. Alhmiedat, O. A. Hassan, and N. M. Turab, "A Survey of Sink Mobility Models for Wireless Sensor Networks," vol. 4, no. 9, pp. 679–687, 2013.

[5] S. Zhou, M.-Y. Wu, and W. Shu, "Finding optimal placements for mobile sensors: wireless sensor network topology adjustment," *Proc. IEEE 6th Circuits Syst. Symp. Emerg. Technol. Front. Mob. Will. Commun. (IEEE Cat. No.04EX710)*, vol. 2, pp. 529–532, 2004.

[6] M. Zhang, X. Du, and K. Nygard, "Improving Coverage Performance in Sensor Networks By Using Mobile Sensors."

[7] G. Wang, G. Cao, and T. La Porta, "Proxy-Based Sensor Deployment for Mobile Sensor Networks," *Proc. 1st IEEE Int. Conf. Mob. Adhoc Sens. Syst.*, pp. 493–502, 2004.

[8] Z. Butler and D. Rus, "Controlling mobile sensors for monitoring events with coverage constraints," *IEEE Int. Conf. Robot. Autom. 2004. Proceedings. ICRA '04. 2004*, vol. 2, no. April, pp. 1568–1573, 2004.

[9] P. K. Varshney, "An intelligent deployment and clustering algorithm for a distributed mobile sensor network," *SMC'03 Conf. Proceedings. 2003 IEEE Int. Conf. Syst. Man Cybern. Conf. Theme - Syst. Secur. Assur. (Cat. No.03CH37483)*, vol. 5, pp. 4576–4581, 2003.

[10] W. Chengdong, C. Shifeng, Z. Yunzhou, C. Long, and W. Hao, "A RSSI-based probabilistic distribution localization algorithm for wireless sensor network," *2011 6th IEEE Jt. Int. Inf. Technol. Artif. Intell. Conf.*, no. 60874103, pp. 333–337, 2011.

Implementation and Verification of QoS Priority over Software Defined Networking

Sun Uk Baek, Chan Ho Park, Earl Kim, Dong-Ryoel Shin
Department of Information & Communications Engineering
Sungkyunkwan University, Republic of Korea
{sunuk100, pch3114, wise0314, drshin}@skku.edu

Abstract - ***Sofware-Defined Networking (SDN) is a new paradigm that can implement various network functions based on the software. SDN technologies are evolving rapidly, and can be implemented to embrace the existing network function. Providing QoS guarantee to various applications is an important issue among the many network functions. The goal is to provide reliable service appropriately allocating bandwidth for each application traffic. In this paper, over Software-Defined Networking environment, we verify and implement QoS guarantee function that is being used in traditional network. Using the OpenFlow protocol and the proper queuing techniques, we present that the QoS guarantee is operating normally with a few scenarios and the traffic of the different conditions over SDN environment.***

Keywords: Software-Defined Networking, OpenFlow Protocol, QoS provisioning, bandwidth guarantee

1 Introduction

Software-Defined Networking (SDN) is a new paradigm that can implement various network functions based on the Software [1]. The traditional network equipment vendors have provided all of the hardware, software and management tools. But, SDN is a totally different way to manage the network in a centralized form. It is based on the principle of separating the control and data planes, and the protocol like OpenFlow describes the information exchange between the two planes [2]. Separating control plane makes a wide variety of applications to be applied not dependent on hardware. This method, it prevents from being dependent on one vender, makes it possible to construct an efficient and intuitive network.

Structure of the SDN is made up of the control plane and data plane. Control plane makes decisions about where traffic is sent and the control plane functions include the system configuration, management, and exchange of routing table information. Above the control plane, the user can be added that the application of the various functions desired via north-bound interface [3], [4]. Various controller performing a control plane exists and has been newly developed. Data plane (a.k.a. Forwarding plane) forwards traffic to the next hop along the path to the selected destination network according to control plane logic. Control plane and Data plane can communicate through the south-bound interface. And OpenFlow protocol among a number of protocols is used in the most popular.

SDN technologies are evolving rapidly, and can be implemented to embrace the existing network function. It is moving in a direction to add more advanced features while using existing network functions. Providing QoS guarantee to various applications is an important issue among the many network functions. Effort for the efficient use of the limited bandwidth and service guarantee for real-time applications has persisted. The goal is to provide reliable service appropriately allocating bandwidth for each application traffic. In this paper, using these controller, OpenFlow protocol, and the proper queuing techniques, we verify and implement QoS guarantee function that is being used in traditional network.

The rest of the paper is organized as follows. Section 2, we briefly discuss the background about our topic such as controller and queuing technique. Section 3 presents our proposed test-bed environment. In Section 4, we show the result of experiment validation. Finally, Section 5 concludes our work.

2 Background and Related Work

2.1 Mininet and Simulation Model

Mininet creates a realistic virtual network, running real kernel, switch and application code, on a single machine (VM, cloud or native) [5]. Because it can easily interact with network using the mininet CLI (and API), customize it, share it with others, or deploy it on real hardware, mininet is useful for development, teaching, and research. And it is also a great way to develop, share, and experiment with OpenFlow and SDN systems. Test bed is composed of several virtual hosts and virtual switches and the controller.

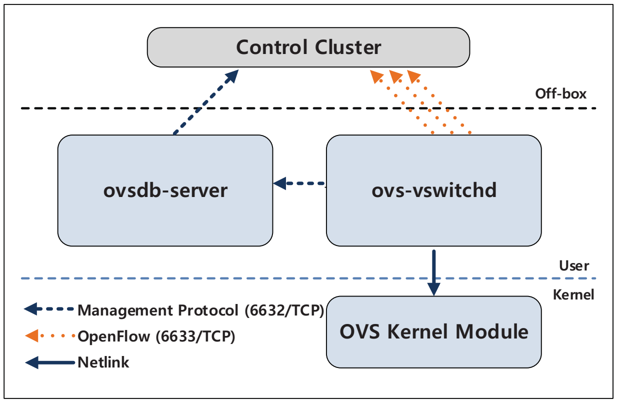
2.2 RYU Controller

Ryu Controller is an open, software-defined networking (SDN) Controller designed to increase the agility of the network by making it easy to manage and adapt how traffic is handled [6]. It provides software components, with well-defined application program interfaces (APIs) that make it easy for developers to create new network management and control applications. Ryu supports various protocols for managing network devices, such as OpenFlow, Netconf, OF-config, etc.

2.3 Open vSwitch

There are various kinds of virtual switches, such as Linc[7], Ofsoftswitch13[8] and OpenvSwitch[9]. Open vSwitch (OVS) is an open-source software-based virtual switch to the functions of the multi-layer network switch. While participating in the Openflow Protocol standardization, some of the functions are rapidly applied. In addition, a variety of tunneling method (GRE, VXLAN, IPSec) also supports. As shown in Figure 1, in Component of OVS is composed of ovsdb-vswitchd, dvsdb-server, kernel module. Ovsdb-vswitchd is a conduit for communicating the switch settings and status information between the kernel module of OVS and SDN controller. Ovsdb-server has the role of data base for storing setting information of the OVS. Kernel module, based on the information received from the controller, is responsible for controlling the flow of the packet.

Figure 1. Main Components of Open vSwitch



OVS supports Hierarchical Token Based (HTB) queuing technique which will be used in this paper. Although Class based Queueing (CBQ) techniques, even if bandwidth can afford, cannot use the more bandwidth as defined. HTB is based on the priority in the proportions defined for spare bandwidth allocated to each class, thereby enabling more efficient use of bandwidth [10].

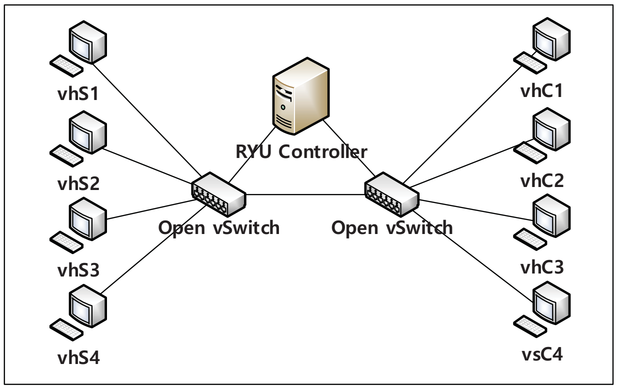
2.4 Iperf

To generate traffic in our experiments, we use the Iperf application. Iperf is a commonly-used network testing tool that can create Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) data streams and measure the throughput of a network that is carrying them [11]. Iperf allows the user to set various parameters that can be used for testing a network, or alternatively for optimizing or tuning a network. Iperf has a client and server functionality, and can measure the throughput between the two ends, either unidirectionally or bi-directionally. It is open-source software and runs on various platforms including Linux, Unix and Windows.

3 Test-Bed Environment

If there is not enough the bandwidth compared to the traffic, to increase the bandwidth is the easiest solution. But, it is required to use the QoS technologies for use in a given bandwidth as efficiently as possible. Bandwidth guarantees and limits depending on the specific features in the traffic flow shall be different.

Figure 2. Test-Bed Environment



In this paper, it distinguishes the traffic to guarantee the bandwidth to the four types. The first type is one necessarily be guaranteed a minimum bandwidth, but the traffic will not exceed more than a certain bandwidth, such as VoIP or Network Management System (NMS) traffic. The second type, except for the bandwidth of first type, is possible to use the remaining bandwidth all the most preferentially, such as streaming traffic. The third type and the fourth type are normal traffic types, but third type traffic is set to a higher priority than fourth type traffic. It can be expressed as Table 1.

Table 1. Characteristics of Flows

Type	*Priority*	*Flow*
1	Highest	vhS1 – vhC1
2	Higher	vhS2 – vhC2
3	Normal	vhS3 – vhC3
4	Low	vhS4 – vhC4

In addition, we set the bandwidth between the Virtual Switches to 1Mbps, and the maximum bandwidth limit was set at 200Kbps in Type#1 traffic. Priority value can be set from 0 to 65533. We set the priority value from 1 to 4 simply, since there are only four flow types. After configuring the test bed of the above conditions, we make three scenarios as follows subjected to verify the operation.

1) Type#1 traffic (300Kbps) occurs in the middle while other traffic flows.
   - Type#1 traffic whose priority is the highest is preferentially performed to guarantee and limit the bandwidth.
   - The bandwidth limit occurs from the lowest priority traffic.
2) Type#2 traffic (800Kbps) occurs in the middle while other traffic flows
   - Except for the highest priority traffic Type#1, use the bandwidth normally.
3) Type#3 traffic (300Kbps) occurs in the middle while other traffic flows
   - Except for the higher priority traffic Type#1 and Type#2, use the bandwidth normally.

4 Experiment Validation

After QoS configuration is set the link between the switches, we apply the configuration of Iperf client and Iperf server to each of virtual host to verify three scenarios.

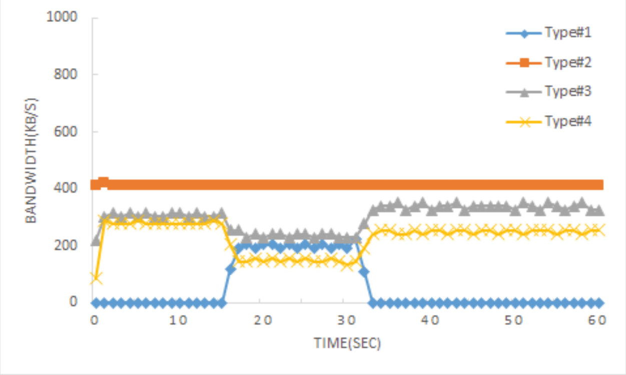
4.1 1st Scenario

Traffic flowed initially set as shown in the table below.

Table 2. Initial Traffic Value of 1st Scenario

	Type#1	*Type#2*	*Type#3*	*Type#4*
Traffic(Kb/s)	-	400	300	300

Figure 3. Validation Test of 1st Scenario



As shown in Figure 3, the Type#1 traffic whose priority is the highest flows in the 300Kbps amount after 15 seconds. Type#1 traffic is guaranteed bandwidth allocation had been set up to 200Kbps. Type#3 and Type#4 traffic with lower priority decreased a lot of bandwidth, especially the Type#4 traffic whose priority is lowest.

In this experiment, if Type#1 traffic come in above the guaranteed bandwidth cannot exceed the limited bandwidth, but it can be confirmed that it is accommodated to the set guaranteed bandwidth.

4.2 2nd Scenario

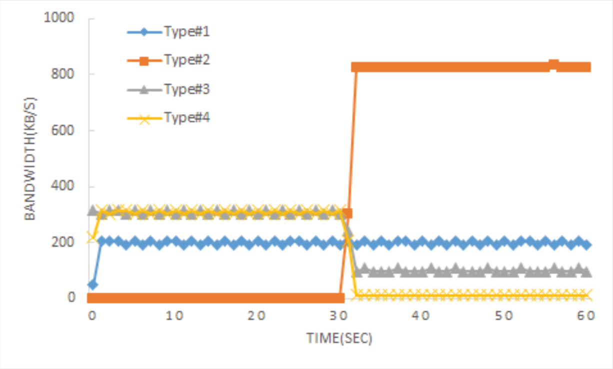
Traffic flowed initially set as shown in the table below.

Table 3. Initial Traffic Value of 2nd Scenario

	Type#1	*Type#2*	*Type#3*	*Type#4*
Traffic(Kb/s)	200	-	300	300

When the traffic was flowing 800Kbps Type#2 traffic after 30 seconds, as shown in Figure 4, without affecting the highest Priority of the Type#1 traffic and it could be confirmed that use both granted 800Kpbs.

Figure 4. Validation Test of 2nd Scenario



4.3 3rd Scenario

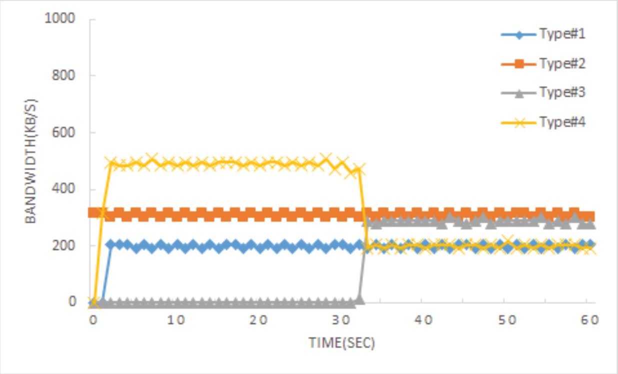
Traffic flowed initially set as shown in the table below.

Table 4. Initial Traffic Value of 3rd Scenario

	Type#1	*Type#2*	*Type#3*	*Type#4*
Traffic(Kb/s)	200	300	-	500

After the initial traffic starts to flow from 30 seconds, Type#3 traffic begins to flow to 300Kbps. As shown in Figure 5, without affecting the higher Priority traffic such as Type#1 and Type#2 traffic, we can see that the Type#3 traffic occupy bandwidth occupied by the Type#4 traffic.

Figure 5. Validation Test of 3rd Scenario



5 Conclusion and Futurework

QoS provision behavior in SDN environment was confirmed that using priority setting operates normally and bandwidth guarantee works effectively. Our future research will involve studies using the meter tables that are provided QoS provision run with OpenFlow version 1.3 in a further variety of conditions.

6 References

[1] Open Networking Foundation, "Software-defined networking: the new norm for networks," white paper, Apr. 2012.

[2] N. Mckeown et. al., OpenFlow: Enabling Innovation in Campus Network, ACM SIGCOMM, 2008.

[3] F. Durr, "Towards cloud-assisted software-defined networking," Technical Report 2012/04, Institute of Parallel and Distributed Systems, Universitat Stuttgart, Tech. Rep., 2012.

[4] M. Mendonca, B. N. Astuto, X. N. Nguyen, K. Obraczka, T. Turletti et al., "A survey of software-defined networking: Past, present, and future of programmable networks," 2013.

[5] Mininet. (2013, Mar). An Instant Virtual Network on your Laptop (or other PC). [Online]. Available: http://mininet.org/.

[6] "Ryu" Component-based software defined networking framework. [Online]. Available: http://osrg.github.io/ryu/.

[7] "Linc" [Online]. Available: https://github.com/ /FlowForwarding/LINC-Switch.

[8] "Ofsoftswitch13" [Online]. Available: https://github.com/CPqD/ofsoftswitch13.

[9] "OpenvSwitch" [Online]. Available: http://openvswitch.org/.

[10] M. Devera, "HTB Linux queuing discipline manual – user guide." [Online]. Available" http://luxik.cdi.cz/~devik/qos/htb/manual/userg.htm.

[11] "Iperf" [Online]. Available: http://iperf.sourceforge.net.

User-Experience of Human Computer Interaction (HCI) in Mobile Phones

Dr. Minerva M. Bunagan[1], **and Prof. Nabil El Kadhi**[2]
[1]Asst. Professor, College of Business, UoB, Buraimi, Oman
[2]Deputy Vice Chancellor for Academic Affairs, UoB, Buraimi, Oman

Abstract – *Users consider some characteristics in selecting a mobile device. These characteristics include interface design, efficiency, reliability, presence of utilities, responsiveness, portability, performance and advanced features. On the other hand, mobile developers apply HCI principles in designing mobile devices and consider the user-experience as well. Mobile users reflected their experience of the 7 HCI principles in their commentaries and it appeared that the most relevant principles are consistency, synthesizability and substitutivity. To specifically determine the effect of a specific design principle on a certain product, the study recommends exploring the prioritization of the HCI principles as implemented by the mobile phone developers. This will pave way for the differentiation of design principles applicable to hardware or software.*

Keywords: Mobile Computing, Mobile Characteristics, Mobile Interaction, Human Computer Interaction, Design Principles, User-Experience,

1. Introduction

Statistics shows that mobile subscribers are widespread all over the world. The International Telecommunication Union estimated that there are 6 billion mobile subscriptions, which make up 87% of the world population [1]. The increase on mobile subscriptions from 2010 to 2011 is 0.6 billion and from 2009 to 2010 is 0.7 billion. It was also estimated that by the end of 2012, there will be 6.5 billion mobile subscribers and 6.9 billion by the end of 2013 [2]. Moreover, in the early 2014 report on mobile users, it was expected that mobile users in 2014 will grow to 4.55 billion worldwide and by the end of 2015 there will be 8 billion mobile subscribers [3]. In addition, it is estimated that the total number of mobile broadband connections will include for almost 70% of the global base by 2020, as compared to 2014, which is around 40% [4]. These data indicate that there is much growth in the use of mobile phones all over the world. This could be attributed to the fact that individuals depend so much on mobile phones today and mobile phones have already become a necessity in the daily lives of the individuals as well.

In the early stage of mobile technology, mobile phones were just used for calls and messages. Today, the advancement in mobile technology is hard to cope with. Mobile phones have a lot of features now, which include text and image messaging, simple calls and video calls, Internet access, GPS-based navigations, music, gaming, social networking and various applications which provide users great help in dealing with daily activities that require easy and reliable information [5].

The presence of various types and brands of mobile phones in the market provides sufficient options to the users. The features available in each mobile phone depend on the operating system, from which it is based. There are mobile devices that are based on IOS, Android, Windows, Blackberry and Symbian, which is now gradually disappearing under the ownership of Microsoft [6].

Most users however choose mobile phones in consideration of the camera, memory capacity, music, video, gaming, downloading, and surfing capabilities. Moreover, they take into account the size, weight, functionality, and ease of use. All of these speak of the experience of the user along mobile interaction.

Mobile interaction is an aspect of Human Computer Interaction (HCI), which deals with the manner to which mobile users interface with computers. It focuses on the understanding of mobile users' requirements and needs [7].

As HCI involves the study on how the users/ human and the computer/ machine interface with each other, it takes into account the relevant aspects of the two (2) elements. For instance, it looks into the techniques in computer graphics, the operating system, the programming language and the development environment of the machine; and the graphic design and industrial design principles, human factor such as user satisfaction and communication theory, among others of the user [8].

The application and involvement of HCI in the design of mobile devices should not be ignored to ensure that mobile devices are appropriately designed. It should also be noted that HCI provides relevant findings along user experience, which help a lot in for the improvement of systems and devices [9].

2. Research Problem

Studies reveal that HCI is a discipline without a fusing design interface, and there is no coherence and consistency in measuring HCI interface performance. However, a study on HCI Principles was conducted to standardize the variety of principles which have been proposed by several authors. The result of the study showed the eight (8) HCI Principles [10], namely: 1) recoverability; 2) familiarity; 3) consistency; 4) substitutivity; 5) task migratability; 6) synthesisability; 7) predictability; and 8) perceptual ergonomics. These principles can be applied to the design of interfaces in mobile devices.

Mobile devices of different kinds and with varied features and interfaces are widely used everywhere. Researches show that there are many challenges faced by those involved in designing mobile devices, which include limited input and output facilities, designing for mobility, hierarchical menus, navigating and browsing, and images and icons [11].

Since mobile devices are used in various disciplines, such as education, medicine and sociology, major emphasis should be placed on how the user interacts with the device. This would only be possible if there is a set of appropriate design principles that can be applied in mobile devices.

Moreover, Mobile phone developers come out with new product releases in response to the growing need of the users along functionality, usability, ease of use, among others. These user requirements are based on the user's experience. Users seek for better ways to improve mobile performance based on their need and personal interaction with the device. Though user experience design is considered in the design of mobile phones, HCI principles should not be taken aside.

As there is no standard set of HCI principles for mobile device design, there is a need to come out with specific design principles or criteria that would be able to meet what the mobile device is intended for. This study explored how HCI principles can be applied to mobile devices by looking into the characteristics users preferred to have from mobile devices and mapping these against the HCI design principles. It also rests on the concept that the design of mobile phones is influenced by the 2 factors – user experience design and HCI Principles.

3. Methodology

The study employed the descriptive research design, where data were gathered in 2 ways - meta-analysis and surveys. The meta-analysis was conducted to review systematically various findings on HCI principles and come out with the concrete principles in HCI. As there were conflicting issues on HCI principles, it is appropriate to extract the conflicting views of various authors and be able to deduce specific principles that serve as basis to continue with the study. Findings were grouped according to the general principles of HCI, which include: recoverability; familiarity; consistency; substitutivity; task migratability; synthesisability; predictability; and perceptual ergonomics. The findings collected along these principles served as basis for comprehensive analysis of the specific descriptions and considerations for each principle. The survey method included on-line literature reviews or feedback on mobile devices from 297 mobile users. They were categorized based on the general characteristics of mobile devices preferred by the users. After taking into consideration the on-line reviews, they were mapped against the HCI principles. An analysis was made as to which HCI principles are applied in the design of the mobile devices.

4. Discussion

4.1 Challenges in Mobile Design

The HCI challenge in the mobile device design lies on its hardware and software design [11]. The limited input and output facilities, and designing for mobility are the hardware challenge, and the hierarchical menus, navigating and browsing, and images and icons are the software challenge.

The three input facilities for mobile devices are keyboard, stylus and the scroll wheel. Among these facilities, the challenge is on the design of the keyboard as the space for it is limited [11]. However, the keyboard allows the user to have more accurate entry than the stylus [12, 13]. An experiment was performed to compare text-based entry using mini-qwerty, hand-writing recognition, quick writing, and a chorded keyboard; and it was found out that the mini-qwerty is the fastest way in entering text and the two-chorded keyboard is the best of the methods in terms of accuracy [12].

With the problem on the space taken by the keyboard, due to the small size of the mobile device, new designs of mobile devices have been out in the market, where the touch screen keyboard replaced the physical keyboard. A very good example for this is the iPhone and the Android Phones. An experiment on the comparison of a physical keyboard and a touch screen keyboard was conducted, where the Palm Treo was used for the physical keyboard and the Samsung i718 was used for the touch screen keyboard. It was found out in the study that the use of physical keyboards proved to be faster and more accurate in entering text as compared to the touch screen keyboard. However, the addition of tactile feedback in the touch screen keyboard gave a comparable result with the physical keyboard. The authors then recommended that tactile feedback be added as feature of the mobile device [13].

A study on tactile feedback was also conducted and it was reinforced that tactile feedback helps a lot in improving the user performance with the mobile. In the study, the users were able to find greater satisfaction in interacting with the touch screen [14].

Various studies have shown that users still prefer to use keyboards for entering text, as it proved to be accurate and faster [11, 13]. The challenge then on the limited space for the keyboard led to the design of touch screen keyboard and the said type of keyboard is almost the same as the physical keyboard with the presence of the tactile feedback, and like the physical keyboard, it brought user performance, accurate and faster text entry.

Moreover, two output facilities which are considered limited are screen and audio [10]. The challenge as regards limited output facilities lies on the size of the screen. A study on the mobile device design in response to the difficulty of viewing huge information from many applications was conducted and it was found out that the combination of fish-eye and semantic zooming techniques would greatly help the user visualize the amount of information, even with a small screen [15]. There is also a challenge in viewing maps, photographs and other huge amount of information, and this is the difficulty in doing spatial activities. Although users can utilize the panning and the zooming techniques for viewing such huge information, these activities are considered tedious in working with spatial tasks. The three techniques for spatial activities – Halo, Arrows and Citylights were compared and it was found out that Halo and Arrow-based visualizations do not vary differently in working on simple spatial activities [16].

On mobility, since a mobile device is carried by the user anywhere and anytime, portability and the power facility should be primarily considered. There are varied instances by which the power is consumed and the consumption is based on the desired applications or performance level of the user [11].

Due to the challenge brought about by the power facility of mobile devices, the challenges of the power and thermal architecture were reviewed, and at the same time design considerations for mobile hardware and software to ensure performance and maximize the power were proposed [17].

Hierarchical Menus are list of options in several levels, available for the user to perform certain tasks. These are typically selected by clicking or tapping the menu, in the case of touch screens [18]. The challenge on hierarchical menus has something to do with its structure. It was mentioned that researches have been conducted regarding the number menus and the submenus in the mobile phones; however, there are conflicting ideas as to which structure provides more efficient interaction [11].

Barell Menu as an alternative to hierarchical menu was proposed. Its prototype was designed and its usability was compared with devices having hierarchical menu. It was revealed that it takes fewer steps to navigate through the menus using their prototype. Though the Barrell Menu proved to be efficient more than the hierarchical menu, the users still prefer the hierarchical menu as they are used to it already [18].

The challenge on navigating and browsing is related to the limited output facility as there is a difficulty along these due to the size of the screen [11]. There are problems in mobile browsing, particularly on viewing information as compared to the desktop. Designs should be taken into account by mobile website designers to improve readability [19].

Images and icons as well pose a challenge in the mobile design as downsizing them from desktop to fit mobile devices should not be ignored. Various studies, which gave considerations in designing images and icons for mobile devices were conducted though [11].

4.2 HCI Design Principles

A review and analysis of HCI design principles was done and 8 design principles were identified. These are recoverability; familiarity; consistency; substitutivity; task migratability; synthesisability; predictability; and perceptual ergonomics [10].

Recoverability principle involves ways by which the user is able to recover from an error he usually makes. There are two types of error recovery -- forward and backward. In forward recovery, the user is able to prevent making the error; and in backward recovery, the user is able to reverse the error made.

The familiarity principle focuses on the experience and understanding of the user in interfacing with the system. It tries to determine how much the user knows about the system and compares to the appropriate or prescribed way of using the system.

Consistency involves how the user uniformly behaves or performs tasks in similar scenarios. It considers the actions of the user in interfacing with the input and output facilities of the system, among others.

Substituvity is providing the user with an option or alternative to execute a task. By choosing either of the options, the user is able to achieve the same thing. An example of this is using a certain icon or a hierarchical menu to open an inbox.

Task migratability pertains to the state of allowing the system to do the task that can be done by the user. The user transfers some responsibilities to the system. A good example of this is allowing the system to check the grammar of the document, which could also bring some drawbacks as there are instances when the system is not able to understand or convey what is exactly meant by the user.

Synthesisability provides the user the ability to simulate what happens in interfacing with the system. The user is able to check or trace what happens after asking the system to do some tasks. An example of this is when the user performs transferring of the files, the user is able to know whether the file is transferred when the file is already in the desired location.

Predictability enables the user to determine what happens whenever the user executes a task. The user is able to perceive the result after selecting a menu or pressing a key.

Perceptual ergonomics allows the system to detect human perceptions. The interface should have the facility of monitoring how the user recognizes things.

4.3 Characteristics of Mobile Phones Required by the User

Mobile phone characteristics required by the user can be categorized into nine (9): interface design, efficiency, reliability, presence of utilities, responsiveness, portability, performance, functionality and advanced features.

Interface refers to the over-all look and feel of the users on the device. It includes the ability to easily use, learn and recognize the appropriateness of the device. [20] The mobile phone users look for simplicity, ease of use, access and operation. They appreciate mobile phones which allow them to easily learn and navigate around the phone. They like mobile phones which are easy to operate, control and conform to their expectations. They attribute this to the type of the Operating System in the mobile and the design of the mobile interface itself. The mobile users specifically mentioned in their testimonies that they stick to a specific mobile device due to their familiarity on the device. Some users also mentioned that the kind of interface relies primarily on the Operating System itself. In addition, mobile phone users focus on the interface aesthetics, where they can find pleasure or satisfaction on the way they interact with the device.

Efficiency includes the ability of the device to accurately and completely allow the users to achieve their goals with the given resources. This can be influenced by 3 factors – performance efficiency, time behavior and capacity. The performance efficiency pertains to how the device performs under the given conditions. Time behavior refers to the amount of time the device is able to respond or process in achieving the goal. Capacity refers to the maximum limit of the devise in meeting the requirements. [20] This characteristic allows the user to be able to achieve the tasks he expects from the phone with minimum effort, time and resources. This includes the ability of the mobile phone to check and edit mails and messages at a certain period, to use the camera as it is expected to function, to have a battery which does not run out of charge for a very short of time, to make and receive calls properly. Mobile users choose phones which can provide them the satisfaction to perform all the tasks they expect from the phone without any hassle or problem. They complain about phones whose battery drains so fast. They appreciate their phones for being able to do the expected tasks at minimal effort and time.

Reliability refers to the degree to which the device is able to perform specified functions or tasks for a certain period of time. This includes the ability of the device to perform as intended in spite of problems on hardware or software. [20] Mobile phone users appreciate phones which have provision for extending battery and memory to allow them to be able to perform various functions with the phone without interruption. Some users also appreciate the unique feature of a removable battery which they can keep charging for times when they do not have access to charger or outlet.

Presence of utilities includes the availability of options or alternative tasks which the user can use as needed. Mobile users require buttons or options for some tasks, such as playback and ring back. They mentioned that it is easier for them to have options to remove unwanted applications, and design. They also want a provision for other browsers in the same phone. Although there are also mobile phones which are redundant in menus and icons, they still would prefer other alternatives to do tasks. For instance, in entering a password for the device, they can use either codes or by the finger prints. Moreover, the presence of options for battery saving, for input of data, and for notification drop down is among the most visible provision of alternatives for the mobile phones.

Responsiveness in this context refers to the ability of the device to work or perform quickly or consistently in the presence of external conditions such as speed and load. Mobile phone users appreciate that their mobile phone is fast to do tasks and responds quickly.

Portability pertains to the manner by which the user is able to smoothly perform the same tasks from one device to another efficiently and effectively. [20] Mobile phone users expressed their satisfaction of their phones which allow them to continue to perform same tasks on similar devices; hence they are able to do integration with other devices. They appreciate that the mobile phones have the ability to automatically synchronize tasks and data, such as contacts and applications.

Performance refers to the responsiveness of the device to execute the tasks required and needed within an expected time. [21] Mobile phone users require that they can have a smooth shifting of task from one option to another. They mentioned that their mobile phones really do what they ought to do, run smoothly, have fluidity in response and their responses are reliable. They also appreciate that there are available icons to access and that the phones do what they are expected to do. They emphasized that their phones handle multi-task and do not disrupt their user experience. They also highlighted that the response is not affected even when charging the phone. They particularly cited the use of maps as being able to give them accurate data.

Functionality includes the properties that show how tasks perform the stated or implied needs. [22] Mobile phone users mentioned that the performance of their mobile phones is predicted, easy to figure out, can back-up all information that they have in their phones and have smooth performance. For instance in some mobile phones, users are able to know what happens when they pinch or slide on the interface. They also said that their phones are stable and have no problem in connecting with others and other devices.

Advanced features include all other functions that mobile users require from the mobile phones. These include auto-detection of movements, auto-adjustment in brightness, and the use of hand gestures for phone functions.

4.4 HCI Design Principles Applicable to Mobile Phones

Mobile Phone Characteristics–HCI Design Principles Map

In the section above, characteristics of mobile phones preferred by the users were explained. These characteristics served as the basis in identifying the HCI design principles applicable in mobile devices as experienced by the users. Table 1 shows the mapping of these mobile phone characteristics to the HCI Principles. These mobile characteristics surfaced in the commentaries of the mobile users extracted from on-line mobile surveys covering 2014, 2015 and early part of 2016.

Considering the given definitions and significance of both the HCI principles as well as the mobile phone characteristics, Table 1 showing the mapping is derived.

In order to evaluate the importance and prioritization of the HCI principles based on user experiences, and as a first step, a one-to-one mapping, where each HCI principle is mapped to the most relevant mobile characteristic is considered. Table 2 presents the considered mapping which was used in the statistical analysis.

Table 1 – Mapping of Mobile Phone Characteristics to the HCI Principles

Mobile Phone Characteristics	HCI Principles						
	FM	CN	SB	TM	SN	PR	PE
1. Interface Design	✓						✓
2. Efficiency		✓	✓				
3. Reliability		✓					
4. Presence of utilities			✓				✓
5. Responsiveness				✓			
6. Portability				✓			
7. Performance				✓	✓	✓	
8. Functionality						✓	
9. Advanced Features				✓			✓

** FM – Familiarity CN – Consistency SB – Substitutivity*
TM – Task Migratability SN – Synthesizability
PR – Predictability PE – Perceptual Ergonomics

As shown in Table 2, the characteristic Interface is mapped to HCI Principle – Familiarity, since familiarity pertains to the understanding of the interface. All responses and commentaries reflecting various instances and experiences dealing with ease of access and in dealing with interfaces were included under this mapping.

The characteristics – efficiency and reliability are mapped to the HCI Principle – Consistency. As this principle speaks about how a user can uniformly behave and perform tasks in similar scenarios, all responses and commentaries dealing with how the mobile phones respond consistently with the tasks given by the users, from simple mails and messages to making calls were included.

The characteristic – presence of utilities was mapped to the HCI Principle – Substitutivity as the principle speaks about availability of options to execute a task. All user responses and commentaries that mentioned about provision of alternative options, such as buttons and short-cuts to perform some commands were included in this mapping.

The characteristics – responsiveness and portability are mapped with the HCI Principle – Task Migratability. Since this principle covers aspects that allow the system to do what the user requires, all responses and commentaries on how the mobile device functions the tasks the user expects it to do were included here.

The characteristic – performance is mapped with the HCI Principle – Synthesizability. Since the principle tackles what happens in interfacing with the system, all user responses and commentaries regarding the ways by which their mobile phones respond to whatever tasks given were mapped here.

The characteristic – functionality is mapped with the HCI principle – predictability. This principle states that the user is able to determine what happens when a task is executed; hence all user responses and commentaries expressing the user's requirements, needs or observations on how the mobile devices are able to operate based on some specific criteria were included here.

The characteristic – advanced features is mapped with the HCI principle - perceptual ergonomics since all those responses and commentaries mentioned have something to do with detecting human perception and sensory.

There is no mobile characteristic required by the user that can be mapped to the principle – recoverability. However, this does not mean that the principle recoverability is no longer applied in mobile devices. It should be noted that the mobile phone developers in their Design Guidelines reflect this principle by giving the user chance to reverse the actions. This may be true for the applications, but not in the mobile phone itself as the users are not able to retrieve or recover loss of data as soon as a mistake has been made. [23] An instance of this is the presence of the icloud storage in some mobile devices, where data, applications, phone settings, homescreen and many more can be recovered [24].

Table 2 – Mapping of Mobile Phone Characteristics to the HCI Principles

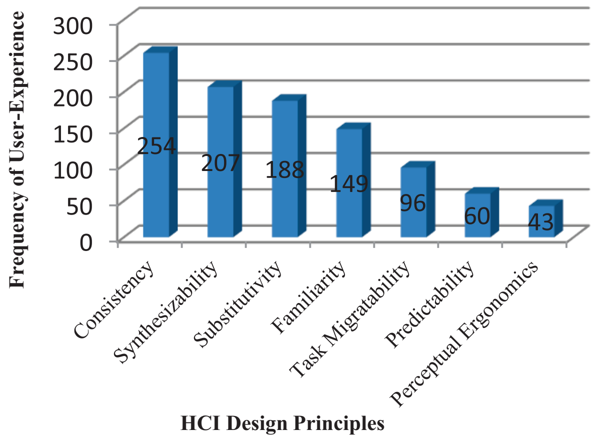
Mobile Phone Characteristics	HCI Principles
1. Interface Design	Familiarity
2. Efficiency	Consistency
3. Reliability	Consistency
4. Presence of utilities	Substitutivity
5. Responsiveness	Task Migratability
6. Portability	Task Migratability
7. Performance	Synthesizability
8. Functionality	Predictability
9. Advanced Features	Perceptual ergonomics

Applicability of HCI Design Principles in Mobile Devices

As presented in the discussion above, there are 7 HCI design principles that were mapped to the mobile characteristics preferred by the user. These principles though do not have the same level of importance or priority when it comes to its applicability in mobile devices. Considering the commentaries of 297 mobile users, which were tallied against the HCI Design Principles, their perception of the HCI design principles applied in mobile devices, was determined. As shown in figure 1, the 7 HCI design principles have varying values, and the principle having the most value is consistency (254/297) and the least is perceptual ergonomics (43/297). The data indicate that users rank consistency as the HCI design principle that should be applied first in mobile devices. This is followed by Synthesizability (207/297), Substitutivity (188/297) and Familiarity (149/297).

The HCI Design Principle – Consistency being the principle which is evidently applied in mobile devices as perceived by the respondents can be attributed to the fact that mobile developers see to it that there is consistency in the design of their products. They ensure that in any interface, the mobile phone user does not have to learn new technique for every application or interface or neither has to be confused in using many icons for the same purpose. [23]

Figure 1: Applicability of HCI Design Principles in Mobile Devices

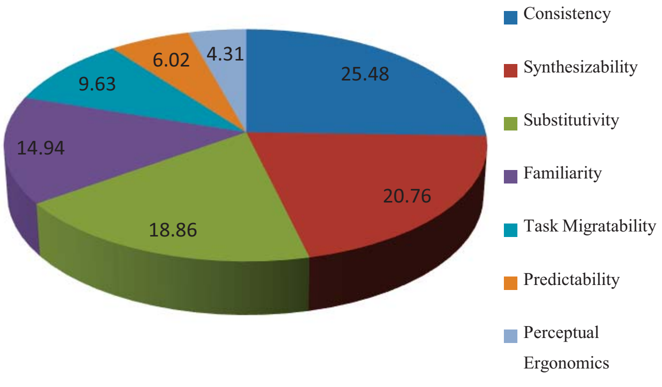


A better analysis can be made by showing how one (1) principle relate to the other six (6) principles., and this can be shown in Figure 2. The figure shows that there is 25.48% user-experience on the principle – consistency, 20.76% on the principle – Synthesizability and 18.86% on the principle Substitutivity. This implies that of the seven (7) principles, users seek for the presence of the principles – consistency, synthesizability and substituvity in mobile devices. These principles are more evident in mobile devices as they make up to 65% of the total user-experience.

4.5 View of Mobile Phone Developers on HCI Principles

Mobile phone developers have their own set of design principles applying the principles for human-computer interaction in coming up with a product that is elegant, efficient, intuitive, and delightful. The HCI principles applied by the mobile developers can be evidenced in the styles of the products themselves. For instance, the HCI principle – Consistency is explained on how the user interacts with the interface. It does not mean that the interface must look the same or must behave the same in all applications, but rather, consistency is seen in the manner of doing things or in behaving with different applications. It means that the user does not have to be forced to learn new techniques in interacting with the application. The concept of consistency can be seen on the labels and features of the interfaces. The same icons behave similarly or perform the same tasks regardless of the interface. The "x" character at the top of every window consistently means cancel.

Figure 2: User Perception on the Applicability of HCI Design Principles



Moreover, the principle HCI – predictability can be manifested on how the user manipulates the device. Mobile phone developers identify actions on the device by explicit and implicit, where they consider the result of the action to an object (explicit) and the result of the action through images or visuals as signs or indicators (implied). For instance, dragging and dropping a set of objects can result to many different things, however nothing is explicitly stated; and with this users rely on the graphical prompts or clues such as change on how the pointer appears to be able to determine the result of the action.

The HCI principle – Recoverability is manifested in situations where the user is able to reverse the action. Mobile developers provide ways to let the users feel that they can try to do things or explore on the device without a losing their data. Some options that mobile developers include in the design is the availability of prompts to cancel or undo a certain task executed. For instance, in editing an image in a certain phone, the user can modify the image as much as he wants without changing the image file until such time that he finalizes or saves the modification. He can leave the changes to the image by the option cancel.

It is also worthy to note that mobile phone developers use the user-focused approached in designing their product. They consider their target audience and use a specific technique to ensure that they deliver a product that meets user requirements.

5. Conclusions and Recommendations

Mobile phone users have a specific set of characteristics they prefer to have from mobile phones and consider consistency as the primary characteristic. The design principles of mobile developers adhere to the HCI principles. However, mobile phone users focus on few of those HCI principles, namely: consistency, synthesizability and substitutivity. The design of the mobile device is a combination of the user-experience and HCI Principles.

Based on the findings and conclusions of the study, it is recommended that a study to explore the prioritization of the HCI principles as implemented by the mobile phone developers be conducted to specifically determine the effect of a specific design principle for a certain product. This will pave way for the differentiation of design principles applicable to hardware or software.

Mobile developers should make the HCI principle – recoverability, be more evident in their devices as this principle is applied primary to software and secondary to hardware, which should not be the case.

Similar study should be conducted for mobile applications, such as games and customized apps to find out whether or not the applications available in the store match the HCI design principles. Similar study can also be done to evaluate websites and student projects and come up with a set of metrics to ensure the appropriateness of the design.

6. References

1. The World in 2011: ICT Facts and Figures. [Online] http://www.itu.int/ITU-/ict/facts/2011/material/ICTFacts Figure2011.pdf.
2. Mobile Subscribers Worldwide. [Online] [Cited: December 16, 2012.] http://mobithinki ng.com/mobile-marketing tools/latest-mobile-stats/a#subscribers.
3. Mobile Phone Users Worldwide. [Online] http://www.emarketer.com/Article/SmartphoneUsersWorldwideWill-Total-175-Billion-2014/1010536
4. Mobile Economy 2015. [Online] http://gsmamobile economy.com/GSMA_Global_Mobile_Economy_Report _2015.pdf
5. Improvements in Cellphone Technology. [Online] [Cited: December 16, 2012.] http://www.eHow.com/about_5542192_improvements-cell-phone-technology.html.
6. Mobile Trends. [Online] [Cited: December 16, 2012.] http://ezinearticles.com/?Follow-the-Mobile-Trends-With-Mobile-Application-Development&id=7421722.
7. Mobile Interaction. [Online] [Cited: December 16, 2012.] http://en.wikipedia.org/wiki/Moile_Interaction.
8. Human Computer Interaction. [Online] [Cited: December 16, 2012.] http://en.wikipedia.org/wiki/Human-Computer_Interaction.
9. User Experience Design. [Online] [Cited: December 16, 2012.] http://www.montparnas.com/articles/what-is-user-experience-design/comment-page-10.
10. Pearson Student Mobile Device Survey 2014. May 9, 2014. http://www.pearsoned.com/wp-content/uploads/ Pearson-K12-Student-Mobile-Device-Survey-050914-PUBLIC-Report.pdf
11. Review and Analysis of Human Computer Interaction (HCI) Design Principles. Hinze-Hoare, V. s.l.: Cornell University Library, 2007. http://arxiv.org/ftp/arxiv/papers /0707/0707.3638.pdf.
12. Challenges in Human-Computer Interaction Design for Mobile Devices. Huang, Kuo-Ying. San Francisco, USA: WCECS, 2009. Proceedings of the World Congress on Engineering and Computer Science.
13. Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. Hoggan, Eva, Brewster, Stephen A. and Johnston, Jody. Florence, Italy: ACM, 2008. 978-1-60558-011-1/08/04.
14. Feel-Good Touch: Finding the Most Pleasant Tactile Feedback for a Mobile Touch Screen Button. Oskinen, Emilia K., Kaaresoja, Topi and Laitinen, Pauli. Chania, Crete, Greece: ACM, 2008. 978-1-60558-198-9/08/10.
15. Personalization and Visualization on Handheld Devices. Zhang, Dongsong, et al., et al. Dijon, France: ACM, 2006. 1-59593-108-2/06/0004.
16. Visualizing Locations of Off-Screen Objeccts on Mobile Devices: A Comparative Evaluation of Three Approaches. Burigat, Stefano, Chittaro, Luca and Gabrielli, Silvia. Helsinki, Finland: ACM, 2006. 1-59593-390-5/06/0009.
17. Advanced Power and Thermal Management for Low-Power, High-Performance Smartphones. Jung, HWisung. California, USA: ACM, 2012. 978-1-4503-1249-3/12/07.
18. Barrel Menu: A New Mobile Phone Menu for Feature Rich Devices. Foster, Greg and Foxcroft, Terence. Cape Town, South Africa: ACM, 2011. 978-1-4503-0878-6/11/10.
19. Mobile Web Browsing: Usability Study. Shrestha, Sujan. Singapore: ACM, 2007. 978-1-59593-819-0.
20. Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models. ISO/IEC 25010:2011. https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en
21. Quality Attributes. [Online] https://msdn.microsoft.com/en-us/library/ee658094.aspx# Performance
22. Quality Model. [Online] http://en.wikipedia.org/wiki/ ISO/IEC_9126
23. Design Principles. [Online] https://digitalcooings.wordpress.com/ 2010/04/13/apples-hci-product-design-secrets/
24. iCloud Storage and Backup Overview. [Online] https://support.apple.com/kb/PH12519?locale=en_US&viewlocale=en_US

Dividing Transmission Method for Multimedia Service Using OpenStack

Sanghyun Park[1], **Linh Van Ma**[2], **Jinsul Kim**[3]
School of Electronics and Computer Engineering, Chonnam National University, 77
Yongbong-ro, Buk-gu, Gwangju 500-757, Republic of Korea
sanghyun079@gmail.com[1], linh.mavan@gmail.com[2], jsworld@jnu.ac.kr[3]

Abstract - *This paper proposes a dividing transmission method for providing an optimal delivery multimedia content using OpenStack. Multimedia data are installed in several servers, these servers has a Management Server which provides an optimal method when user requires content. In addition, we develop a partition algorithm so that users can receive multimedia data from these servers in an appropriated way through Management Server. In this paper, a fast algorithm for getting data in the multimedia server have developed based on the transmission of each server.*

Keywords: OpenStack, Division Transmission, Optima Multimedia Transmission, Management Server

1 Introduction

Various techniques have been researched in order to provide multimedia services globally. In addition, many agencies and suppliers are developing their business along with cloud services so users can require multimedia services at any time over the Internet [1-2]. However, building a real-time multimedia system service is difficult. In addition, to serve large amount request from users, we need manage the system efficiently. If the user requests data from a multimedia server that is far away or it is overloaded, users cannot have a seamless service. Therefore, in this paper, we propose an effective management resources method and a dividing algorithm that provides an optimal multimedia service in OpenStack [3]. Dividing transmission algorithm is a method that transmit data rapidly from a nearest server that contains data to a user [4]. If the server has no desired multimedia contents, it dedicates other servers that provide the requesting data even it has fast connection. In general, the transmitting multimedia data from the server that holds the content cannot provide an optimum if the server is far away from a user. Therefore, this paper uses the optimal selection server method user-oriented-based for the transmission of multimedia content that uses the transport techniques OpenStack. We describe the algorithm for split delivery and test the proposed system as well.

2 Related works

Many studies have been conducted with multimedia services. Among them, the multimedia transmission and related articles "Caching and optimized request routing in cloud-based content delivery systems" paper [5]. It limits storage costs and physical resources with a variety of uploading content and provides user-requested content quickly. It also described a method of how to reduce the delay time. The proposed method in this scheme serves the content using dynamic caching, the Elastic storage resources that are less costly than the existing algorithm. The "Volly Automated Data Placement for Get-Distributed Cloud Services" paper [6], describes an increasing in the size of the distributed data center services, data distribution considerate the costs and limitations of the WAN bandwidth with data center capacity. The proposed method has been described that an optimized algorithm development and connection request access patterns according to the data center and the log acquisition of the client location. In addition, this paper analyzes the log data acquisition using its proposed algorithm, then the data analysis was applied to a cloud service. Figure 1 shows a flowchart of data analysis in the paper.

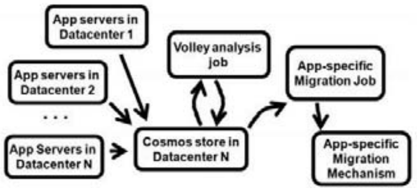

App servers in Datacenter 1
App servers in Datacenter 2
. . .
App Servers in Datacenter N
Volley analysis job
Cosmos store in Datacenter N
App-specific Migration Job
App-specific Migration Mechanism

Fig. 1. Data flowchart analysis of Volly automated data placement

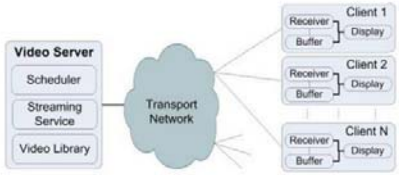


Fig. 2. Structure of video delivery network

"In the paper "Improving Scalability of VoD Systems by Optimal Exploitation of Storage and Multicast" stated that VoD system users and an amount of data increases continually in accordance with the scalability and bandwidth VoD services. Therefore, in order to solve the above problem, Prepopulation assisted batching with multicast patching (PAB-MP) developed for End User Devices to read an initial segment of the pre-reading and video multicast server with fast and simple structure. Figure 2 shows the structure of the network proposed in the paper. Multimedia transmission method of the above paper is that no matter how good they are based on the server, the server does not provide built even useless to users. In this paper, we developed an algorithm that provides multimedia content then selected the optimum server to incorporate each one in the center server for the user and proceed an experiment.

3 Operating structure of whole system

3.1 Scenario of the whole system

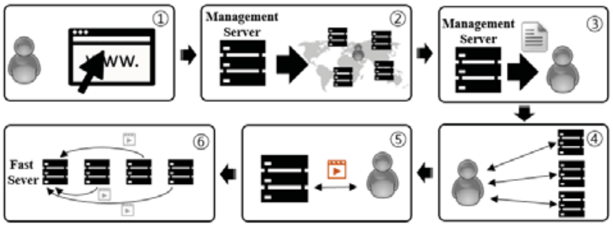


Fig. 3. Proposal system sequences

The overall structure of the system is shown in Figure 3. Fig 3-① user first uses the Web and requests the server for selecting the desired content. Fig 3-② OpenStack-based Management Server retrieves the servers which near the user's location. Fig 3-③, the Management Server sent the user the list of the adjacent VM Data Servers. Fig 3-④, users request the respective data by using the list of VM DataServer which received from the Management Server, then the servers are requested to check a transmission rate of the user. Fig 3-⑤ a user check the transfer rate of the server, and then select the server volume. And it receives the multimedia data transmission request from the selected server. Fig 3-⑥ If the server does not provide optimum content transmission which requested by the user in the VM DataServer, user requests a file from the other server. At this time, the user transmits data that divided proportionally according to the data rate at which the file is requested to provide a respective server in multiple servers. When the split delivery to the user, the best portions of the fast server sends to the user multimedia content, it also receives the content from some other server. By this way, the user requests the content easily using the Web and the internal server system provides a good quality content to the user. We use a server-centric rather than user-centric data transfer with data transmission faster than any existing data server. In general, the service rate of the server cannot provide a seamless multimedia content due to the switch, router, and relay device that connects them from the remote server. Therefore, if user's surrounding servers provide a data center to the user may send faster and better quality multimedia data than the existing data servers.

3.2 Division transmission method of real-time multimedia

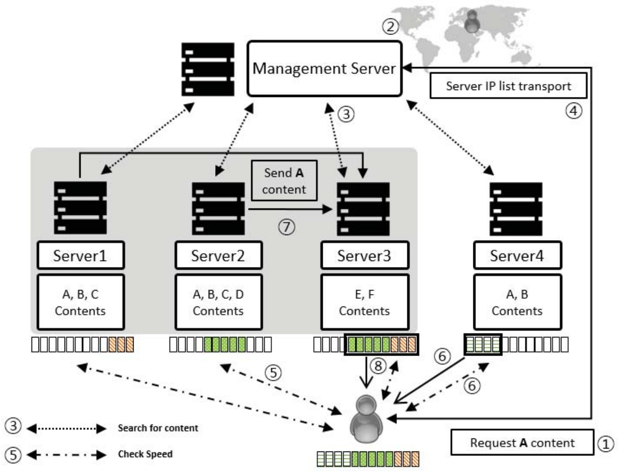


Fig. 4. Data division for optimal multimedia scenario transmission

Figure 4 shows the data division transfer scenarios for optimum multimedia transmission. Fig 4-① users use a variety of devices with Web access and select multimedia content. Fig 4-②, ③ After analyzing the location of the user device by using the IP, the server searches for the nearby multimedia content providing servers. Then, it checks the status of the multimedia content that requested by the user with the retrieved destination server. Fig 4-④ Management Server transfers the list of discovered servers to users. Fig 4-⑤ The user checks the response speed of the respective server, using the list received from Management Server. Fig 4-⑥ user checks the response rate of the data piece on server in order to request the content from the second fastest if there is no content in the second fastest Server4. The second fastest Server4 sent real-time multimedia content to users earlier than the rest of the content that requested by the user from the other, such as Fig 4-⑦ are partially transfer the content from the fastest Server3. At this point, the partial rate based on the amount of transmitted data servers that typically divides the multimedia data. In this way, it can transmit multimedia data of high resolution and high quality. Finally, in Fig 4-⑧ if the content transfer is completed with fastest connections and terminate at the server (Server 3) then data is merged from data piece in the second fastest server (Server4) and the fastest server (Server 3). This is a user-oriented optimum method for transferring data in between servers and user, the user finally has a high resolution and high-quality transmission of multimedia data.

3.3 Dividing transmission algorithm for optimal transmission of multimedia content

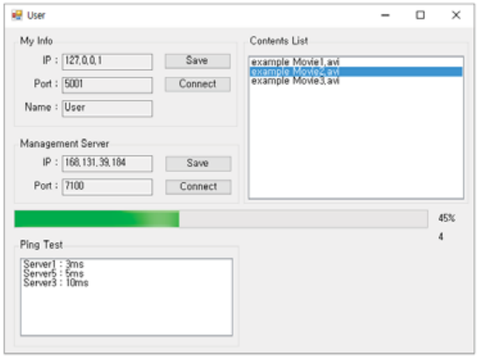
Algorithm 1 is divided transmission algorithm for optimized transmission of multimedia content. $S_c(i)$ refers to a server that has the content requested by the user. After the server is ready to transmit data, the server finds in Management Server, the fastest server in the list Sort(0) which will transmit data from the fastest optimum server $S_c(0)$. The server will provide the requested content to the user, if the server has different content then a next fastest server is used to provide data to users. The remaining servers are the servers that provide the optimal rate content on the size of the user-requested content, which is transmitted to the user proportionally transmitted to the server list Sort(0). At this point, the expression (1) means that the entire file, and divided according to the percentage of file size S_f. $S_f(0)$is the first front part of the multimedia data to be transferred. $S_b(i+n)$ is sent directly to the user, except for servers that belong to the servers $S_c(i)$ based on the speed ratio S_b. The transferring data in server $S_c(0)$ ends or the best Sort(0) server transfer data is completed, then the user receives the provided requested-content. Division transmission algorithm is shown in the Fig. 4-④ ~ ⑥.

Algorithm 1. Division transmission algorithms for multimedia transferring content

```
WHEN $\boldsymbol{S_c(i)}$ Server ready for data transmission THEN
  IF $\boldsymbol{S_c(0)}$ ! = Sort(0) THEN
  // If you do not have the fastest server Content
     $\boldsymbol{S_c(0).send}()$;
     // Earlier data transmission of the multimedia data
    WHILE $\boldsymbol{S_c(i+1)}$ != NULL THEN
    // Data partitioning according to server ratio
      $\boldsymbol{S_f(i+1) = (1 - S_f(0))\frac{s_b(i+n)}{s_b(i+1)+ \ldots +s_b(i+n)}}$     (1)
      $\boldsymbol{S_c(i+1).send}()$;
      // Data transmission according to the ratio
    END WHILE
  END IF
  IF $\boldsymbol{S_c(0)}$ == Sort(0) THEN
  // If you have the fastest server content
    $\boldsymbol{S_c(0).send}()$;
  END IF
  IF (Complete $\boldsymbol{S_c(0)}$ data transfer) or (Complete Sort(0) data
transfer) THEN
    Sort(0).Send();
    // If you have the fastest server content
  END IF
END WHEN
```

4 Division transmission experiment and evaluation result of multimedia content

In this paper, the algorithm was developed by dividing the actual data transfer. It is possible to use the simulation tools for testing, but we would rather develop a test program to find out how the rate comes out in actual application test. Fig. 5 shows the user is receiving data from the test program. Data Management Server uses a Linux-based OpenStack environment, the user client program test was developed using a Windows-based environment of the C# language. When a user connects to the server as shown in the figure, it shows a content list from the Management Server. A user chooses content by clicking on the content list which receives from the server that holds on the Management Server. The testing programs divided multimedia data transmission proportionally.



IP : 127,0,0,1
Port : 5001
Name : User

example Movie1,avi
example Movie2,avi
example Movie3,avi

Management Server
IP : 168,131,39,184
Port : 7100

45%
4

Server1 : 3ms
Server5 : 5ms
Server3 : 10ms

Fig. 5. Test program for division transmission speed

In order to receive offers from servers, we must measure the speed of servers. In this paper, we use ping testing method. First, the test-program sorts servers in the order which is proportional to the speed and transmits the data according to the split command. Secondly, the user requests the multimedia data from the data server. Third, multimedia data is sent to the user after the playback, it sends all the rest of the data files. Finally, the combined data is provided to the user. As such, the user receives the transferred data is partitioned on the basis of the fastest server in order, which is shown in sort expressions (2) with share proportionally data.

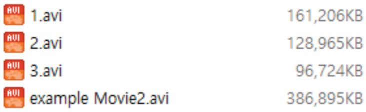
$$\frac{S(i)_{speed}}{Sum_{S\_speed}} F_{size} = S(i)_{F\_size} \qquad (2)$$

$S(i)_{Speed}$ refers to the speed of each of the data server, $Sum_{S_speed}$is a server that provides data to users throughout the speed, $F_{size}$is the size of the data requested by the user, $S(i)_{F\_size}$ refers to the amount of data to receive data transmitted from the server. For example, if the total file size of 360MB in server 1 then the partition is allocated about 130MB. If we divide the capacity of data four times in the server, we get 32MB data allocation. If the data is allocated to the data server informs the start and end points. The data server uses the start and end points if the pieces of data are sent to the user. The first part of data is sent from the fastest server to user meanwhile user also receives other pieces of data from the other servers.

Table 1. Experiments environment network condition

	Server1	Server2	Server3	Server4	Server5
Providing speed	6Mbps	3Mbps	2Mbps	1.5Mbps	4Mbps

Fig. 6. Partitioning multimedia data



1.avi	161,206KB
2.avi	128,965KB
3.avi	96,724KB
example Movie2.avi	386,895KB

Fig. 7. Data paritioning and its combination

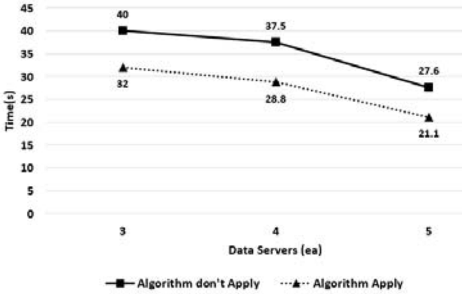


Fig. 8. Division transmission test result of multimedia data

Fig. 6 shows a picture of reproducing the first data piece. It is possible to play the video partially as shown in the rectangle because it transmits the data into small pieces. Then, when part of the data transfer is complete, the data is combined into a single file can reproduce the entire multimedia file. Figure 7 shows the three received data pieces which deliver from multimedia data servers with 1-3.avi, and the combined multimedia data file is Movie2.avi. Fig. 8 shows the results of the speed test multimedia data transferred with the graph partitioning. The vertical axis represents the amount of time (in seconds), the horizontal axis represents the number of the server. Because the data is partitioned with rate 1:1 as shown in the image which makes the transmission typically faster, when we apply the algorithm to split data into the proportional piece, then we achieve transfer data faster than keeping the original file. The above program uses an algorithm in the real to measure whether the tests and data transfer, we might have some errors depending on the speed and the surrounding environment. Therefore, we use the extension number of the same data and use the server to get the accurate measurement under the same conditions, and measure the transmission rate.

5 Conclusions

We proposed a method for the user to find the optimal server center dividing multimedia content delivery in order to provide optimal multimedia services. In the experiment in this paper, we divide the transmission data based on a rate from Ping Test, The division number of the receiving server through the experimental results confirm that the speed of the content to be provided is reduced. In order to provide a service we use OpenStack, OpenStack only uses for server environments, In order to evaluate the proposed method, we developed the program on C#-based. In the future, we will continue the research for delivery data from dynamic VM environment resources efficiently.

6 Acknowledgements

This work (Grants No.C0218369) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2014, and also supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2013R1A1A2013740)

7 References

[1] Jieyao Liu, Ejaz Ahmed, Muhammad Shiraz, Abdullah Gani, Rajkumar Buyya, and Ahsan Qureshi, "Application partitioning algorithms in mobile cloud computing: Taxonomy, review and future directions," Journal of Network and Computer Applications, Vol.48., pp.99-117, 2015.

[2] Nan, Xiaoming, Yifeng He, and Ling Guan, "Towards optimal resource allocation for differentiated multimedia services in cloud computing environment," Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, pp.684-688, 2014.

[3] Yanagawa, Toshihide, "OpenStack-based Next-generation Cloud Resource Management," Fujitsu Sci. Tech. J, Vol.51 No.2., pp.62-65, 2015.

[4] Ubarhande, Vrushali, Alina-Madalina Popescu, and Horacio Gonzalez-Velez, "Novel Data-Distribution Technique for Hadoop in Heterogeneous Cloud Environments," Complex, Intelligent, and Software Intensive Systems (CISIS), 2015 Ninth International Conference on. IEEE, pp.217-224, 2015.

[5] Niklas Carlssona, Derek Eager, Ajay Gopinathana and Zongpeng Li, "Caching and optimized request routing in cloud-based content delivery systems," Performance Evaluation, Vol.79., pp.38-55, 2014.

[6] Sharad Agarwal, John Dunagan, Navendu Jain, Stefan Saroiu and Alec Wolman, "Volly: Automated Data Placement for Get-Distributed Cloud Services," Networked Systems Design and Implementation, pp.17-32, 2010.

Crowd-Sourced, Cloud-Based Applications Support Real-Time, Decentralized, Ad-hoc, Emergency Management Services at the Individual Level: A Case Study

J. McKinney Young[1], L. Etzkorn[2]
[1,2]Department of Computer Science, University of Alabama in Huntsville, Huntsville, Alabama, USA
300 Technology Hall
University of Alabama at Huntsville
Huntsville, AL 35899
[1]julienmckinneyyoung@mac.com, [2]EtzkorL@uah.edu

Abstract - This case study examines how data from cloud-provided social media and cell phone service was interleaved with real-time Virginia Department of Transportation road closure data to help individuals collaborate to pickup children and a stranded motorist during an ice emergency. This was done using heterogeneous technology and apps connected to the cloud. This study also highlights the need for intensified research focused on integrating the smart phone based, cloud-aware apps data into vehicles "hands-free" sound systems and visual displays. The case study shows existing capabilities of communication between people and machines using mobile devices and wireless access to cloud-based, crowd-sourced real time data has changed how humans respond to emergencies. The events in this case study also show that wireless vehicle-to-mobile device integration, however, lags significantly compared to wireless device-to-cloud integration.

Keywords: Mobile Cloud, Social Media, Emergency Management

1. Introduction

On Jan 20, 2016, black ice on most roads in Northern Virginia brought evening rush hour traffic to a standstill. Accidents, stalled cars, and closed streets caused numerous people to be stranded in their cars. Many motorists abandoned their cars stuck in the stalled traffic if they were close enough to walk home through the snow and ice. Even as late as 1:30 the next morning, motorists were stranded on frozen streets. This case study first examines how data from cloud-provided social media and cell phone service was interleaved with real-time Virginia Department of Transportation (VDOT) road closure data to help individuals collaborate to pickup children and a stranded motorist during an ice emergency. Heterogeneous technology and apps connected to the cloud supported ad hoc, real-time collaboration to provide emergency services at the individual level. Secondly, this case study highlights the lack of integration between the smart phone based, cloud-aware apps into vehicles "hands-free" sound systems and visual displays. The existing capabilities of communication between people and machines using mobile devices and wireless access to cloud-based, crowd-sourced real time data has changed how humans respond to emergencies. Wireless vehicle-to-mobile device integration, however, lags significantly compared to wireless device-to-cloud integration.

2. Related Work

Emergency management - helping other humans in times of disaster - is an ongoing effort in research. Geumpana, et al. study emergency management from the perspective of the technology that supports the human endeavor. Communication - a key feature to emergency management [3] - through global connectivity is increasing through access to the internet, social media and particularly mobile phones [3]. The authors reviewed 53 papers to develop a taxonomy of issues for implementing mobile cloud computing (MCC) for emergency management. The paper partitions the group of MCC issues into subgroups that include connectivity issues, communication technology differences and context awareness, among others [3]. The combination of reliable support from wireless network technologies and backend cloud services plus mobile smart devices form the critical parts in delivering support during emergency situations [3].

Sanaei, et al, define MCC as

> *...a rich mobile computing technology that leverages unified elastic resource of varied clouds and network technologies toward unrestricted functionality, storage, and mobility to serve a multitude of mobile devices anywhere, anytime through the channel of Ethernet or Internet regardless of heterogeneous environments and platforms based on the pay-as-you-use principle* [9].

The challenge of providing and managing MCC is detailed in [1, 2, 9, 10]. The chief issues as they relate to this paper are:

- the threat of intermittent connectivity [10],
- the lack of hardware and software integration due heterogeneous technology [9],
- network latency issues due to geographic location of real-time location data and services and the remote geographic location of the requesting user [2, 4, 6].

Shi, et al, describe the challenges of providing data and computation services to a mobile device in [10]. The authors write that the process of offloading computation from a device to the cloud is in wide use. The activity suffers from several shortcomings: the inflexibility of dividing work between the mobile device and the remote cloud, the latency inherent in accessing remote cloud services and the intermittent connectivity feature of communication due to device mobility [10]. Shi, et al, terms this type of spotty connectivity area a “cirrus” cloud [10] - using the name of thin, wispy cloud formations that stretch across the sky with intermittent open spaces of sky showing. The open sky represents lack of connectivity to the internet cloud.

Sanaei, et al, write that mobile cloud computing is a convergent technology made up of three fundamental technologies: mobile computing, cloud computing and networking [9]. Mobile device heterogeneity is defined as differing platforms, differing operating systems and different wireless network carriers [9].

The third challenge of mobile cloud computing that is relevant to this paper is location data - both the geographic location and the cloud location of the users [2] and the location specific data of interest [4]. Shi, et al, describes the issues of user/mobile device location and mobility as a function of interacting with cloud resources. Chang, et al, describes how the location of the mobile device within the cloud topography can impact the delivery of data and services in [2]. The idea of emerging cloud services follows the trend to extend the cloud to the edge of the network [2]. Stojmenovic calls this network edge area of services the “fog” in [14].

Jeong, et al, discuss user location data - both the actual physical location of the user and the location data that the user uploads to and downloads from the cloud in [6]. Cellular networks provide mobile devices like smart phones mobile wireless network access to cloud services [6]. Smartphone navigation apps like Waze provide crowd-sourced real-time traffic condition information to cloud services [6]. Google cars roam cities and map road topology [4]. The topology information is combined with actual photographs of buildings [4].

Guerla, et al, discuss how users tend to request data content by geographic and time relevance [4]. The authors propose the concept of the Time-Space Validity of data. An example would be users requesting current road condition data - the condition data of the road two hours ago or the condition of a road 500 miles away is of no interest to the driver who is currently commuting ten miles from work to home. The authors discuss the idea of a vehicle cloud made up of networked vehicles sharing sensor data [4]. In this case study, however, the vehicle was only a mobile platform for the human carrying the smart phone. The communication network was the cell-phone system to internet cloud. The human network of actors was comprised of a social network that existed as a micro-community of Facebook and real-life friends.

3. Background

There are several features of the Northern Virginia area that support ease of voice and data communication for mobile devices. The geography is not very mountainous. The major telephone service providers offer high-speed 4G LTE network coverage throughout the region. The data that the users in this event uploaded and downloaded could be accessed from data centers placed geographically near to the users.

The social media service provider used in this case study, Facebook, maintains an datacenter in Ashburn, Northern Virginia [17]. The real-time road condition map service provider, Google, has a data center in Reston, Virginia [19]. The Google Maps app receives traffic and incident data such as accident reports from the navigation app Waze (which Google bought in 2013) [7]. Waze uses crowd-sourced data posted by its users [7]. Google also posts information from local departments of transportation such as road closings or scheduled road construction work [12]. Google real-time traffic flow data is crowd-sourced by Android phones that have location services activated as well as iPhones that have Google Maps open [12]. These apps send anonymous data back to Google.

The individuals involved in the events of the evening did not experience issues with two of the three main challenges of MCC listed in the preceding section; intermittent connectivity and location-based latency. The emergency was caused by icy roads - not by a failure in the communications network. Wireless network access was consistent and access across the internet was uninterrupted. Plus, the apps and data that the group of friends used were resident at two major data centers placed geographically only a few miles from them. The humans and their cell phones were placed figuratively & geographically near the center of cloud resources rather than at the ”edge” of the cloud. Data uploading and downloading speeds were not affected by geographical “remoteness”. Technology heterogeneity did not impact the communication ability of the group of friends to provide help to those friends stranded in the ice storm via wireless smart devices connecting to the cloud. Technology heterogeneity did impact viewing current road conditions and possible routes through the affected area. It excluded any “hands free” viewing capability of Google Maps data within embedded vehicle display hardware. Drivers did not use embedded vehicle static map data at all. They needed real-time road open status from VDOT and up to the minute traffic condition information crowd-sourced from other drivers.

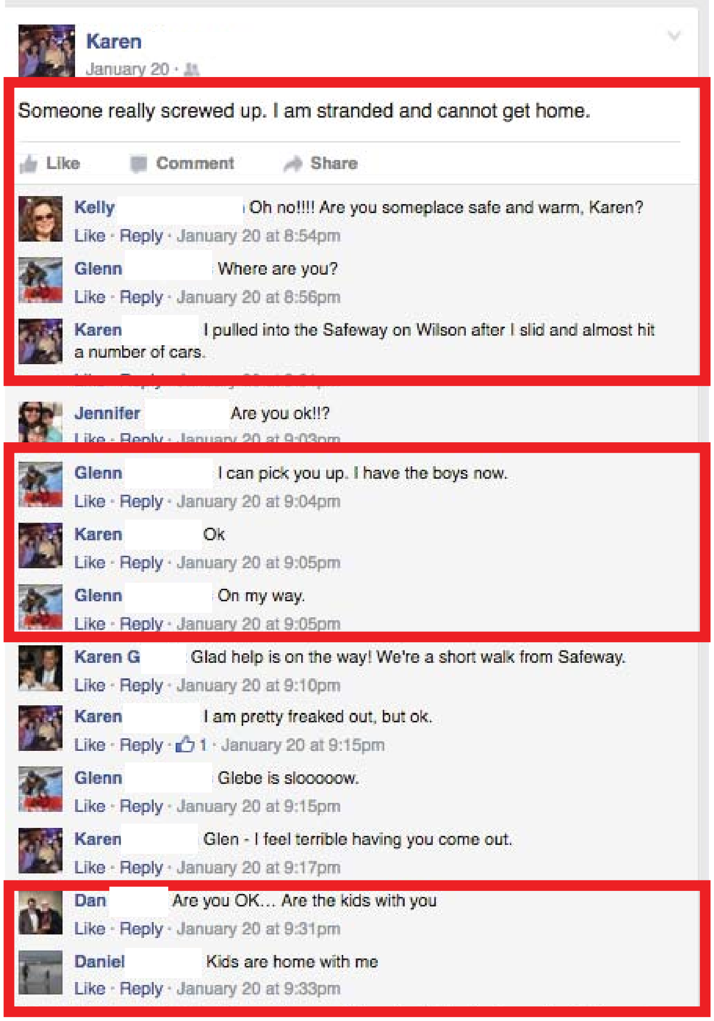
The case study discussed in this paper occurred on January 20, 2016, in Northern Virginia, just before the big East Coast blizzard “Jonas” immobolized the Mid-Atlantic and Northeast United States [11]. Two days before Jonas arrived, a small cold weather system pushed through the Northern Virginia area causing “massive traffic delays” [20]. Although only about one inch of snow fell, most roads were untreated. Any snow that melted quickly refroze into “black ice” - invisible to motorists - and immediately rendered roads inpassable [21]. Cars were stuck, bumper to bumper, during the DC area rush-hour on icy roads. Fire and Rescue services were inundated with calls. Where their normal velocity of calls is about 325 calls per 24-hour period (or one call every 4-5 minutes), the department reported handling about one call per minute - a four-fold increase. [20]. There were still gridlock issues including traffic stoppages as late as 1:30 am the next morning. Virginia State Police reported responding to 767 accidents and 392 reports of disabled vehichles [21]. This case study highlights one call that was NOT made to local authorities asking for help. Help came from a cloud-connected friend commuity communicating wirelessly in real time via smart phones.

4. Case Study

There are four people in cars (Paresh, Glenn, Janet and Karen) and two children (Kian and Luke) that the case study follows by name. In the beginning of the evening, one parent (Paresh) is driving to the sports center to pick up two children (Kian & Luke.) Her car is in a traffic accident because of the icy roads. The next plan is for Glenn to drive to the sports center and pick up the two boys. While Glenn is en route, the plan changes. Due to weather conditions, the sports center will close before Glenn can get there. Another parent (Janet), who is currently at the sports center to pick up her son when the decision is made to close it immediately, uses her cell phone while inside the building to call the other families whose children are still waiting to be picked up. The decision is made for her to take the three remaining children in her ca.. She will drive the four children either to a house that is quite close to her own to drop off one and then on to her house where another parent (Glenn) will pick up both his son and a neighbor's son. At this point in the timeline (at 7:58 pm), Glenn specifically requests an address to be texted to him. From the smart phone's text message screen, he can touch the address and Google Maps will automatically show the location of the address on the Google Map screen. Real-time, crowd-sourced road condition data near the address as well as official VDOT road status information will also be automatically rendered to the screen. (Note that Glenn never uses the vehicle provided static map data during this event.) Glenn will then drive both boys back to their neighborhood and pick up a stranded neighbor (Karen) on the way. Everyone makes it safely back to their own home that night due to constant communication through the wireless cell phone network and internet cloud access.

Details of Mobile devices/OS/Wireless Carrier [11]:
Glenn: Samsung Galaxy S5 running Android on Verizon Wireless.
Janet: iPhone5s running iOS9 on AT&T
Karen: iPhone5s iOS9 on Verizon Wireless

Timeline of Events [11]:
6:53 pm: Simolunas leaves work in Rosslyn, Va. Gets stuck in traffic. Google maps show many roads colored red.
7:30 pm: Abandons/parks car near 1008 N. Arlington Mill Dr., Arlington VA 22205, and walk the rest of the way home.
7:41 pm: When Glenn arrives at home, his spouse tells that their neighbor, Pareesh, had a car accident on her way to pickup her son and Glenn's son (Kian & Luke). No information about her accident or her condition. Glenn takes second car (an SUV with All-Wheel-Drive) to go pick the boys up from sports center (S Four Mile Run Dr, Arlington, VA 22206).
7:53 pm: Glenn stuck in traffic. Google Maps shows most roads highlighted in red indicating standing traffic. Spouse calls on cell phone. Change in plan. The coaches can't leave until all the kids leave. Some parents are still a long way away, so one parent (Janet) coordinates to take the rest of the kids: her son, another boy, Kian & Luke so the staff can go home. New plan is for Glen to meet the other parent at her house and pick up Luke & Kian (Kian lives near by Glenn).
7:58 pm: Spouse texts Glenn other parent's address (S. Irving St. Arlington VA 22204) because it made it easier to enter it into Google Maps with one touch on the smart phone's touch screen (road is icy and full of sliding cars - need both hands on wheel). He then uses Google Maps to find an open road to Janet's house,
8:20 pm: Arrive at S. Irving St. No one's home. Janet is still dropping off a child near by.
8:35 pm: Glenn surfs on Facebook. News is posted that Pareesh is OK and that it was just a fender bender.
8:56 pm: Karen mentions she's stranded on Glenn's Facebook timeline. Glenn queries for her location. No response. Visit her timeline and ask again. She's at Safeway which is near route to Glenn's house.
9:04 pm: Children arrive. Tell Karen (via Facebook post) that car/ help is on the way. Use Google Maps to find a collection of roads without a VDOT "road closed" symbol on them that lead to Karen's location.
9:28 pm: Arrive Safeway parking lot and collect Karen (Wilson Blvd., Arlington VA 22204).
10:25 - 10:35 pm: Glenn makes it to neighborhood, drops off Karen at her house, Kian at his house & finally makes it home with own child. After a quick dinner, Glenn checks back in on Facebook to see if everyone is ok.



Karen
January 20 ·

Someone really screwed up. I am stranded and cannot get home.

Kelly [redacted] Oh no!!!! Are you someplace safe and warm, Karen?
Like · Reply · January 20 at 8:54pm

Glenn [redacted] Where are you?
Like · Reply · January 20 at 8:56pm

Karen [redacted] I pulled into the Safeway on Wilson after I slid and almost hit a number of cars.

Jennifer [redacted] Are you ok!!?
Like · Reply · January 20 at 9:03pm

Glenn [redacted] I can pick you up. I have the boys now.
Like · Reply · January 20 at 9:04pm

Karen [redacted] Ok
Like · Reply · January 20 at 9:05pm

Glenn [redacted] On my way.
Like · Reply · January 20 at 9:05pm

Karen G [redacted] Glad help is on the way! We're a short walk from Safeway.
Like · Reply · January 20 at 9:10pm

Karen [redacted] I am pretty freaked out, but ok.
Like · Reply · 👍 1 · January 20 at 9:15pm

Glenn [redacted] Glebe is slooooow.
Like · Reply · January 20 at 9:15pm

Karen [redacted] Glen - I feel terrible having you come out.
Like · Reply · January 20 at 9:17pm

Dan [redacted] Are you OK… Are the kids with you
Like · Reply · January 20 at 9:31pm

Daniel [redacted] Kids are home with me
Like · Reply · January 20 at 9:33pm

Figure 1: Facebook Group Conversation 1/20/2016, 8:54 pm - 9:33pm
(Last Names Hidden for Privacy)

The red highlighted comments in Figure 1 show the emergency management activity taking place by several people. In the first highlighted area, Karen alerts the friend community that her car (and, by association, herself) is stuck. Glenn asks her location and, in the second highlighted area, volunteers to pick her up. In the third highlighted area, another member of the friend community (Dan) inquires about her status and, using the extra information that is part of a social/friend network, asks about Karen's children. Karen's husband (Daniel) who is also part of this group conversation thread reassures all that their children are with him (implying that the children are safe).

Not shown in the Facebook conversation in Figure 1 is the number of cell phone calls and text messages made during the evening that re-planned and rescheduled tasks such as the collection of various children from cancelled after-school activities. Most of the communication that evening occurred via cell phones from inside a vehicle but some originated from cell phones and tablets inside buildings.

Figure 2 and Figure 3 are the Google maps that members of the friend community involved in the conversation shown in Figure 1 posted to the Facebook thread. The first two maps show most of the roads in the Northern Virginia area either closed (information originating from the Virginia Department of Transportation [12]) or with standing traffic (gathered by Google through real-time data collection from cell phones in the cars on the actual roads [12]).

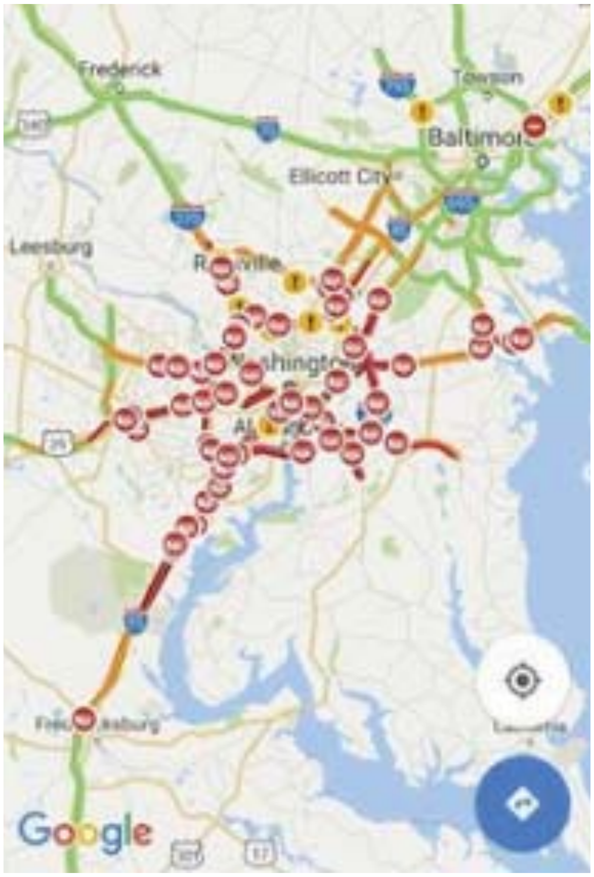


Figure 2. Google Map 1/20/2016 around 8:40pm.
⛔ indicate road closures [18].

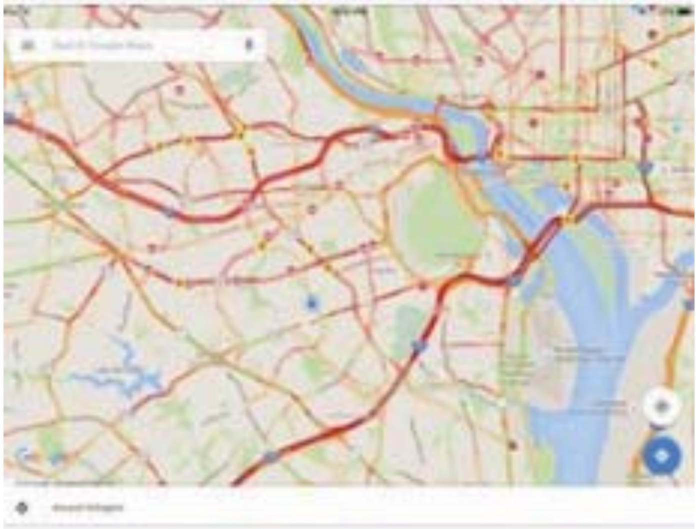
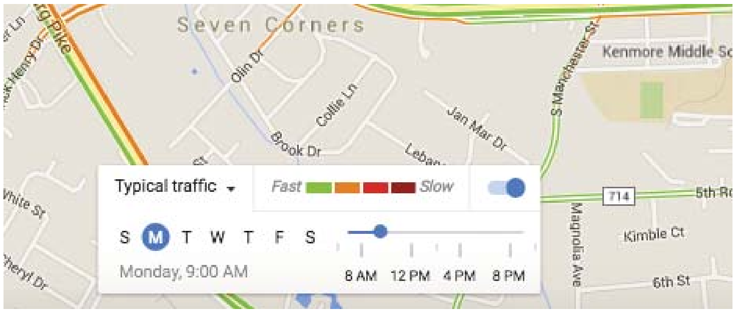


Figure 3. Google Map 1/20/2016 at 10:30pm

Figure 4 shows what a Google Map typically looks like in normal traffic conditions. Many roads are marked in green to show typical fast moving traffic flow and only one or two marked in orange indicating slight traffic slow-down. Compare Figure 4 with Figure 2. In Figure 3 almost all roads are either a dark red or a less intense red showing very slow traffic. In Figure 2, the view is "zoomed out" to show major arteries and the map is splattered with Road Closed symbols. These are the real-time updated maps that drivers studied (while standing in traffic) in order to better understand the conditions of nearby roads.



Typical traffic ▾ Fast 🟩🟧🟥🟥 Slow
S M T W T F S
Monday, 9:00 AM
8 AM 12 PM 4 PM 8 PM

Figure 4. Google Map of normal road conditions 5/9/2016 at 9:00am

Drivers used these types of screens to make navigation decisions while en route to their destinations. The real-time crowd-sourced data from Waze users interleaved by Google Maps with official VDOT road closing information proved invaluable to the driver's in this case study. Sadly, this information was only available on the small screen of the driver's smart phone.

5. Conclusion

This case study illustrates two interesting things. First, it shows how the cars that actually delivered the human-provided emergency response remained simple mobile platforms unintegrated into the wireless smart phone-to-cloud network. All communication between humans via text, Facebook or voice and real-time data systems such as Google Maps occurred through wireless cell phone network and internet cloud access. While a certain amount of communication that evening originated from cell phones and tablets inside buildings, most of the communication occurred via cell phones inside vehicles. One of the vehicles involved had vehicle-bound cell phone support capabilities such as "hands-free" car-cell phone connectivity. It did support the actual humans attempting to drive iced-over roads and navigate multiple car pile-ups [11]. This feature allowed a driver to keep both hands on the wheel while the phone conversation was carried over the vehicles sound system and driver spoken responses were captured and transmitted to the cell phone by a vehicle-based microphone. However, it is clear that visual data display such as real-time traffic and road conditions and possible navigation routes were only available from the display of the cell phone (held in one hand while the driver gripped the wheel with the other hand.)

There is discussion in on-line popular media about an Android Auto tool that allows a driver to view Google Maps map data on the car provided screen that is in beta testing. However, it is not wireless. It is tethered by both a USB cable and connected via Bluetooth [22]. It also does not include the critical feature of this case study's success story - real time crowd-sourced road and traffic data from Waze integration [22]. It requires purchasing an entire car with it preloaded or purchasing an aftermarket Pioneer unit that can be installed. In either case, the car starts by default in the proprietary interface. You must then tap the screen button to move to Android Auto [22.] This is not the effortless, fully mobile paradigm that users experience with a smart phone. The chief weakness of the Android Auto approach is that it continues the heterogeneous (and proprietary) hardware and software paradigm that is a main stumbling block to MCC as discussed in Section 2. Another weakness is that the approach does not capitalize on the vibrant and innovative mobile device app market. Improvements to both hardware and software for hand-held mobile devices will far outpace the development and deployment schedule for upgrades for vehicle-bound hardware and software.

There is also something called the Open Automotive Alliance that is a mix of technology and auto industry members who are working on a way to bring the Android platform to cars [23]. The list does not include several big name car manufacturers like Toyota and BMW [23]. The Alliance, while interesting, suffers from the same flaw as the Android Auto beta tool - the vehicle display screen is tied to one smart phone OS - and as such continues the practice of exclusive technology heterogeneity in MCC access.

A much better solution, if perhaps still a pipe dream, is an open-source vehicle display hardware and software design that could render whatever map graphic is on the user's smart phone straight to the car display via wireless connection. The capabilities that exist on the smart phone (access to either Google Maps or Apple Maps data, for example) would be fully visible on the car display screen. The data would come from whatever map app the user used on whatever smart phone (iOS or Android - based) the user chose. The smart phone would still be the originator of the data but the display would be rendered on the larger and more conveniently placed car display screen. This would be similar in concept to the "hands free" bluetooth cell phone to vehicle sound system connectivity capability that many vehicles have that is OS and mobile device independent (at least in the user's perspective.) The topic of vehicle-to-smart device integration presents a rich research area.

The second interesting development that the case study demonstrates is how how currently deployed wireless network and cloud services can support spontaneous collaboration within an existing social-media-enabled friend community to provide real-time emergency services at the individual level. There was a small community of friends - many of whom live in the same general neighborhood - who also belonged the same "friend" community on Facebook. All the people involved carried smart phones with them. In the traffic gridlock crisis in Northern Virginia, on January 20, 2016, the existing communication infrastructure and real-time cloud collection and dissemination of crowd-sourced data, allowed the "friend" community members to respond collaboratively and effectively to rapidly changing, often dangerous, circumstances.

References

[1] P. Bahl, R. Han, L. Li, M. Satyanarayanan, "Advancing the State of Mobile Cloud Computing," Proceedings of the third ACM workshop on Mobile cloud computing and services (MCS'12), Low Wood Bay, Lake District, UK, June 25, 2012. pp 21-28.

[2] R. Chang, J. Gao, V. Gruhn, "Mobile Cloud Computing Research - Issues, Challanges, and Need," Proceedings of the 2013 IEEE 7th International Symposium on Service-Oriented System Engineering (SOSE), San Francisco, Ca, USA, March 25-28, 2013, pp 442 - 453.

[3] T. A. Geumpana, F. Rabbi, J. Lewis, "Mobile Cloud Computing for Disaster Emergency Operation: A Systematic Review," Proceedings of the 2015 IEEE International Symposium on Technology and Society (ISTAS), Melahide, Ireland, November 11-12, 2015, pp 1-8.

[4] M. Guerla, E. Lee, G. Pau, U. Lee, "Internet of Vehicles: From Intelligent Grid to Autonomous Cars and Vehicular Clouds," Proceedings of the 2014 IEEE World Forum on Internet of Things (WF-IoT), Seoul, South Korea, March 6-8, 2014, pp 241-246.

[5] D. [5],"Facebook Places: Here's how it works," *CNN - Social Media*, August 19, 2010.

[6] J. Jeong, H, Jeong, E. Lee, T. Oh, D. Du, "SAINT: Self-Adaptive Interactive Navigation Tool for Cloud-Based Vehicular Traffic Optimization," IEEE Transactions on Vehicular Technology, Vol. PP, Issue: 99, 2105, 16 pages.

[7] M. McLaughlin, "WAZE (for IPhone)," *PCMag.com*, October 16, 2015.

[8] M. McLaughlin, "GoogleMaps (for IPhone)," *PCMag.com*, October 20, 2015.

[9] Z. Sanaei, S. Abolfazli, A. Gani, R. Buyyi, "Heterogeneity in Mobile Cloud Computing: Taxonomy and Open Challenges," I*EEE Communications Surveys & Tutorials*, Vol. 16, No. 1, First Quarter 2014, pp 369-392.

[10] C. Shi, M. Anmar, E. zegura, M. Naik, "Computing in Cirrus Clouds: The Challenge of Intermittent Connectivity," In Proceedings of the First Edition of the MCC workshop on Mobile Cloud Computing(ACM MCC'12). ACM, New York, NY, USA, 23-28.

[11] G. Simolunas, emails, conversation threads and Google Map screenshots from Facebook, Jan 21, 2016, May 8-9, 2016.

[12] T. Stenovec, "Google has gotten incredibly good at predicting traffic - here's how," *TechInsider.com*, November 18, 2015.

[13] C. Stobling, "The New Apple Maps vs. Google Maps: Which Is Right For You?," *HowToGeek.com*, September 29, 2015.

[14] I. Stojmenovic, "Fog Computing: A cloud to ground support for smart things and machine-to-machine networks," Telecommunication Networks and Applications Conference (ATNAC), 2014 Australasian, Melbourne, Australia, Nov 26-28, 2014.

[15] JR. Yu, J. Ding, X. Huang, M. Zhou, S. Gjessing, Y. Zhang, "Optimal Resource Sharing in 5G-enabled Vehicular Networks: A Matrix Game Approach," IEEE Transactions on Vehicular Technology, Vol. PP, Issue: 99, 2106, 13 pages.

[16] A. Zelchick, "How Facebook Works," *MIT Technology Review*, June 23, 2008.

[17] Facebook Website - data centers, http://www.datacenterknowledge.com/the-facebook-data-center-faq/, last visited 5/9/2016.

[18] Google Website - traffic symbols, https://support.google.com/maps/answer/3092439?co=GENIE.Platform%3DDesktop&hl=en, last visited 5/9/2016.

[19] Google Website - data center locations, https://en.wikipedia.org/wiki/Google_platform, last visited 5/9/2016.

[20] Washington Post Website - weather/traffic news for 1/20/2016, http://wtop.com/winter/2016/01/snow-storm-likely-to-dump-inches-of-snow-friday-night-saturday-morning/slide/1/, last visited 5/9/2016.

[21] Weather Website - Blizzard Jonas news, https://en.wikipedia.org/wiki/January_2016_United_States_blizzard, last visited, 5/9/2016.

[22] F. Ion, "Android Auto review: The best way to get Google Maps in your car," *greenbot* by IDG Communications, June 4, 2015, (http://www.greenbot.com/article/2931099/android-auto-review-the-best-way-to-get-google-maps-in-your-car.html, last visited May 12, 2016.)

[23] F. Ion, "11 things you need to know about Android Auto," *greenbot* by IDG Communications, April 30, 2015, (http://www.greenbot.com/article/2914832/11-things-you-need-to-know-about-android-auto.html, last visited May 12, 2016.)

[24] R.Popely, "Which Cars Have Android Auto?," cars.com, February 9, 2016, (https://www.cars.com/articles/which-2016-cars-have-android-auto-1420681175443/, last visited May 12, 2016.)

A Flexible Method for Sharing Network Bandwidth Utilization in Local Network by Applying the Advantage of Fuzzy Control: An Example on WebRTC Streaming

Linh Van Ma[1], **Sanghyun Park**[2], **Jinsul Kim**[3]
School of Electronics and Computer Engineering, Chonnam National University, 77
Yongbong-ro, Buk-gu, Gwangju 500-757, Republic of Korea
linh.mavan@gmail.com[1], sanghyun079@gmail.com [2], jsworld@jnu.ac.kr[3]

Abstract - *Limitation of resources and uncertainty of environment in the network communication field, leading us to a problem of how to use the most effective utilization of bandwidth. This research, therefore, focuses on a work of sharing and balancing bandwidth in a local network, some prior users subscribe its service first and consume much bandwidth capacity, it makes new coming users cannot gain its service because it does not have enough requiring amount of the service bandwidth. Thus, we try to distribute the bandwidth of the users who have much utilization, to the new coming, by that way we make a proportional usage with a negotiating resource process between users. In order to do that, we use fuzzy logic control method and set it up to each user, in which each one estimates its network state, and sends a bandwidth request to others, or shares its consumption to others by using methods of decreasing bandwidth usage. In this paper, we design a fuzzy system which uses some equations to calculate input, output system parameters; we also prove that the equations work well in all conditions of the network. In the experiment, we simulate to show how our method works by using OPNET with the fuzzy system in each user. A comparing result in an experiment between our method and original implementation of WebRTC, which is one of the most recent real-time communication technologies, shows that this approach improves video streaming between peers efficiency.*

Keywords: Low bandwidth, Sharing bandwidth, Fuzzy control logic, WebRTC real time communication, Reduce frame rate, Reduce resolution

1 Introduction

Along with the development of the Internet, bandwidth has a vital role in providing and qualifying network quality. Recently, streaming media over the Internet presents many challenges. On one hand, the current network provides fixed up and down bit stream for a group of local users, but they always demand a high quality of serving services with various kinds of requirements such as big data processing, parallel processing, HD streaming video services. On the other hand, the network has its own problem with inherent infrastructure which causes the work of changing the structure of the network is not easy with the involved of many components, otherwise, we must improve network management [1, 2]. Many previous studies have been conducted in the field of media transmission for optimizing bandwidth usage with streaming media, low transmission [3, 4]. Thus, improving efficient utilization of bandwidth is necessary. The more effective bandwidth uses, the better quality service is, in network communication application. For the above reason, this paper concentrates on improving bandwidth utilization for a group of users in the local network where each user knows other. Each user needs a fuzzy system which has an ability to estimate its current network state whether it has low, medium or high speed. Then the system designs a method to reduce or increase resource consumption. To form the input system, we compare values between current usage and service requirement. However, the process to produce the fuzzy output system depends on services which user subscribes to. Hence, we choose WebRTC video streaming service [5, 6, 7] to demonstrate and illustrate our research in the most straightforward way. Our work is primarily focused on network transmissions, such as file transfer and multimedia streaming with two main contributions. First, we proposed bandwidth sharing system, which uses the fuzzy control system to share bandwidth between different users on the local network according to requesting bandwidth users. The second contribution is a smoothing method for decreasing bandwidth usage, such as frame rate and resolution in the case of video streaming.

The rest of this paper is organized as follows. In Section II, related works on the packet scheduling problem and some other related fields, are introduced. In Section III, we describe our proposal system using Mandani fuzzy model, which includes a description of sharing and requiring bandwidth estimation with fuzzification and defuzification processes. In Section IV, we present the way of how our method works with the simulation of OPNET - a tool that provides performance management for computer networks and applications. Section V provides conclusions and discussions.

2 Related research

Making a plan to distribute bandwidth is an important task in a large computing system, based on [8], a decentralized, accurate, and low-cost system that predicts pairwise bandwidth between hosts [9] proposed an algorithm which constructs a distributed tree that embeds bandwidth measurements without any centralized component requirement. By that way, they determine the performance of distributed computing applications in a computing system.

Network coding brings substantial improvements in terms of throughput and delay in collaborative media streaming applications. To overcome this critical redundant transmission problem [10] addressed the problem of finding a suitable asynchronous packet scheduling policy in collaborative media streaming applications.

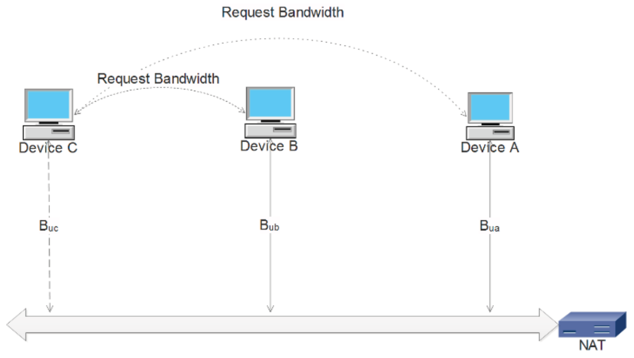


Fig.1. A local network with negotiating bandwidth proposal method.

In the field of fuzzy control logic [11] presented a connection admission control method that uses a type-2 fuzzy logic system. The system can combine the input rate of real-time voice, video traffic, and non-real-time data traffic in the decision of connection admission combine the experiences from lots of experts so that an acceptable decision boundary can be obtained. Also, it provides an interval decision, so that a soft-decision can be made based on a design tradeoff between cell loss ratio and bandwidth utilization.

3 Proposal system overview

In the proposal system has three devices A, B, and C in a local network, each device consumes bandwidth with value represents B_{ua}, B_{ub}, B_{uc} correspondingly. At first, device A and B are active at some services and they take much of bandwidth resource in the network, which describes by $B_{ua} \approx B_{ub} \gg B_{uc}$. For the above reason, device C cannot get a desired service because it does not have enough bandwidth for gaining the service. Hence, device C will negotiate with A and B for getting network resources by sending a request as shown in a dashed arrow Fig. 1. For better explanatory our research, we do not mention any case of complexity in the network infrastructure, and outside affection of the network.

3.1 Fuzzification input system

The proposal uses Mandani Fuzzy Model, and triangle shapes membership function. It has two input variables which are sharing ability and resource requirement of a client. The output variable of the system is the amount of bandwidth that user can reduce. Our method focuses on the client itself, thus, each client has their system for calculating and requesting resources. Depending on the current state of network utilization, each client estimates its fuzzy usage with a value; the process is so called Fuzzification. Then, the human knowledge and constraint rules involve for calculating output; the process is called Fuzzy Inference. Finally, the Defuzzification defuses output fuzzy values, from that, the system makes a decision decreasing bandwidth.

Assuming that, B_S is a current utilization bandwidth of a prior device which established a connection with some services and was consuming the bandwidth of the local network. B_R is a current utilization bandwidth of a device which tends to be subscribed to a specific service, however, it cannot reach the service because the bandwidth almost consumed by the device has B_S. Furthermore, we name B_M is a minimum required bandwidth to gain a specific service in the user. Thus, different users have different B_M values of their subscribed service.

By comparing the current consuming bandwidth value with minimum requirement, we form scalar values for the input system by using two equations below:

$$S = \frac{B_M - B_S}{B_S} \times 100. \qquad (1)$$

$$R = \frac{B_M - B_R}{B_M} \times 100. \qquad (2)$$

The sharing variable in (1) and requiring variable in (2) has range [-100, 100], the negative value in (1) means that it can reduce bandwidth usage, the positive value in (2) means that it need more bandwidth, and it cannot share bandwidth to other devices on the network and vice versa. In (2), if the users already established services and consumed most of the bandwidth in the network, then the device which tends to establish a new connection has a small bandwidth consumption when compares with minimum requirement, $B_R \ll B_M \Rightarrow R \approx 100$, it makes the sharing users reduce much more resource, and the serving service become low. In this case, we make a new equation which supports the user gains a service by making a comparison with B_M and B_S, and equation (2) becomes:

$$R = \frac{B_M}{B_S} \times 100. \qquad (3)$$

3.2 Fuzzy inference process

As mention above, we define the symbol as positive or negative in accordance with the increasing or decreasing bandwidth utilization. The input variables are the requirement and sharing ability, the output variable is the amount of bandwidth that the sharing user needs to decrease bandwidth usage. The fuzzy values of the input and output variables describe as Negative High, Negative Low, Negative, Medium, Normal, Positive Low, Positive Medium, Positive High or their abbreviation words are NH, NL, N, PL, PM, PH by increasing order.

In the system, we have two input variables, *R* represents the fuzzy value of requiring bandwidth, and *S* represents the fuzzy value of sharing ability. The output of the system *Op* is an amount of a device which can reduce bandwidth usage depends on the request from another device. Therefore, we have a set of fuzzy rules as follows,

If R is PH and S is NH, then Op is NH

If the requiring is positive high (it needs much bandwidth), and the sharing is negative high (it is eager to share resource), then the output is negative high (the sharing user reduces much of its utilization). Probably, we do nothing if R is kind of Negative because we do not want to decrease the bandwidth of the required device, meanwhile, it is requesting. Also, in the case S is kind of Positive which means the sharing devices need more bandwidth, they cannot share bandwidth with others. Thus, the output is N (keeping its current state). The discussion of those rules between input and output is shown in Table I.

Table 1. Fuzzy Inference of bandwidth

R/S	NH	NM	NL	N	PL	PM	PH
NH	N	N	N	N	N	N	N
NM	N	N	N	N	N	N	N
NL	N	N	N	N	N	N	N
N	NL	NL	N	N	N	N	N
PL	NL	NL	N	N	N	N	N
PM	NM	NM	NL	NL	N	N	N
PH	NH	NH	NM	NL	N	N	N

3.3 Defuzification output system

Based on human knowledge, triangle shapes membership function, and fuzzy rules which were given above, the system calculates output variable $\mu_O \in [-1, 1]$ by using the centroid method. Finally, we have the crisp value of the output system but it is still fuzzy value. Thus, we use (5) to calculate a real output amount *Dec*, of the system.

$$\mathrm{Dec} = (\mathrm{B_S} - B_M) \times \mu_O. \qquad (5)$$

The max-min inference method produces output μ_O variable, is formed as:

$$\mu_O^i = max\left[\min[\mu_S^i, \mu_R^i]\right]. \qquad (6)$$

With $\mu_O^i, \mu_R^i, \mu_S^i$ is values of output membership function corresponding to input value i^{th}, R and S. The relationship between input and output is shown in Fig. 2.

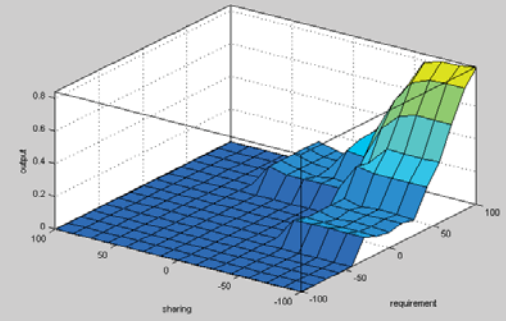


Fig. 2. Relationship between input and output of the fuzzy system

If the user does not have priorities in consuming bandwidth which means all users have equality to take resources, then we reform equation (1), (2) with Avrg is the average bandwidth consumption of the local network. Assuming that Avrg ≥ $\mathrm{Min_r}$,

$$\mathrm{S} = \frac{Avrg - B_S}{B_S} \times 100. \qquad (7)$$

$$\mathrm{R} = \frac{Avrg - B_R}{Avrg} \times 100. \qquad (8)$$

Using the same strategy above with equation (7) and (8), we have a final balance bandwidth usage after the negotiation in the local network because each sharing and requiring user tends to come *Avrg* value.

4 An example on WebRTC

Aiming to prove that our method stands out from other methods, we choose a technology which is one of the most advanced real-time communication technologies recently, WebRTC. While our method works well in low bandwidth condition, WebRTC also works well in the similar condition, such as the WebRTC audio and video engines work together with the underlying network transport to probe the available bandwidth and optimize delivery of the media streams [6]. However, DataChannel in WebRTC API transfers require additional application logic: the application must monitor the amount of buffered data and be ready to adjust as needed. Thus, it takes some time to evaluate and may cause package loss. In this section, we discuss an analysis of a negotiating algorithm, after that, we simulate our method for balancing bandwidth utilization in OPNET and illustrate the result by using open source EasyRTC [12].

A sending rate of video streaming with *r* represents frame rate or refresh rate, *y* represents total vertical lines of active pixels, *x* represents total horizontal of active pixels, and $\alpha \in \mathrm{R}^+$ is a factor when the original rate is affected by network environment, it calculates by equation (9),

$$\mathrm{f(x, y, r)} = \alpha \times xyr. \qquad (9)$$

The strategy for reducing bandwidth utilization is that we reduce the frame rate until it gets minimum value Min_r (Max_r is the maximum frame rate), and then reduce the resolution. However, in the real world, the resolution of the digital camera has a discrete type, and the total horizontal, vertical pixel are somehow fixed in a resolution list corresponding to the type of camera, we cannot choose it randomly.

In the sharing device, from original resolution and frame rate (x_0, y_0, r_0), we reduce the rate to a new point (x_i, y_i, r_i) with several steps which is called smoothing process. From (5) and (9), we yield $\mathrm{Dec} = \alpha(x_0 y_0 r_0 - x_i y_i r_i)$, we divide the smoothing process into two cases:

First, we keep the resolution $x_i = x_0, y_i = y_0$ and reduce frame rate if it satisfies (10) with $[a] = \min\{z \in Z, |z| \le |a|$, we move to the second step if the condition does not satisfy.

Second, finding next resolution in the resolution list which has lower rate and satisfying (10),

$$\mathrm{Min_r} \le r_i = \left[\frac{\alpha \times x_0 y_0 r_0 - Dec \times \beta}{x_i y_i \times \alpha}\right] \le Max_r. \qquad (10)$$

$$\mathrm{Min_r} \le r_i = \left[\frac{Dec \times \beta}{x_i y_i \times \alpha}\right] \le Max_r. \qquad (11)$$

In the requiring user, after the sharing user decreases bandwidth usage, it has an amount Dec × β bps with $\beta \in R^{+}$is a factor which the environment impacts on the reducing amount of the network. If $\beta \sum_i Dec_i > B_M$ which means, it satisfies the requirement of the service with *Dec_i* is an amount of decreasing bandwidth at i^{th} sharing user. Thus, the resolution of requiring device is calculated by equation (11). When we decrease the rate from (x_0, y_0, r_0) to (x_i, y_i, r_i) in just a while, it makes inconvenience to the user because of video vibrant resolution. Thus, we must smooth the decreasing process by spreading out the decreasing time *t*. Suppose that (x_i, y_i, r_i) is an intermediate step between step 0 and i^{th}, if the resolution changes in step *j* then $r_j = Max_r$. Otherwise, (x_j, y_j) is invariable and r_j is given by (13), $r_0 > r_i$ because the resolution does not change. Assigning B_{Step} is the number of decreasing resolution steps. A resolution takes a higher efficient quality than the frame rate. Thus, we assign B_{Step} with a weight higher than the weight of step for decreasing the frame, the total steps in the process is,

$$\text{step} = B_{Step} + \left(\frac{[r_0 - r_i]}{4} + 1\right). \qquad (12)$$

$$r_j = \left[\frac{r_0 - r_i}{step}\right] + r_i. \qquad (13)$$

In equation (12), step ≥ 1 even if $B_R = 0$ and $r_0 = r_i + 1$ or $r_0 = r_i - 1$. Given a time *t*, each step must take $\Delta t = t/step$ seconds.

At the beginning of this section, we propounded the method which describes how the new coming user, C gets streaming video in the case bandwidth almost saturated by A and B. The process for gaining resource is a *negotiating algorithm* with following steps:

Step 1: Each device determines its current state in the network, such as streaming video resolution and frame rate, minimum video streaming requirement, camera resolution list, and network bandwidth capacity.
Step 2: According to the current streaming state, new device requests resource from current devices.
Step 3: In each current device do a fuzzification depending on its current state.
Step 4: In each current device do a defuzzification to estimate the decreasing utilization bit rate.
Step 5: Calculating new resource and frame rate for all devices.
Step 6: In this step, the algorithm drive into two ways. First, if new device can get streaming video after the previous steps, then we move to the next step. Otherwise, we back to step 4. Secondary, we try to make a proportion of resource utilization for every device. If the condition satisfies, then we move to the next step. Otherwise, we back to step 4.
Step 7: Assigning new resolution and frame rate for preparing step 8.
Step 8: In case of sharing device, we will make a smooth decreasing process to make the convenience. If it is a new coming device, its bandwidth usage is not decreased, and then we end the process.
Step 9: Comparing the new value and initial value to make a smoothing process which is given in equation (12).

4.1 Simulation and experiment

To show how our method works, in this section we use network simulation tool OPNET, to illuminate the negotiation process between devices in the network. The network topology in the simulation is shown in Fig. 3 with three devices A, B, and C. We conduct the experiment with an initial camera resolution. For less complexity, we choose α = 1, β = 1 in (9), (10), and (11). Both video source and destination have no encoder and decoder, the transmit rate sets to 8 Mbytes.

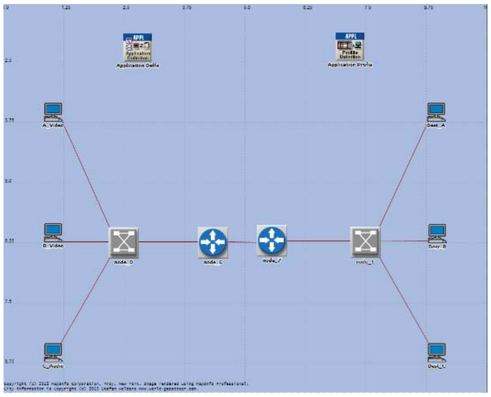


Fig. 3. Network topology of simulation

Three devices A, B, and C have video conferencing service with the imbalance bandwidth utilization. Each device negotiates with others and takes an average bandwidth utilization of the current network as the step of result is shown in Table II with one smoothing step between step 0 and 1.

Table 2. Balancing bandwidth usage in the local network

Step	Stream A (H/V/F)			Stream B (H/V/F)			Stream C (H/V/F)		
0	400	240	30	320	224	35	0	0	0
-	-	-	-	320	224	33	-	-	-
1	400	240	27	320	224	31	160	144	24
2	400	240	25	320	224	29	160	144	38
3	400	240	24	320	224	28	256	192	21
4	400	240	23	320	224	27	256	192	24
5	400	240	22	320	224	26	256	192	27

WebRTC works well in low bandwidth condition because it has an API for negotiating and assigning bandwidth amount to each peer bases on the network condition. The frame rate and resolution automatically change to match the available bandwidth. Thus, in the experiment, we only demonstrate the proportional bandwidth process and compare our research with the current implementation of WebRTC in EasyRTC [12]. The experiment uses EasyRTC server, which is installed on Ubuntu 12.04, the limited uploading speed of the network sets to 200Kbits by using an open-source Trickle. To create a virtual webcam device we use WebcamStudio open-source, in which we can control the resolution and frame rate of a video. Two hosts run on Chrome browser. It makes video streaming with two hosts in another computer which has Chrome browser. When the connection established between two hosts and its destination, we access to chrome WebRTC statistic for getting parameters measurement at *chrome://webrtc-internals/*.

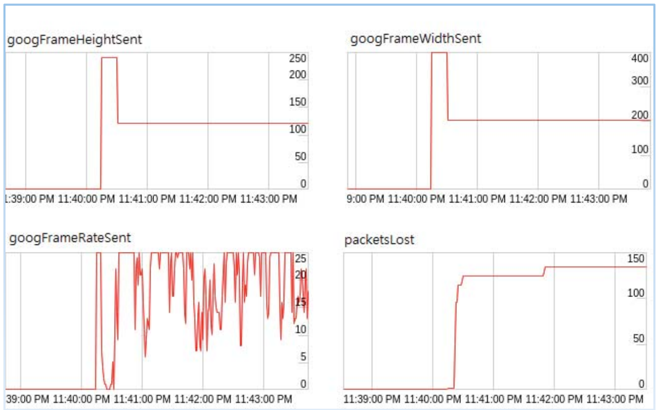


Fig. 4 (a): A statistically in an original host using EasyRTC for video calling to another host in other computers.

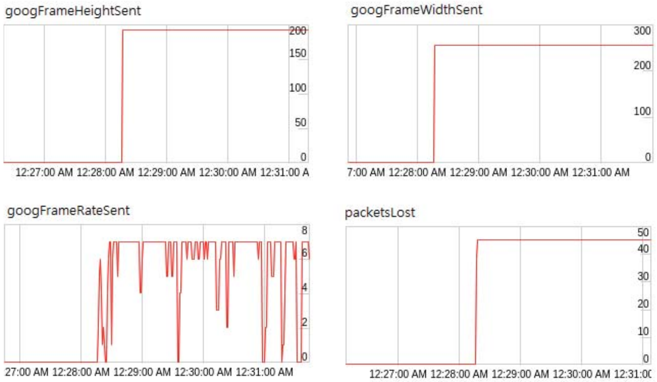


Fig. 4 (b): A statistically in a host which uses the negotiating algorithm for detecting resolution and frame rate before sending using EasyRTC for video calling to another host in other computers.

In Fig. 17 and Fig 18 show that proposed method reduces a certain package lost with the same quality of information per frame, also, the frame rate is less variant than original EasyRTC implementation. By using the method, we detect the initial of the matching resolution before making streaming service. It does not only reduce package lost but it also reduces the computing resource in both sides of streaming service.

5 Conclusions

In this paper, we have presented the flexible method for sharing and balancing bandwidth of the local wired and wireless network using fuzzy control theory with the simulation using OPNET and the experiment using the implementation of WebRTC, EasyRTC. The method does not provide the best quality for each individual, but every user in the same group deserves to have a beneficial service with the lowest quality instead of losing service. By such way, we proportionate the quality service for every user in the same group where has no priority of serving service. Throughout the experimental section, the proposed method solved completely the problem lead on introduction section. Each sharing device just gives a small amount of bandwidth that does not have much effect on its service, but the total decreasing amount is higher when we have more and more involving users. So far, the experiment demonstrated the proportion of the bandwidth utilization beyond the requirement of the introduced problem.

In the future works, this method does not only apply for video streaming application, but it can also use to make a sharing protocol between devices, such as computer, mobile, printer, Set Top Box etc., in the same network.

6 Acknowledgements

This work (Grants No.C0218369) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2014, and also supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2013R1A1A2013740)

7 References

[1] H. Kim and N. Feamster, "Improving network management with software defined networking," Communications Magazine, IEEE, Vol. 51., pp. 114-119, 2013.

[2] M. Baldi, S. Gai, and G. P. Picco, "Exploiting code mobility in decentralized and flexible network management," in Mobile Agents, pp. 13-26, 1997.

[3] S. H. Kang and A. Zakhor, "Effective bandwidth based scheduling for streaming media," Multimedia, IEEE Transactions on, Vol. 7., pp. 1139-1148, 2005.

[4] H. Qi, M. Stojmenovic, K. Li, Z. Li, and W. Qu, "A low transmission overhead framework of mobile visual search based on vocabulary decomposition," Multimedia, IEEE Transactions on, Vol. 16., pp. 1963-1972, 2014.

[5] A. B. Johnston and D. C. Burnett, WebRTC: APIs and RTCWEB protocols of the HTML5 real-time web, Digital Codex LLC, 2012.

[6] A. Bergkvist, D. Burnett, and C. Jennings, "A. Narayanan," WebRTC 1.0: Real-time Communication Between Browsers," World Wide Web Consortium WD WD-webrtc-20120821, 2012.

[7] A. Johnston, J. Yoakum, and K. Singh, "Taking on WebRTC in an enterprise," Communications Magazine, IEEE, Vol. 51., pp. 48-54, 2013.

[8] LU SH, YIN JP, CAI ZP, ZHAO WT, "Efficient available bandwidth estimation for network paths," Journal of Harbin Institute of Technology, Vol. 1, No.032., 2008.

[9] Song S, Keleher P, Bhattacharjee B, Sussman A, "Decentralized, accurate, and low-cost network bandwidth prediction," INFOCOM, 2011 Proceedings IEEE, pp. 6-10, Apr 10. 2011.

[10] T S. H. Hao, E. Izquierdo, and Pengwei, "Bandwidth-efficient Packet Scheduling for Live Streaming with Network Coding," IEEE Transactions on Multimedia, 2016.

[11] Liang, Qilian, Nilesh N. Karnik, and Jerry M. Mendel, "Connection admission cont1rol in ATM networks using survey-based type-2 fuzzy logic systems," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, Vol. 30 No. 3., pp.329-339, 2000.

[12] Priologic Software Inc, “EasyRTC Documentation”, 2014.

SESSION

NOVEL ALGORITHMS, APPLICATIONS AND DEVICES + AMBIENT INTELLIGENCE

Chair(s)

TBA

Ranged Filtering of Streaming Numeric Data, or Geolocation Filtering of Streaming GPS Data, using Topic-Based Pub/Sub Messaging

Aaron W. Lee
© Solace Systems Inc., Ottawa, Canada

Abstract— *This paper describes an innovative method to implement location-based filtering of streaming GPS point data without the use of a geospatial database or GIS engine. The filtering algorithm takes a set of polygons as input, and generates an approximating shape of a specified relative accuracy as output, consisting of a union of rectangles. These rectangles can be used to perform point-in-polygon comparisons quickly using text-based pattern matching in a pub/sub messaging system, which excels at handling streaming data at scale. This method has applicability in transportation, logistics, smart cities, security, and surveillance, and can be generalized to one or three dimensions.*

Keywords: geolocation, GPS, filtering, messaging, pub/sub, MQTT

1. Introduction

The number of connected devices and associated generated data continues to rapidly increase. Finding efficiencies in handling this growing volume of data, and deriving value and utility from information in real-time is becoming essential. Messaging systems are emerging as a central technology in the domain of the Internet of Things for connectivity at scale, as well as being able to efficiently route the torrents of information in the Big Data realm.

To filter streaming data based on location, or to provide geo-fencing capabilities (detecting an entity entering or exiting a defined area), one may consider using a Geographic Information System (GIS) or some form of geospatial database to comprehend location-based semantics. Retrieving data from these systems is often poll-based, running queries at set intervals, and may not scale to large numbers of devices or modern real-time, event-based requirements.

As an example, a transportation/delivery company may have tens of thousands of vehicles, hundreds of depots, and hundreds of thousands of potential delivery sites; or a municipal transit service with 5,000 buses and many times more passengers. If every vehicle was generating GPS updates every second, and if every depot and every passenger wanted to be able to receive the live streaming location of the vehicles within a set distance from their location, consider the sheer number of updates and queries these information systems would have to be able to support.

I present a method by which to filter streaming coordinate data to a specified area using a publish-subscribe messaging system that utilizes text comparison rules for routing of information. This paper shall:

- Define messaging systems, pub/sub, and nomenclature around routing with topics
- Demonstrate how topics can be used to encode coordinate data, and how topics can be used to match a rectangular geographic area
- Provide an algorithm to generate a filtering mechanism to approximate a given area of interest

The method presented here is in the context of 2-dimensional latitude/longitude GPS coordinates, but could be extended to any n-dimensional bounded real number range. Additional applications could include: streaming 1-dimensional numeric readings from pressure, temperature, or vibration sensors (e.g. pipeline or bridge monitoring); 2-dimensional Military Grid Reference System (MGRS) coordinates (e.g. ground Moving Target Indication (MTI) radar data); or 3-dimensional coordinates (e.g. aircraft position).

2. Topic-Based Pub/Sub Messaging

Message-Oriented Middleware (MOM) infrastructure allows the sending and receiving of data, in the form of *messages*, between distributed components using the concept of a shared bus [1]. Many MOM systems use a centralized message *broker* to route and filter the data between components, which architecturally decouples producers and consumers of the message data [2]. While there are several ways to exchange message data, one typical pattern used is the publish-subscribe pattern, or *pub/sub* [3].

As illustrated in Figure 1, when data is published onto the message bus for distribution in a topic-based pub/sub system, it is associated with a *topic* to help describe the data, which provides a mechanism by which to help route it. Components, or *clients*, that wish to receive data about a particular subject will register one or more *subscriptions* with the message broker to indicate their interest. When the data, or message, is published onto the bus, the broker compares the topic of the message with the known subscription list of every connected client, and for every client that has at least

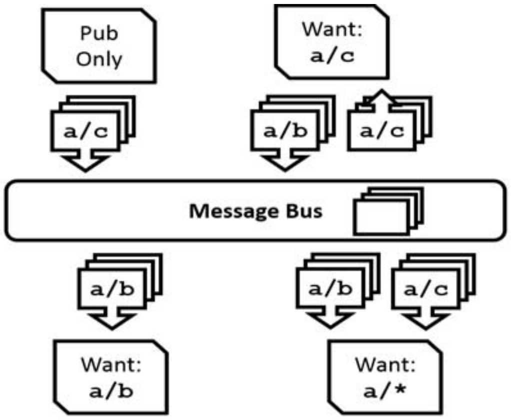


Fig. 1: Simplified diagram of pub/sub architecture

one matching subscription, the message will be delivered to them. This message delivery is typically done in a push-based asynchronous manner without the need to poll the message broker.

2.1 About Topics

In some topic-based publish-subscribe messaging systems, topics are simply text labels, and routing based on subscriptions is essentially a straight string comparison. Other pub/sub systems use hierarchical topics and subscriptions to route the data to interested parties [4]. The exact implementation and syntax of the topics can vary from one messaging system to another, but many of them share similar features. For the purposes of this paper, the following definitions will be used:

Topic string: defined by the publishing client, this is the entire topic belonging to the published message, consisting of one or more levels. E.g. `animal/dog/husky`

Topic delimiter, or *topic level separator*: a (typically) single character used to separate the topic string into multiple levels, giving rise to a hierarchy. In the above example, the topic delimiter is "/" or slash.

Topic level: a portion of the topic string contained within the topic hierarchical delimiters. E.g. "`animal`" or "`dog`"

Topic subscriptions are sent to the message broker by consuming applications to indicate the particular data they are interested in receiving. When a message arrives on the bus that matches a particular subscription, that data will be sent to the corresponding client. A topic subscription can contain one or more wildcards.

Wildcards allow a single topic subscription to match more than one topic string, using rules similar to very basic pattern-matching regular expressions. Although different pub/sub system can implement wildcard matching differently, for the purposes of this paper, define the following topic subscription wildcard:

> * (star): a single-level wildcard, matching zero or more text characters within a single level (between topic delimiters). Only one * wildcard can appear within each topic level. This wildcard can have characters preceding it within the level, thereby providing a prefix match at that level.

Subscription:	`animal/dog/*`
Matches topic:	`animal/dog/husky`
	`animal/dog/poodle`

Subscription:	`animal/do*/h*`
Matches topic:	`animal/dog/husky`
	`animal/dolphin/hector`
Doesn't match:	`animal/dog/poodle`
	`animal/dragon/horntail`

The outcome of the topic subscription match operation is entirely determined by successive text string comparisons: for the match to be successful, each topic level is considered individually and must match the corresponding level of the subscription exactly (if no wildcard), or match the prefix characters exactly (with a wildcard).

Architects and users of a hierarchical topic-based messaging system can specify a standard format for the topics published onto the bus for applications to use. Below are some real-world examples of topic formats:

- `APP_ID/PUB_ID/PRI/DEST_SYSTEM/SUBJECT`
  e.g. `TALON/053B/3/RISK/EOD_REPORT`
- `VEH_TYPE/NUM/LAT/LON/MSG_TYPE`
  e.g. `CAR/0034/_45.382/-075.751/UDPATE`

Building descriptive topic string formats using relevant parts of the accompanying data allows complex and powerful routing rules to be defined. For example:

```
TALON/*/3/*/EOD*
```

A client application subscribing to this would receive all Priority 3 End-of-Day messages published by any client in the Talon application, regardless of the intended destination.

2.2 Encoding Geolocation in a Topic String

I will now define a format to encode location data into a topic string. The following assumes coordinate data will be in decimal degrees [5], but a similar approach could be used to route based on Universal Transverse Mercator (UTM) or MGRS coordinates, Eastings and Northings within a specific grid.

Define two levels within the message topic string to denote the latitude and longitude of the data in decimal degrees format. The range of possible values are therefore:

- Latitude: -90.0 to +90.0
- Longitude: -180.0 to +180.0

Hence, for the latitude coordinate, we need 2 digits before the decimal point, and one character for a plus or minus sign; similarly, 3 digits before the decimal point are required for longitude. If we only consider these two levels of the topic hierarchy, we can provide some examples of encoding:

- Topic: `_45.38/-075.75/`
  Location: 45.38°N, 75.75°W, in Ottawa, Canada
  2 decimal places of accuracy (approx. 1km resolution)
- Topic: `_51.5160/-000.0707/`
  Location: 51.516°N, 0.0707°W, in London, UK
  4 decimal places of accuracy (approx. 10m resolution)

The use of the underscore symbol "_" is used to differentiate non-negative coordinates; the plus symbol "+" could be used, however some messaging standards have special meaning for this character [6]. Note that the coordinates must be zero-padded to prevent accidental subscription matches; this will be explained more fully in the next section.

2.3 Geolocation Topic Subscriptions

After defining a method to describe the data's geographic location in the topic string, it remains to provide a method to subscribe to the data. Consider the two levels for latitude and longitude individually, as the following applies to both equivalently.

As the message broker treats the topic string and topic levels as text, the geographic coordinates must be matched using the text-based comparison in the topic subscription. This is not a numeric comparison: a topic "45.38000" would not match a subscription "45.38" although they are numerically equal. However, using a single-level wildcard, a continuous range of numbers can be matched to the given topic subscription. Consider the following examples for latitude:

- Subscription: `_45.38*/` would match any number that starts with these characters, such as: 45.38, 45.38000, 45.387312, or any number in the range 45.38 to $45.38\overline{9}$; or in interval notation: $[45.38, 45.39)$
- `_45.3*/` would match anything in $[45.3, 45.4)$
- `_45.*/` would match anything between 45°N inclusive and 46°N exclusive
- `_4*/` would match anything in $[40, 50)$, but due to the zero-padding mentioned at the end of Section 2.2 above, this will correctly *not* match numbers between 4 and 5, which is handled by: `_04.*/`
- `-45.3*/` would match anything between $-45.3\overline{9}$ and -45.3, or $(-45.4, -45.3]$
- `-45.*/` would match anything in $(-46, -45]$

Even though this is a text-based comparison, this method using wildcards permits the matching of a range of real numbers using a topic subscription.

By using two levels within the topic hierarchy to represent latitude and longitude, a range of two-dimensional coordinates can be matched using topic subscriptions with wildcards. That is, each subscription corresponds to an area.

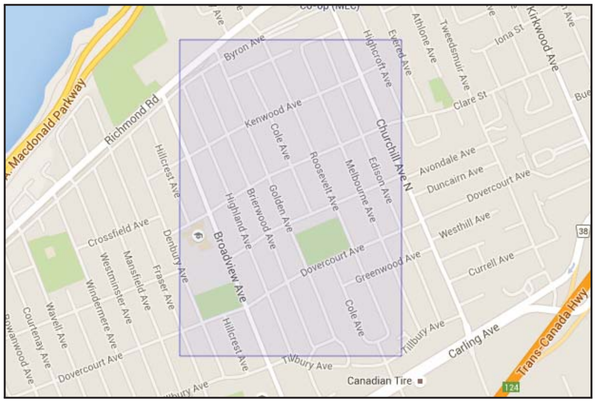


Fig. 2: Axis-aligned rectangle for `_45.38*/-075.75*/`
Background Map data ©2016 Google

Referring to Figure 2, a subscription defined as `_45.38*/-075.75*/` would match any coordinate inside the square area bounded by 45.38°N to 45.39°N, and 75.75°W to 75.76°W. Data published onto the message bus with a coordinate that intersects this box would be sent to the recipient via this subscription. Note that the displayed area in Figure 2 is not perfectly square due to longitude degrees becoming compressed towards the poles; this area would be square at the equator.

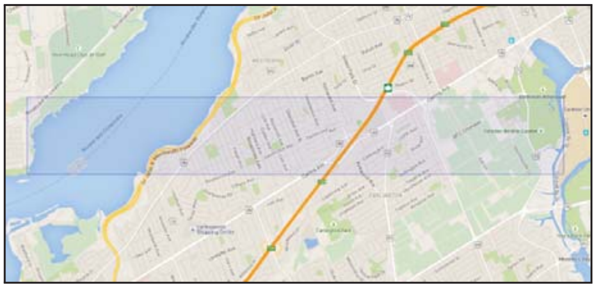


Fig. 3: Rectangle for `_45.38*/-075.7*/`
Background Map data ©2016 Google

Consider Figure 3. Subscription `_45.38*/-75.7*/` would match any coordinate in the long, wide rectangular area bounded by 45.38°N to 45.39°N, and 75.7°W to 75.8°W. As the wildcard on the second level is placed one decimal place to the left, this rectangle's longitude range is 10 times larger than its latitude, or 10 timers "wider" if drawing with north up. By changing where the wildcard is placed within the subscription string, it allows the creation of different shaped areas as follows:

- Square rectangle: decimal accuracy for both latitude and longitude portion of the subscription is the same.
  E.g. `_45.38*/-075.75*/`

- Horizontal rectangle: spans a larger range of longitude coordinates; the decimal precision for the longitude is at least one decimal places less.
  E.g. `_45.38*/-075.7*/`
- Vertical rectangle: spans a larger range of latitude coordinates; the decimal precision for the latitude is at least one decimal place less. E.g. `_45.3*/-075.751*/`

Observe that the square rectangle illustrated in Figure 2 is contained wholly within the horizontal rectangle in Figure 3, and is in fact exactly $^{1}/_{10}$ of the larger rectangle; this occurs as the corresponding topic subscription has one extra decimal degree of accuracy. More generally, it is possible to split any rectangle corresponding to a geolocation topic subscription described above, either horizontally or vertically, into ten equally sized sub-rectangles, or *child* rectangles, by replacing the larger topic subscription with ten smaller subscriptions that have the wildcard character moved one position to the right in either the latitude or longitude level respectively.

From these examples, it should be clear that there exists a 1-to-1 mapping between the defined wildcard topic subscriptions and axis-aligned rectangles that line up with decimal boundaries of the degree coordinates.

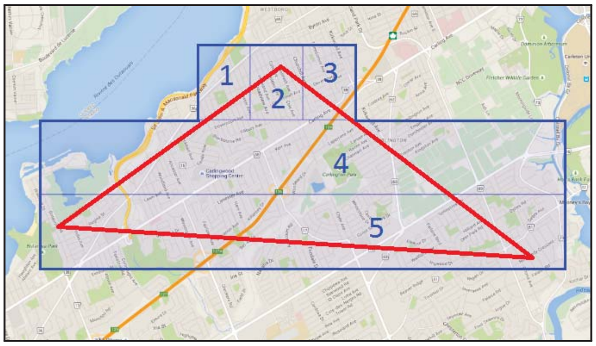


Fig. 4: Composite geometry of five rectangles
Background Map data ©2016 Google

Finally, by subscribing to multiple topic subscriptions simultaneously that cover different areas, it is possible to produce a composite geometry of boxes. Figure 4 shows a union of five rectangles that approximate a triangle. A message whose coordinates lie within the union of all the rectangular shapes will match one of the corresponding geolocation topic subscriptions.

3. Iterative Algorithm for Generating Subscriptions

I now present an algorithm for generating a set of rectangles to approximate a given input geometry or geometries. As demonstrated in the previous section, a geolocation topic subscription can be represented by specific a rectangular axis-aligned shape, whose edges line up with the decimal boundaries. Hence, by providing an algorithm to generate a set of these rectangles which can be converted back to corresponding topic subscriptions, the solution to the problem of filtering streaming point-location data to an input geometry will ultimately be shown.

Due to using decimal degrees as the coordinate system, it effectively divides the surface of the earth into four quadrants, defined by the location of the origin at $(0, 0)$ and the respective signs of the latitude and longitude coordinates. Namely:

- North-East quadrant: origin in lower-left corner
  Coordinates: $(+lat, +lon)$
  Topic strings of the form: `_yy.yyy/_xxx.xxx/`
- North-West quadrant: origin in lower-right corner
  Coordinates: $(+lat, -lon)$
  Topic strings of the form: `_yy.yyy/-xxx.xxx/`
- South-East quadrant: origin in upper-left corner
  Coordinates: $(-lat, +lon)$
  Topic strings of the form: `-yy.yyy/_xxx.xxx/`
- North-East quadrant: origin in lower-left corner
  Coordinates: $(-lat, -lon)$
  Topic strings of the form: `-yy.yyy/-xxx.xxx/`

As the format of the topic strings varies in each quadrant, the starting state of the algorithm will be four large rectangles to cover the earth, with coordinates of the form:

$$(lat_1, lon_1), (lat_2, lon_2) = (0, 0), (\pm 100, \pm 1000)$$

The basic premise of the algorithm is as follows:

- Start with 4 quadrant rectangles that cover everything
- If the approximation is not accurate enough (based on the percent of coverage of the input geometry, or some other measure), pick a rectangle and subdivide into 10 equally-sized rectangles, splitting either vertically or horizontally as appropriate. This translates into replacing one topic subscription with 10 more-accurate subscriptions.
- Discard any rectangles that do not intersect the input
- Repeat, splitting larger rectangles into smaller ones, until arriving at the desired result

3.1 The Algorithm

<u>Define:</u>

- *Coverage Ratio*: the percentage of how much the inputted target shape geometry covers a particular rectangle, or also how much the target covers the union of all rectangles. This will be the accuracy of the approximation, and determines how many false positives will match; an 80% coverage ratio, or 0.8 will result in approximately 20% false positives, or 1 out of 5.
- *Child*: when a rectangle is split into 10 smaller rectangles, they are the children

- *Split*: the act of subdividing a larger rectangle into 10 equally sized rectangles, either vertically or horizontally

Constraints:

- *Subscriptions*: the number of subscriptions a topic-based pub/sub messaging system can maintain is not infinite. Having more subscriptions requires more work to be done by the system to match data. Hence, it may be necessary to limit the number of subscriptions (rectangles) generated as a result of the output, even if the desired coverage ratio is not achieved.

Inputs:

- T = target shape geometry that we are trying to approximate. The target is a multi-polygon: it can have holes and multiple pieces, but cannot be complex (i.e. no crossings)
- k = maximum number of rectangles allowed in the output, a terminating condition; an integer ≥ 4
- z = minimum coverage ratio desired for the output, i.e. the accuracy, a terminating condition; a scalar $[0, 1)$

For example: given a geometry T, compute an approximation to T, where T covers at least $z = 80\%$ of the approximation or consists of no more than $k = 50$ rectangles.

Computed Values and Outputs:

- $L = \{R_1, \ldots, R_n\}$ = list of all rectangles (the algorithm output)
- $n = |L|$ = number of rectangles in the list L
- R_i = the rectangle at position i in L; a geometry
- $S = \bigcup_{i=1}^{n} R_i$ = union of all rectangles; a geometry
- $A(T)$ = area of the target shape; a scalar
- $A(R_i)$ = area of rectangle at position i in L; a scalar
- $A(S)$ = area of the union of rectangles; a scalar
- $c(R_i)$ = coverage ratio of a rectangle; a scalar $[0, 1]$, given by ${}^{(A(R_i) \cap A(T))}/_{A(R_i)}$. That is, what percentage of the area of the rectangle is covered by the target:
  0 = no intersection
  1 = the rectangle is completely contained by the target
- $c(S)$ = coverage ratio of the union of rectangles; a scalar $[0, 1]$, given by ${}^{(A(S) \cap A(T))}/_{A(S)}$
- $R'_{i,j}$ = the j^{th} child rectangle of R_i after splitting, with j in $[1, 10]$; a geometry

Pseudocode:

```
L = [$R_1$,$R_2$,$R_3$,$R_4$]
while (c(S) < z && n < k) do {
  sort(L)
  split($R_1$) → [$R'_{1,1}$..$R'_{1,10}$]
  for each $R'_{1,j}$ (j in 1..10) {
    if ($R'_{1,j}$ ∩ T != ∅) {
      L = L + $R'_{1,j}$
    }
  }
}
```

That is:

- Start with four (large, completely covering) rectangles
- While the coverage area of the union of rectangles is less than the specified accuracy, and more rectangles are allowed…
- *Sort* the current list of rectangles by some metric
- *Split* the "worst" rectangle into 10 smaller child rectangles (either horizontally or vertically, as specified in Section 3.3)
- Add each one that intersects the target area to the list of candidate rectangles, and recalculate

3.2 *sort*() Function

A sorting function must be derived in order to determine the next rectangle to split. The exact function used can vary, and is left to the discretion of the implementer: different functions will produce results of varying quality. Below are examples of factors to consider when deciding to split a given rectangle R_i, and these can be combined and weighted together to produce a sorting function:

- Coverage ratio of R_i
- Number of potential child rectangles that would intersect the target shape after a split. Define this as $|d(R_i)|$. More formally:

  $$d(R_i) = \{x \in R'_{i,j} | x \cap T \neq \varnothing\} \tag{1}$$

  This metric is important as each rectangle corresponds to a topic subscription, and generally these are considered a "scarce" resource.
- Coverage ratio of each of the child rectangles. E.g. how many children have a coverage ratio of 1?
- Size of the rectangle: $A(R_i)$. E.g. prefer to split larger rectangles before smaller ones.
- Range between largest and smallest rectangles in the union. E.g. only want to have one order of magnitude difference between largest and smallest rectangles.

One example of a weighting for the sort function could be defined as (bigger numbers are more likely to split):

$$(1 - c(R_i)) \cdot \frac{A(R_i)}{1.2^{|d(R_i)|-1}} \tag{2}$$

That is, the "ratio inverse" of the coverage ratio (rectangles that are less covered are more likely to get split), times the area of the rectangle (split bigger ones first), divided by a weighting between 1 and 5.16 (1.2^0 to 1.2^9) which favours splitting rectangles with less children. The constant value of 1.2 was chosen through experimentation.

The reasons for defining the algorithm as above:

- Ratio: when comparing two rectangles of the same size and with the same number of possible resulting children, it is preferable to split the rectangle that is only 10% covered rather than one that is 60% covered, as this will improve the total coverage ratio more, while resulting in the same number of children (i.e. subscriptions)

- Area: if both the coverage ratios and the resulting number of children are the same, then splitting a larger sized rectangle will improve the total coverage ratio more, while resulting in the same number of children
- Number of Children: if the rectangles' sizes and coverage ratios are the same, it is preferable to choose to split the rectangle that has less children, as this corresponds to less topic subscriptions on the pub/sub system. Something more complex could have been used, such as calculating the coverage area of each child as well and incorporating that into the sort function.

Note that it is often possible to get a split "for free": when considering a rectangle to split, and the target area intersects only one of the resulting 10 children rectangles, that is $|d(R_i)| = 1$, then the rectangle can be split immediately, and replaced by the single child rectangle. This follows as the total number of rectangles doesn't change, and the coverage ratio improves.

3.3 *split*() Function

Similar to the sort function, different split functions can impact the final result of the algorithm. When it is decided to split a particular rectangle into 10 smaller child rectangles, the orientation in which to split the rectangle depends on a few factors. For this implementation this paper is based on, the split function is defined to follow the rules below:

- If the rectangle is horizontal / wide as in Figure 3, split it with vertical cuts into 10 square-shaped grids
- If the rectangle is vertical / tall, split it with horizontal cuts into 10 square-shaped grids
- If the rectangle is square-shaped, then the shape of the area defined by the intersection of the target and rectangle must be considered:

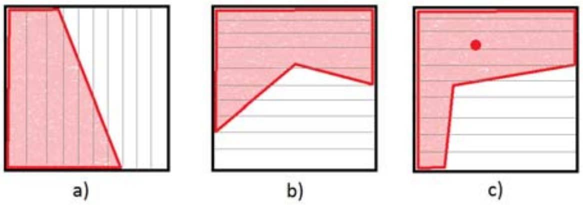


Fig. 5: Choosing a vertical or horizontal split orientation

a) If the intersecting shape has more of a vertical tendency (i.e. height > width), the square should be split with vertical cuts into 10 vertical rectangles due to the likelihood of being able to discard child rectangles from future consideration if they do not intersect the target.
b) Similarly, if the target shape is more horizontal, split the square with horizontal cuts into 10 horizontal rectangles.
c) Otherwise, another metric is employed, such as using the centroid of the intersection, or possibly splitting the square into 100 smaller squares (double split). Again, the exact implementation on how to choose which way to cut is left to the implementer.

After the split has occurred, the coverage areas for each of the 10 new children rectangles are calculated:

- If the child's coverage ratio is 0 (meaning the child doesn't intersect the target T at all), it therefore does not need to be considered, can be discarded, and this child is not added to the list L of rectangles
- If the child's coverage ratio is 1 (meaning it is completely covered), then it would make no improvement to the overall coverage ratio to split this child further. The sort function should disregard any such rectangles with coverage ratio = 1, yet it is still added to L.
- Else, the rectangle is partially covered by the target shape, and is added to the list of potential split candidates, with a weighting defined by the sorting function.

Iteratively perform the steps outlined in Section 3.1 until one of the two terminating conditions are met: either the number of subscriptions is reached, or the desired accuracy of the output is achieved.

3.4 Further Implementation Details

This section discusses some insights gleaned during the implementation of the algorithm.

As mentioned at the end of Section 2.2, the GPS coordinates within the topic string must be zero-padded, both before and after the decimal point. The amount of padding after the decimal must be at least the maximum number of digits in the topic subscription. That is, if the subscription had two decimal places (e.g. `_45.30*`), the topic must have at least two as well (e.g. otherwise `_45.3` would not match, as it requires the 0). Civilian GPS systems have accuracy around 3-4 metres, so using five or six decimal places of accuracy should be sufficient for most applications; this yields a resolution of approximately 1 metre or 10 cm respectively [5].

At the termination of the algorithm, the various rectangles that are generated as output must be converted into topic subscriptions for the filtering to occur in the message bus. As stated at the beginning of Section 3, the topic subscription format has different signs, whether + or −, depending on which quadrant (NE, NW, SE, or SW) the rectangle is in. The quadrant will determine the "inner corner" of the rectangle, or corner of the rectangle closest to the earth's origin $(0, 0)$, which is very useful when converting to the topic subscription.

The algorithm does not specify how to calculate if a particular rectangular area intersects the given input target geometry. A simple approach would be to treat the latitude/longitude coordinates as planar coordinates, and perform planar geometry intersection calculations. This approach would provide sufficient accuracy at small scales (e.g. urban centres), but very large polygon edges would not follow the great circle route. A more advanced option would be to use spherical geometry, or ensure the various areas are represented by geodesic polygons.

Care must be taken for polygons that traverse the International Date Line, and for polar contained areas.

4. Conclusion and Discussion

The use of this algorithm and technique has broad applicability across industry and government. It allows the filtering of location data to a specified area within the message bus layer, thereby reducing data bandwidth and CPU load on clients by not having to filter locally.

As mentioned in Section 1, this technique can be extended to cover ranged queries in other domains, as well as to more dimensions. For example, it would be easy to add a 3rd topic level to describe elevation, and could therefore be used for aviation geo-fencing.

Algorithm runtime complexity was not evaluated as part of this paper due to the rather small (e.g. low thousands) number of iterations through the main loop to produce very accurate approximations to the input geometry, even with a very complex shape with many vertices and edges.

This algorithm seems to work well for "boxy" or convex shapes, especially those that are more axis-aligned. The algorithm tends to require a larger number of subscriptions to approximate star-shaped polygons, or polygons that have long diagonal edges, as these require many "staircase steps" to approximate sufficiently.

One disadvantage to this approach is that it only works for matching point coordinates to polygons: it cannot do line-to-line intersection, or polygon-to-polygon matching.

For a working demonstration of this algorithm, please visit: http://london.solacesystems.com/aaron/geo/

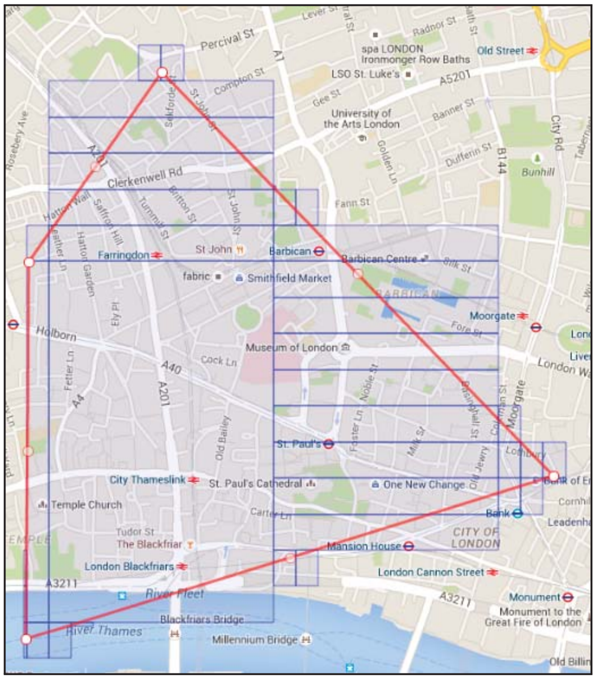


Fig. 6: 37 rectangles, corresponding to 37 subscriptions, approximating a target polygon with 75% accuracy
Background Map data ©2016 Google

References

[1] Wikipedia: "Message-oriented middleware", retrieved March 12, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Message-oriented_middleware
[2] Wikipedia: "Message broker", retrieved March 12, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Message_broker
[3] Wikipedia: "Publish-subscribe pattern", retrieved March 12, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Publish-subscribe_pattern
[4] Solace Systems: "Controlling Information Flow with Topics", May 2012. [Online]. Available: http://www.solacesystems.com/techblog/controlling-info-flow-with-topics
[5] Wikipedia: "Decimal degrees", retrieved March 12, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Decimal_degrees
[6] MQTT version 3.1.1, OASIS Standard, December 10, 2015. [Online]. Section 3.3.2.1, Topics: http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html

DIY VR: The Development of an Inexpensive Headset for Makers

Ronald P. Vullo, Ph.D.[1]**, Andrew Phelps, M.S.**[2]**, and Michelle A. Catalfamo B.S.**[3]
[1] Department of Information Sciences and Technologies,
RIT Center for Media, Arts, Games, Interaction & Creativity (MAGIC)
Golisano College of Computing and Information Sciences
[2]Director, RIT Center for Media, Arts, Games, Interaction & Creativity (MAGIC)
[3]Alumna, Motion Picture Science, Minor in Web Design and Development

Rochester Institute of Technology
Rochester, New York 14623

Abstract - *Virtual Reality looks like "the next big thing" but between the disposable Google Cardboard and the pricey Oculus Rift there is still much room for makers to experiment and students to learn the nuts and bolts (both literally and figuratively) of VR. This paper is a report on an ongoing project to develop an inexpensive, customizable VR headset kit appropriate for use in the classroom and beyond. The project's goals include using off-the-shelf materials, keeping costs low (current cost is approximately $25), and supporting student software and media development. Conference attendees will have hands-on access to assembled headsets running student-developed software, as well as being able to see the unassembled components and assembly instructions.*

Keywords: Virtual Reality, VR, DIY, headset, smartphone, Molly

1 Introduction

While Virtual Reality has been the big story in computing for the past year or so, it is not a new area of research and development. One of the authors (Vullo) began experimenting with Virtual Reality Markup Language (VRML) in the mid 1990s as a way of visualizing dentition and restorations in electronic dental records. It is, however, only recently, with the rapid development and immense popularity of smart phones, that the technology for truly immersive virtual reality experiences have become practical and inexpensive enough for use in the classroom as a student development environment. It is our belief that as this rapid development continues, VR will soon be a mainstream consumer technology. At the very least, VR is an interesting combination of technologies that engages students across multiple disciplines in learning and development.

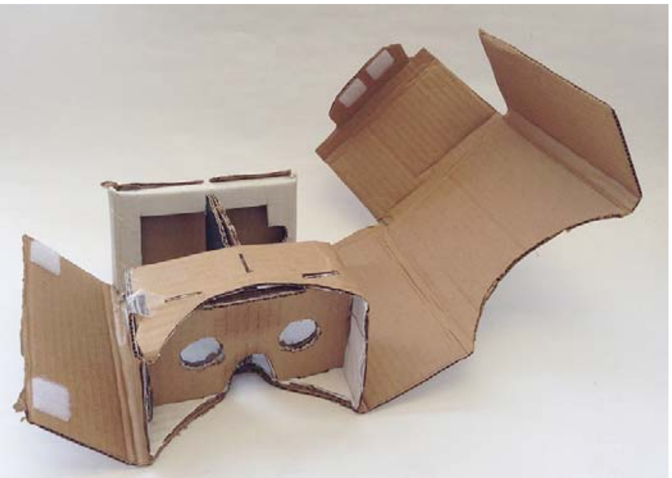
The goal of this project has been to develop a quality, yet inexpensive VR system design available to students and makers interested in exploring this exciting emerging field. Designed by Professor Vullo, it uses LASER cut plywood, acrylic, and fabric components making for an easy to assemble and customize headset. This approach allows both VR pros and hobbyists to gain easy insight into the hardware side of the VR equation.

2 A Series of Prototypes

This project has been one of successive prototypes and tests; experimenting with various materials, optics, phones, and software. The history of these is hopefully instructive.

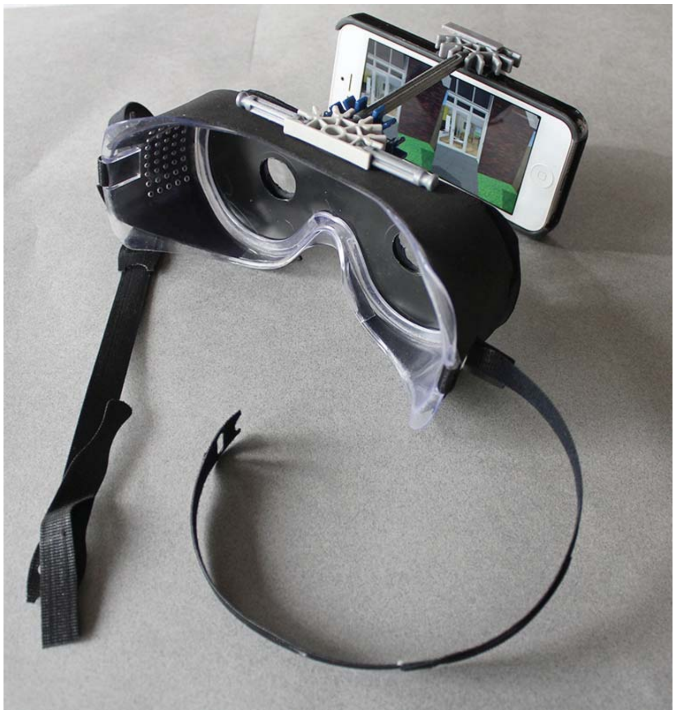
2.1 The Beginning

The Initial design was inspired by Google's "Cardboard" design, which was originally made for the Android platform. That design was modified to work with an available iPhone 5 and the lenses, which were on hand from a previous 3D photography project.



2.2 VR "Goggles"

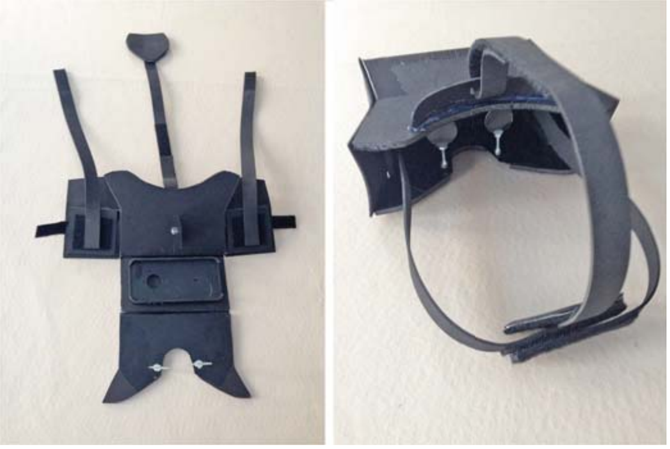
The initial prototype proved awkward and difficult to assemble, so for the second prototype the design was based on modifying an inexpensive pair of safety goggles.



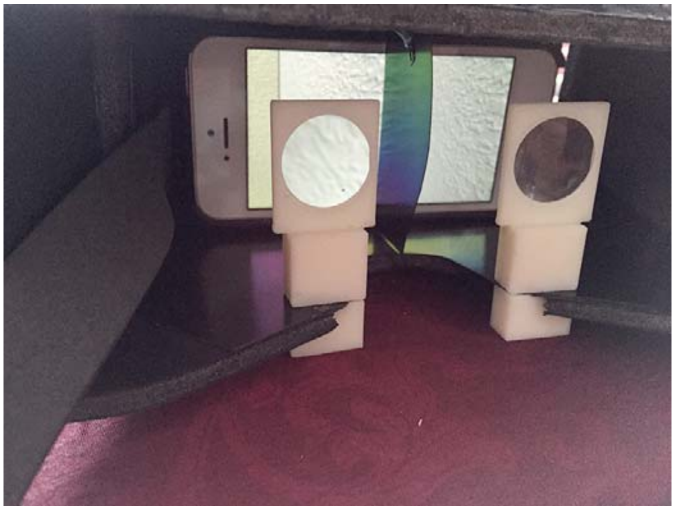
Better lenses, removed from a pair of binoculars, with a shorter focal length (about 3 1/4 inches) allowed for a smaller design and a better field of view in early tests. However, lens mounting diminished this field of view advantage. Development of a lightweight, yet rigid structure proved problematic, therefore this design approach was abandoned.

2.3 It's Not Cardboard

This prototype combined the box-shape of the original cardboard design, the face-hugging shape of the safety goggles, and a new material: foam core.



Foam core proved to be lighter and more rigid, as well as easier to cut and assemble with hinges of duct tape, compared to the cardboard design. Mounting the lenses, now extracted from inexpensive loupes, via foam core, then 3D printed mounts secured via rare earth magnets allowed for adjusting focus and inter-ocular distance of the lenses.

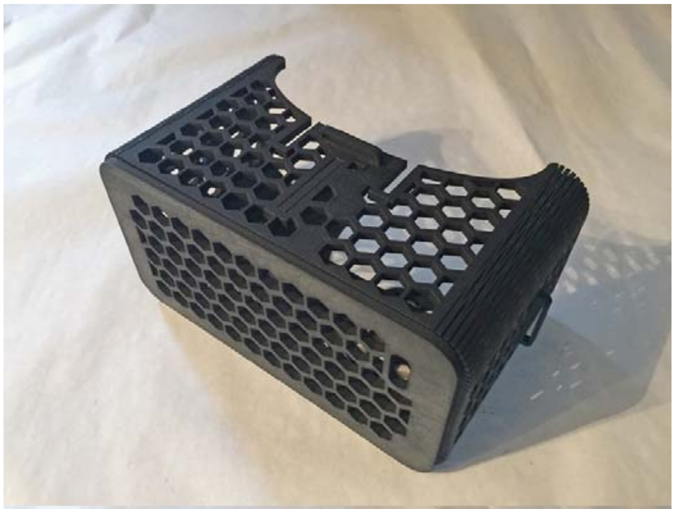


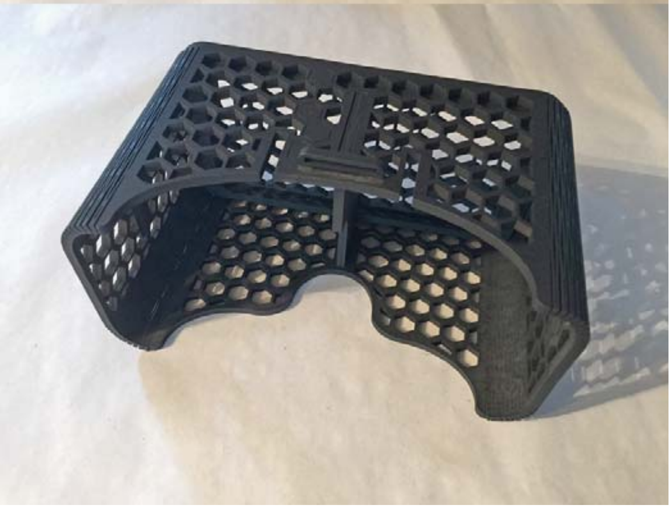
The availability of foam core with plastic rather than paper coating gave hope that this approach would result in a design that would be much less disposable than cardboard.

2.4 Burning Holes with Light

Hand cutting components from both cardboard and foam core proved slow, tedious, and imprecise. Taking advantage of a colleague's lab's new LASER cutter became the obvious solution. Cutting foam core on a LASER cutter is, at best, dangerous however, as the material tends to catch fire. Returning to cardboard, which is easily cut on a LASER cutter was not appealing, for other than quick throw-away tests, however the ability to cut thin inexpensive plywood was promising.

Plywood is considerably heavier than either cardboard or foam core, so this needed to be addressed. Taking a hint from honey bees, the wooden frame was honeycombed with hexagonal holes such that rigidity was maintained, but weight was greatly reduced. Instead of the square corners of earlier box-based designs, the ability to precisely cut accordion slits in the plywood made it flexible enough to bend into rounded corners.



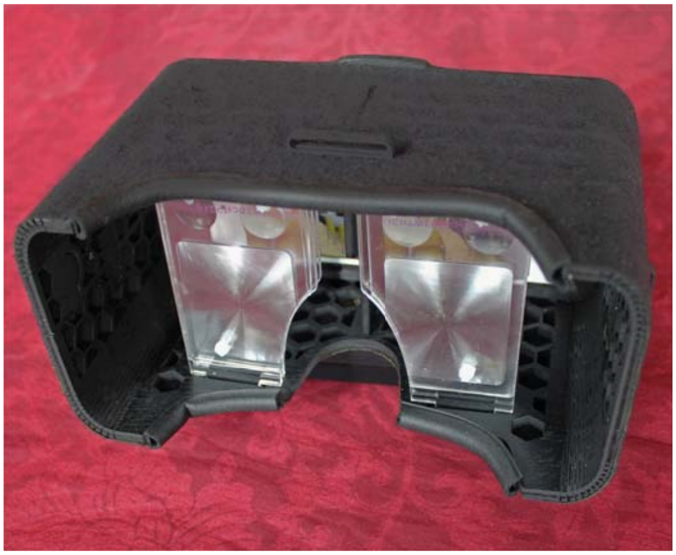


Honeycombing lets light into the headset, so LASER cut fabric is used to seal out the light.

Lens mounts for this design were LASER cut from acrylic, held together with rare earth magnets and fastened to the frame with wing bolts and nuts. While this design was adjustable, in the end it still was not optimum.

2.5 Pursuit of Optics

One of the most difficult aspects of this project has been the optics; specifically, identifying and adapting inexpensive and off-the-shelf optical components suitable to the task. Many lenses were tested, most of which are not detailed here. The current design uses what has thus far proven to be the best solution. For each eye we LASER cut a pair of acrylic 2X Fresnel magnifiers to provide space for the user's nose. We sandwich them between two LASER cut acrylic shields which slot into the headset allowing adjustable intraocular distance. This compound Fresnel lens design gives us excellent field of view and the nature of Fresnel magnification reduces the screen-door effect of the magnification of the phone's pixels as observed with traditional lenses.



3 Software

The other half of a Virtual Reality solution is the software. We have used existing software, primarily designed for Google's Cardboard, for testing during the development process. However, we have also used our own 3D images and begun developing 3D and VR content. It is worth noting that similar to VR hardware platforms, the software and engines to drive VR content is also in a state of rapid-evolution with a wide range of capabilities: low-end or casual-end solutions exist using basic web technologies such as MozVR or Three.js which utilize WebGL rendering techniques, and scale to professional 3D engine technology such as Unity, Unreal, and Cryengine, all of which are leading industry tools that output to multiple platforms including PCs, game consoles, mobile devices, and hand-held systems. RIT has recently joined the VRFirst initiative, and is working with hardware partners at Oculus, HTC, Microsoft, and others in support of exploring this range of devices and capabilities.

3.1 Non-VR 3D Content

Production of 3D images is an old and well known process, and allows for easy testing of the headset's optics. Similar, but somewhat more complex and labor intensive is the production of 3D video, which we (Vullo and Catalfamo) have begun experimenting with. This requires mounting and aiming a pair of video cameras, to which end we have built several twin camera tripod mounting adapters which have proven successful in early tests.

We have also experimented with mounting paired IP cameras to test use of the headsets for telepresence. Early experiments show promise, though there are some technological hurdles to overcome, primarily in combining two audio-video streams on a phone's screen.

3.2 Virtual Reality

As of the writing of this paper we have two VR software development projects underway. The first, is based in the School of Interactive Games and Media's Research Studio course (Phelps). This course allows students to work as domain specialists on teams completing one or more research and development projects. Typically, the faculty member teaching the class will provide the research themes, with the exact project goals and objectives emerging from these base ideas or pre-production materials. The goals of the course are to engage students with research methodology and production-level design theory to implement, test, deploy, and evaluate the results of digital-media projects. Students complete research reports and final assessments of themselves and their teammates, as well as the studio processes and communications paradigms of the team, in addition to completing their assigned responsibilities on the main projects. Recently this course has produced two well received commercial videogames: Splattershmup and Hack, Slash, and Backstab that have garnered numerous awards and recognitions in the professional community.

The second project (Vullo) is exploring the use of web technologies, specifically AFrame and Molly, to develop an experimental prototype VR social media environment. (Still in early stages as of the submission of this paper.)

4 References

http://diyvr.com/

https://www.google.com/get/cardboard/

http://magic.rit.edu/studios/hsb/

http://molly.rit.edu/

http://splattershmup.rit.edu

Vullo, Ronald P., Ph.D.; Molly: Simplifying Development of Complex Web Apps, invited presentation to the Rochester Joint Chapters of the IEEE Computer and Computational Intelligence Society, Rochester, New York (June, 2012)

Vullo, Ronald P., Ph.D., (December, 1998). The Future of Information Technologies In Dentistry. Invited Presentation at the International College of Dentists Annual New York Section Luncheon, New York, New York.

Vullo, Ronald P., Ph.D., (March, 1997). Does Your Reach Expand Your Grasp? Innovative Solutions: Extending the Educational Web. Uses of the Internet for Clinical Instruction Symposium Section, American Association of Dental Schools' Annual Session, Orlando Florida.

5 About the Authors

Dr. Vullo is Associate Professor, Department of Information Sciences and Technologies, creator and director of the Minor in Web Design and Development for non-computing Majors, and MAGIC Center faculty affiliate.

Mr. Phelps is Professor and Founder, School of Interactive Games and Media, and Director and Founder, MAGIC Center.

Miss Catalfamo is an Alumna of the Rochester Institute of Technology's Motion Picture Science program and a member of the Society of Motion Picture & Television Engineers (SMPTE). She also earned a Minor in Web Design and Development during which she was first exposed to the DIY VR project.

HomeAutomation
- Using OpenSource to fulfill EU-Directives -

Olaf Droegehorn
Harz University of Applied Sciences
38855 Wernigerode, Germany
odroegehorn@hs-harz.de

Jari Porras
Lappeenranta University of Technology
53850 Lappeenranta, Finnland
Jari.Porras@lut.fi

Fisayo Caleb Sangogboye
University of Southern Denmark
DK-5230 Odense M, Denmark
fsan@mmmi.sdu.dk

Abstract – Within the Kyoto protocol and the Paris agreement, as a follow up, the world's countries have agreed to limit global warming to a maximum of 2°C. The European Union, being an active member in these discussions, has passed directives to mitigate emissions from buildings that still constitute 36% for CO_2. In order to reach this goal home automation is a major element needed to be implemented for existing households. But looking to the existing market it becomes clear that none of the available vendors/solutions can cover a sufficient end-user scenario alone. Therefore open, integrating approaches have to be used to foster the uptake of home automation by end-users. This paper investigates the existing technologies and vendor solutions and points out several OpenSource solutions, which can be suitable to fulfill end-user scenarios and therefore to implement the EU-directives.

Keywords: EU-directives, world climate protocol, energy saving, building automation, OpenSource

1 Introduction

Today, energy conservation constitutes a major socio-economic discussion amongst committees of nations in the world. This is because the increasing greenhouse gas emission from industrial activities has led to significant depletion of the Ozone layer and an evident global climate change. One such energy conservation discussion, widely known as the Kyoto Protocol, was held in Kyoto, Japan in 1997. This discussion was held to extend the commitment of the 1992 United Nations Framework Convention on Climate Change (UNFCCC) of member states to reduce their greenhouse gases. This protocol instruments a common but differentiated responsibility on participant states by putting more obligations on developed states to reduce their current emissions based on the premise that they are historically responsible for the greenhouse gases (United Nations, 1998) [15]. The Kyoto protocol has been renewed in the 21st United Nations Framework Convention on Climate Change, already known as the Paris agreement, stating that the worlds countries will limit global warming at a maximum of 2°C, better at 1,5°C. To achieve this, the emission of greenhouse gases needs to be reduced to zero in the years between 2045 and 2060.

The European Union, being an active participant of this protocol/agreement, identified that buildings constitute 40% of the energy consumption and 36% of CO_2 emission in the EU and it is the largest end-user sector followed by transport (32%), industry (24%) and agriculture (2%). Thus two main legislations mitigating emissions from buildings were approved by member states (European Commission, 2015).

These legislations include:
1. Energy Performance of Building Directives (EPBD)
2. Energy Efficiency Directives (EED)

The EU directives serve as a framework for developing the national implementation strategies for member states to significantly reduce the energy consumption at both residential and public buildings. Also a concerted action was launched to enable the exchange of member states best practices about energy conservation among themselves. One of the strategies within this action is the installation of building automation systems.

The SMARTer 2030 report [2] given in the Global eSustainability Initiative identified that the installation of building management systems (smart systems) by occupants to automate building functions such as lighting, heating and cooling could offer a major opportunity to reduce the global CO_2 emissions of buildings by a ratio of 15% percent. Also according to the report, 42% of home energy expenditure comes from house conditioning, however much of this energy expenditure are often used for space conditioning when the home is unoccupied. It was highlighted in [3] that the installation of programmable devices could significantly mitigate energy wastefulness from negligent occupants and could save approximately 10 to 30% of their overall energy bills.

In the next chapter we scope the need and problem of home automation based on specific, vendor driven solutions. In section 3 we give an insight on the literature and on known works in the field. Chapter 4 is giving an overview of the general building automation infrastructure that might be needed to fulfill EU-directives. In chapter 5 an OpenSource approach for building automation is presented, highlighting several specific solutions and covering the open approach to interconnect different vendors. Before the paper closes with a conclusion a general discussion on possible implementation scenarios of EU-directives is given in section 6.

2 Motivation

Improving the performance of a building through investment in building automation is associated with significant costs. Results from observations and product research for residential homes indicate, that the investment cost of building automation ranges from 500 to several thousand Euros (depending on building type). Several authors have pointed towards the significant chain of environmental degradation (in terms of CO_2 and greenhouse gases reduction) such investments could mitigate and have highlighted the social impact and human consideration of these technologies (in terms of its inherent comfort and control), however:

1. According to the report in [3], it is still unproven and unclear how much these technologies could save in terms of energy and cost
2. There has been no sufficient economic justification for these investments based on any economic metrics (for instance investment return and payback time)

This paper aims to contrast vendor specific implementations and approaches against open-source solutions and to discuss the necessity of open scenarios for end-users.

3 Literature review and related work

The report given in [4] provides a summary of the energy usage for residential and non-residential buildings in EU states and a comprehensive analysis of how the effects of the economic, energy prices and occupant's behaviors affect energy usage. The analysis is based on the energy usage data and energy efficiency indicators provided by the ODYSSEE database and website. The energy usage in buildings may vary per country, however this consumption represents in average a total of 41% of the energy usage in the European Union (EU) and from this lot, residential buildings accounts for 65.9% of the total energy usage of EU buildings and 27% of the energy consumption in the EU. For Finland, Spain, Portugal, Cyprus, building energy usage represents 33.33% of their total energy usage while for Germany, Denmark, France, Poland, building energy usage represent 45% of the final energy consumption. Also, while the distribution of building energy consumption between residential and non-residential buildings may vary per country, the share for residential building from the total building consumption for Germany and Finland ranges between 60-70% and the annual consumption per (kWh/m^2) for these two countries are 210 and 325 respectively. This disparity is associated to climatic differences between the two countries, and therefore in Germany not so much energy is used overall but the percentage of space heating is larger due to the lack of other needs, as water heating of lighting. A breakdown of the energy consumption per household for both Finland and Germany in table I reveals that space heating represents the largest share of the total household energy usage.

Distribution	Germany (%)	Finland (%)
Space Heating	75	66,7
Water Heating	12	14
Electric Appliances and Lighting	12	19
Cooking	1	0,3

TABLE I: Distribution of building energy consumption per usage category

A comparison of the energy usage for space heating from the year 1990 to 2009 reveals a reduction trend for the EU average usage with a ratio of 30-60%. This reduction was attributed to the implementation of thermal regulations from EU countries for new buildings. However, the data provided by [5] for heat consumption per m^2 at normal climate conditions reveals that between the year 2000 and 2012, Germany recorded a 17.38% decrease in energy usage with figures 17.472koe/m^2 and 12.436koe/m^2 - respectively while Finland recorded a 2.18% increase with figures 15.583koe/m^2 and 15.923koe/m^2 - respectively. This implies a 21% energy usage difference for space heating for Finland and Germany for the year 2012.

Comparing the energy usage for electric appliances per dwelling for the year 2000 and 2012, the data given in [5] reveals that Germany recorded a slight 8.81% increase from 2078kWh to 2261kWh respectively and Finland recorded a significant 30.23% decrease from 4548kWh to 3173kWh respectively. This implies a 29% energy usage difference for electricity for Finland and Germany for the year 2012.

The ecoMOD project by the University of Virginia given in [6] entails the design, construction and evaluation of houses for energy efficiency. This project aims to achieve three objectives: academic, environmental, and social. To achieve energy monitoring, a monitoring system was installed to retrieve sensory and actuation data every second and stores them with timestamps. This monitoring system comprised of cost effective sensors that measure temperature, humidity, air quality, water flow, electric usage for appliances, carbon dioxide level and wind speed. Sensory and actuation data were retrieved through a wireless connection and these were stored on a remotely accessible database. A detailed data analysis was conducted on a 20 day stored data using a custom developed web data-analytical application software and the data analysis results indicates that the HVAC[1] and

[1] HVAC (heating, ventilation, and air conditioning) is the technology of indoor and vehicular environmental comfort. Its goal is to provide thermal comfort and acceptable indoor air quality. HVAC system design is a sub discipline of mechanical engineering, based on the principles of thermodynamics, fluid mechanics, and heat transfer.

water heating system constituted the larger portion of the energy consumption with both measuring 38% and 21% total energy consumption respectively. Also the result indicates a 50% and 45% reduction in the envisaged energy consumption of the building. The discrepancies between the envisaged consumption and the analysis result for the hot water heater and HVAC was not justified with measured data, however it correlated with the result of a similar study given by [2].

Utilizing various wired and wireless media approaches for implementing smart gateway architectures for home automation were extensively discussed in [7], [8], [9], and [10]. However, setting up an architectures doesn't necessarily provide a way how to implement that architecture. Although many technologies are available to connect

4 Design and implementation of a smart home scenario

In order to realize a smart home scenario, different levels of technology and abstraction need to be covered, as depicted in Figure 1. At the very first level, the Field-Layer, sensors and actuators need to be selected. Already here several different technology lines and hundreds of different sensors, measurement approaches, as well as actuators are available, even covering the same task with different approaches.

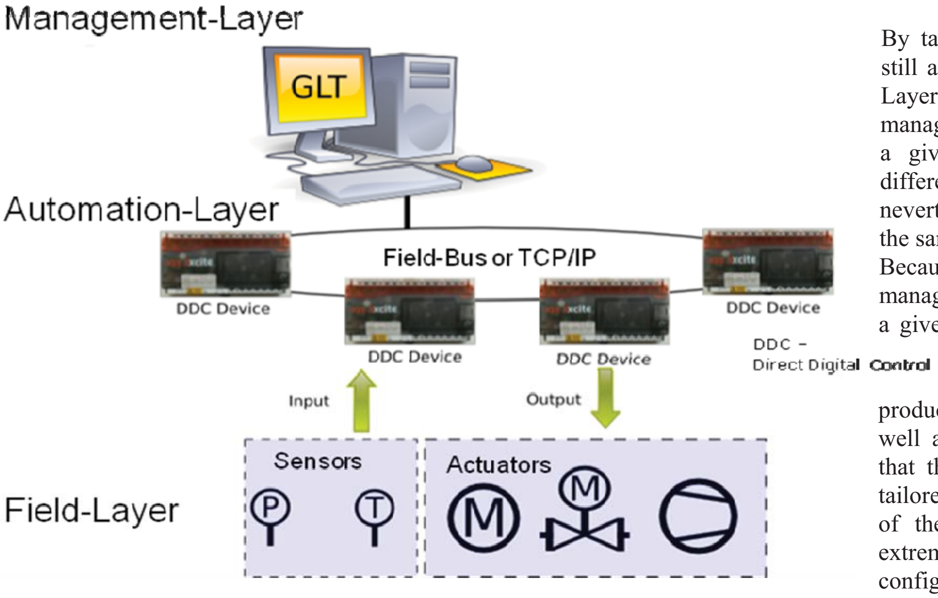


Figure 1: Architectural levels of Smart Home systems

Typically, in correspondence to the Field-Layer devices, the Automation-Layer is selected or already sold in combination. This is due to the fact that the sensors and actuators are using either a proprietary protocol towards their DDC device or are developed and produced by the same company. Therefore many different wired or wireless protocols are used between these two layers, offering a wide variety of possibilities to end-users. But at the same time this makes it necessary that end-users are dealing with the decision, which system or protocol they should use. Just to name the most prominent protocols/device families, the end-user can choose from:

X10, KNX/EIB, C.Bus, LONWORKS, ONE-NET, ZigBee, EnOcean, FS20, HomeMatic, OneWire, SCS Bus, EHS, 802.15.4, PowerLine, IP, InfraRed, RS485, etc.

A detailed comparison of different technologies has been done already [1]. The typical approach of these comparisons is based on performance aspects of the different protocols to bridge the gap between the Field-layer and the Automation-Layer. This might give end-users a guideline, which technology to choose, but if and only if the selected technology can cover all desired functions. E.g. the EnOcean technology, itself been quite popular due to the fact that it uses battery-less sensors and actuators, covers only the typical HVAC scenarios as it delivery actuators for switching lights and controlling heaters. Beyond this the EnOcean technology can't provide more functionality.

A similar observation can be made for several other technologies, which leads to two possible conclusions for the end-user:

1. Stay within the boundaries of a single technology, and therefore tailor the scenarios towards the available components.
2. Combine several technologies and therefore several protocol systems in order to cover the desired scenario

By taking those aspects into account still an approach for the Management Layer is needed, in order to be able to manage and control the devices within a given scenario. Also here many different possibilities arise; nevertheless the dilemma is basically the same as on the Automation Layer. Because in most cases a specific management software is delivered with a given set of devices, using a single technology and being tight to the components within this product line. This has advantages as well as disadvantages in such a way that the management solution can be tailored to the necessities and specifics of the selected products, making it extremely ease for end-users to configure the setup. And as the Management Layer software can hide a lot of complexity of the devices and their specifics the end-user is quite motivated to follow this approach. As long as this covers the needs of the scenario the only possible drawback of that approach may be an advanced price as such a system is typically sold as a fully integrated solutions. Companies like Siemens, Bosch and others [11] are offering these solutions with a fairly good user experience. But as soon as the end-user wants to integrate devices, based on different technologies or from different vendors, several aspects have to be taken into account. When combining different Field-Bus systems or even different Automation- or Field-Layer technologies the major question for end-users will be how to monitor and control all these devices.

Although it is possible to install different Management-Layer applications, each one dealing with its related lower layers, this leads to a very diverse set of control points and finally to a bad user experience, as the end-user has several front-ends and systems he needs to handle.

The second option might be to install gateways between different Automation-Layer components in such a way, that a single, dedicated management system is able to handle all devices. This might need a certain way of device simulation, as unknown devices for the management software need to be mapped to products being integrated in the specific software application. This works for a certain set of devices and related gateways, as long as a simulation/mapping towards existing elements in the management software can be made.

The third, and most suitable solution, would be to use a Management-Layer system that integrates the multitude of Automation- and Field-Layer systems in a single and manageable system. As this typically goes beyond the boundaries of a single vendor mainly OpenSource solutions are covering these requirements.

5 OpenSource systems for the Management-Layer

In order to handle a multitude of different Automation-Layer devices, and related Field-Layer elements, a vendor independent Management-Layer approach is needed. This is typically provided by OpenSource solutions, driven by a community that integrates different hardware-systems as needed. At the moment, several possible solutions are available, differing in many technology-related aspects as well as in the level of maturity.

One solution is the OpenHAB community and the related software solution [12]. The system consists of an OSGi core and a set of related libraries around that. This gives a good flexibility within the system and allows for several developers to integrate new functions without interfering with others.

Beyond this it gives a good abstraction from technology specific aspects and handles sensors and actuators in a common way. A general overview of the openHAB architecture is depicted in Figure 2, showing the layered approach of the system. The benefit of this architecture is its flexibility and the possibility to use different hardware configurations to interface to a multitude of different Automation-Layer and Field-Layer systems. Although this advantage makes the approach favorable it leads to a complex system with many components and needs a core platform with a powerful processor. As the system builds on OSGi an underlying JAVA engine is needed and therefore an additional virtual machine is used, which leads to quite some computational needs of the host system.

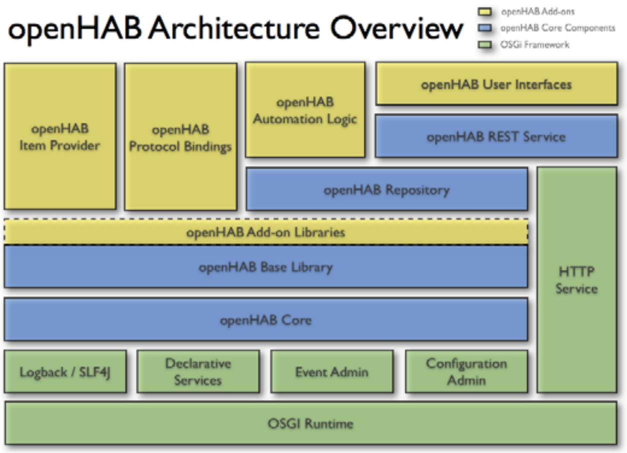


Figure 2: The openHAB architecture [12]

Another approach is the FHEM [13] system, a GPL'd Perl server for house automation. It is used to automate some common tasks in the household like switching lamps / shutters / heating / etc. and to log events like temperature / humidity / power consumption, etc. FHEM is also a community based solution, which adopted a flat architecture, using a single main core as an event loop supported by many modules for different tasks. Each module represents a certain device from the Automation- or Field-Layer and integrates the specific functions into the FHEM system. Beyond that so called helper-modules are used to provide additional functions, like web-servers, JSON/XML providers, etc.

Due to this simple architecture FHEM is more a mediating server, collecting all the different devices form the available modules and combines them within a common repository, which can be used and controlled by different front-ends. Figure 3 gives a rough overview how this approach is organized.

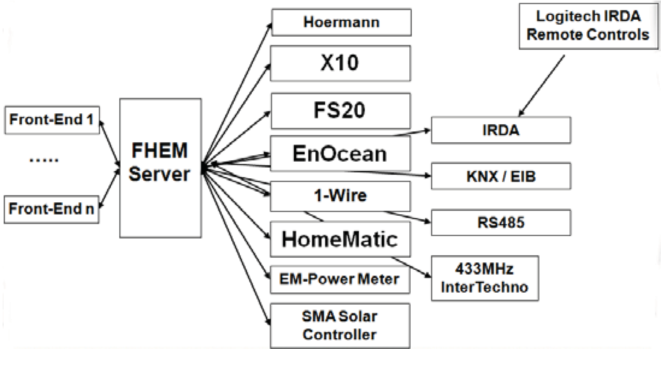


Figure 3: The FHEM Server infrastructure

Due to this reduced architecture and the fact that FHEM is written in PERL, a very basic language, being available on most devices, even embedded ones, the resources used by FHEM are minimal. Therefore it's common to see FHEM installation on embedded or integrated devices like NASsystems (network attached storage), media players or any other tiny board.

In accordance to FHEM another OpenSource project, called “culfw”, can be used for an open adaptation towards the Automation- and Field-Layer. The “culfw” project is developing a firmware for an open hardware-platform, typically based on ATMEL processors, which provides an interface to popular, although proprietary protocols, mainly used in Europe, like HomeMatic, FS20, etc. [14]. This open firmware project, in combination with one of the OpenSource Management-Layer solutions, leads to a powerful and open environment, covering nearly all available technologies for sensing, metering, and controlling appliances and therefore builds and excellent starting point to fulfill the given EU-directives.

6 Discussion

Although it is surely possible to fulfill the given EU-directives also without home automaion approaches, by building new, highly energy efficient houses, or even by a given single technology, it has been shown that common scenarios typically include more than one technology or vendor. In order to foster the uptake of home/building automation it doesn’t make too much sense to just concentrate on a single vendor or a specific technology. Like in the internet, a multitude of technologies might and should be used to fulfill the users needs and to enable the end-user to make choices as he wishes. Given the fact, that many low-priced proprietary solutions are already offered in different stores it should be taken into account, that end-users really have the choice on how much money they want to spent and which vendors they might want to choose.

Therefore an open solution at least for the Management-Layer is needed to incorporate different technologies and vendors.

At the moment ist is fulfilled in the best way by several open source soulutions, mentioned here, or elsewehre. The increased uptake of home automation is much more important than the market defvelopment for a single vendor, therefore open solutions should be favoured in the future.

That an open solution can reduce the individual energy consumption a lot has been shown already in [16].

7 Conclusion

Following the Paris agreement, as a follow up of the Kyoto protocol, the worlds countries want to limit global warming at a maximum of 2°C, better at 1,5°C. Thus two main legislations that mitigate emission from buildings were approved by european member states (European Commission, 2015).

In order to achieve this, several things have to change. It will for example be the way houses are built, but also the way existing buildings are used and energy is put into them. Therefore home/building automation might be used to adapt the energy usage towards the real users needs not to the average household. This is only meaningful, if end-users can build their own scenarios independent of the boundaries of specific vendors or technical protocols. Therefore, at least on the Management-Layer of home automation systems, an OpenSource approach should be used to integrate many given technologies and approaches to fulfill the end-users needs. Several possible solutions for integrating different vendor-specific approaches have been highlighted and therefore it should be monitored carefully how the parameters of the Paris agreement will be turned into reality in the future. For sure enough technology is available to make this happen and with a good sense of openness also the end-users can be taken on board.

8 References

[1] Chathura Withanage, “A comparison of the popular home automation technologies”, Dev. Pillar, Singapore Univ. of Technol. & Design, Singapore, Singapore, 2014 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA), pp. 600 – 605, ISSN: 2378-8534

[2] Global eSustainability Initiative, “GeSI SMARTer2030: Enabling the low carbon economy in the information age,” Global eSustainability Initiative, UK, [Online], Available: http://smarter2030.gesi.org/, accessed 04.2016

[3] Energy Star, “ENERGY STAR for Programmable Thermostats”, [Online]. Available: https://www.energystar.gov/products/heating_cooling/programmable_thermostats, accessed 04.2016

[4] Odyssee-Mure, “Energy Efficiency Trends in Buildings in the EU,” 2012. [Online]. Available:http://www.odyssee-mure.eu/publications/br/Buildingsbrochure-2012.pdf, accessed 02. March 2015

[5] Enerdata, March 2015. [Online]. Available: http://www.indicators.odyssee-mure.eu/onlineindicators.html, accessed 04.2016

[6] S. Foster, A. Tramba and L. MacDonald, “ecoMOD:Analyzing Energy Efficiency in Affordable Housing,” Charlottesville, VA, 2007.

[7] C. Wei and Y. Li, "Design of energy consumption monitoring and energy-saving management system of intelligent building based on the Internet of things," in Electronics, Communications and Control (ICECC), 2011 International Conference on, Zhejiang, 2011.

[8] J. Skon, O. Kauhanen and M. Kolehmainen, “Energy consumption and air quality monitoring system,” Adelaide, SA, 2011.

[9] C.-Y. Chen, Y.-P. Tsoul, S.-C. Liao and C.-T. Lin, “Implementing the design of smart home and achieving energy conservation,” Cardiff, Wales, 2009.

[10] D.-M. Han and J.-H. Lim, “Smart home energy management system using IEEE 802.15.4 and zigbee,” Kongju, South Korea, 2010.

[11] Siemens Home Automation: http://w1.siemens.ch/web/bt_ch/de/products_systems/building_comfort_hvac/home_and_building_automation/home_automation_system/Pages/home_automation_system.aspx, accessed, 2013

[12] openHAB reference, [Online] available: http://www.openhab.org/features/architecture.html, accessed, 04.2016

[13] FHEM online source, [Online] available: http://www.fhem.de, website of the FHEM project, accessed. 04.2016

[14] Culfw online resource, [Online] available: http:/www.culfw.de, website of the culfw project, accessed 04.2016

[15] KYOTO PROTOCOL TO THE UNITED NATIONS FRAMEWORK CONVENTION ON CLIMATE CHANGE, http://unfccc.int/resource/docs/convkp/kpeng.pdf, accessed 04.2016

[16] F. Sangogboye, O. Droegehorn, J. Porras: Analyzing the payback time of investments in building automation, International SEEDS conference: Sustainable Ecological Engineering Design for Society (SEEDS), 2015

A Collaborative Approach to Website Translation

Zongjie Tu and Chia-Yung Han
Department of Electrical Engineering and Computing Systems, University of Cincinnati, Ohio, USA

Abstract - *As more and more non-English speaking users begin to use the World Wide Web, translation services start to flourish. Current popular web translation services like Google Translate still suffer a few drawbacks inherent in machine translation and conventional client-server model. A collaborative approach is proposed in this paper. It can be adopted to aid both professionals and non-professionals in translating websites in a way similar to peer-to-peer data sharing. A web browser plugin backed by Distributed Hash Table was developed to enable web users frequenting the same website to perform in-place translation and share their works in real-time via WebRTC.*

Keywords: Web Translation, Internet Collaboration, P2P Network, Distributed Hash Table, Augmented Browsing

1 Introduction

Today's web browser is so versatile that it forms an integral part of daily life. Thanks to a plethora of web applications, web users can now read news, check emails, manage bank accounts, book air tickets, get driving directions, watch video clips, chat with friends and deal with other routines, all within the same browser window. As the web has started to flourish in countries where English is not the first language, the growth of non-English websites is gaining momentum. The percentage of native non-English speakers on the Internet has reached 74.1% [1]. The resultant linguistic diversity has stimulated demand for web translation, bringing into being online translation services backed by natural language processing and machine translation.

Nowadays, web users can readily take advantage of an online translation service whenever they encounter a web page in a language other than their own. Upon request, the service in question is able to translate a portion of or the entire web page in place within seconds. Conventional web translation is, for the most part, based on client-server model and machine translation. A web user can install a browser extension from a translation service provider and trigger translation on any web page by clicking a "Translate" button or menu item that is embedded into a browser. Although the translation quality may not be as high as that of a professional human translator, web users can at least get a general idea of what the web page in question is about.

At present, Google Translate [2] proves by far the most successful translation service on the web. Google Translate is integrated into various Google products such as Google Chrome. It also provides a widget called "Website Translator" that can be easily plugged into any web page [20]. For occasional use, the response time is totally acceptable and the translation quality is usually sufficient for a rudimentary grasp of general ideas. It supports translation between more than 100 languages, and it is free of charge. It is a versatile tool that can be integrated into a website by a webmaster, incorporated into a web browser by a software developer, or accessed interactively on Google Translate website by a web user.

None the less, many potential problems can arise from existing web translation services. For instance, data privacy and security may loom large in a client-server model. Not to mention that centralized services run by for-profit organizations are not guaranteed to be infallible or ever-lasting.

In the case of Google Translate, user data need to be sent to Google's centralized servers. How those data are handled is at the discretion of Google, as pointed out in Google's Terms of Service [21]:

> *"...When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content."*

That said, data privacy is not guaranteed in all Google's services, including Google Translate. "Free lunch" as Google Translate appears at first glance, users pay a hidden cost to Google when using its service. In addition, centralized data in a client-server model pose a risk to data security. Although Google has so far kept an almost impeccable record with respect to data security, millions of users could be affected if a data breach would happen to any one of Google's datacenters.

Google Translate may be discontinued at any time in Google's sole discretion. Despite the numerous free services

Google offers online, Google is a for-profit corporation after all. It would not hesitate to shut down any service that is unlikely to generate revenue, no matter how popular it has become. For example, Google Reader [22] was discontinued in 2013 amid objection from over 150,000 users [23], for the mere sake of declined usage [24]. There is no guarantee that Google will not do the same to Google Translate. Furthermore, Google Translate may be inaccessible at times. In spite of the multitude of servers Google possesses, Google, as a service provider, is not infallible. For example, Google cached links [25] can occasionally return invalid results [32] Likewise, it is entirely possible that Google Translate becomes temporarily inaccessible.

Language translation is a challenging task. Machine translation is far from perfect. Under many scenarios, only with a human translator that understands the context and has the pertinent background knowledge can accurate and elegant translation be had. Our research goal is to develop a framework to avoid relying solely on server-centered solutions and offer the capacity to incorporate human collaboration into the translation endeavor.

2 Background

The success of web applications can be attributed to cross-platform compatibility, server-centric software maintenance and ubiquity of wireless Internet, among others. State-of-the-art web technologies have even made it possible to achieve inter-browser communication, laying the foundation for any prospective browser-based peer-to-peer network. The collaborative approach we are proposing has been made possible by advances in several technologies. They include Distributed Hash Table (DHT), WebRTC, Node.js and userscript.

2.1 Distributed Hash Table

With a view to boosting system robustness and avoiding copyright lawsuits, a growing number of peer-to-peer applications, including eMule and BitTorrent, begin adopting serverless approaches to peer discovery, Distributed Hash Table [4] in particular. A Distributed Hash Table is a distributed system that resembles a hash table, providing key-value storage and lookup service [3]. Since 2001, numerous DHT implementations have been proposed [3][5][6][7][8][9]. Yet only a few, most notably Kademlia [8] and its variant Mainline DHT [9], have gained popularity in practice. In essence, Kademlia is a structured overlay network [10] built atop the User Datagram Protocol (UDP). A node in the overlay network reaches another via a logical link, which may correspond to a complex path in the physical network.

One striking characteristic of Kademlia is its unified ID space for both nodes and files. Each node and file is represented by an unsigned 160-bit binary ID, respectively. In a naive implementation, a node ID can be obtained by generating a 160-bit random number. A file ID is its SHA-1 hash [11], which is also a 160-bit number. Thus, both node and file can be located by looking up, in an iterative style, distance between two IDs, which can be obtained with an exclusive-or (XOR) operation

$$\text{Distance}(A, B) = A \oplus B\,. \tag{1}$$

The closer a pair of nodes or files are to each other, the less the unsigned value obtained from the exclusive-or operation.

2.2 WebRTC

WebRTC [12] is designed for direct or server-assisted multimedia communication between browsers. It is specially tailored for real-time communications, such as voice calling or video chat, yet is also capable of handling exchange of conventional data, such as peer-to-peer file sharing. WebRTC is currently natively supported by major PC browsers like Google Chrome, Firefox and Opera, as well as the most popular mobile platform, namely, Android. [13]

2.3 Node.js server-side web applications

Conventional web applications such as Gmail and Microsoft Office Online rely on JavaScript [18] for client-side functionality. The server-side code is usually not written in JavaScript. With the advent of Node.js, however, it is now possible to adopt JavaScript as the sole programming language for both client and server. Thanks to enormous advantages such as non-blocking I/O, fast execution and high scalability, Node.js is pervading the Internet industry [15] at an astonishing rate. An extra benefit of Node.js is that most server-side applications written in Node.js can be easily converted to client-side applications via Browserify [16], which is a Node.js application itself. Node.js finds valuable application in real-time collaboration such as peer-to-peer video streaming [17].

2.4 Userscript

A userscript [19], or user script, is a client-side browser plugin written in a scripting language such as JavaScript. It is able to modify a web page on the fly when executed in a web browser. Any web user having a good command of JavaScript can write a userscript to add a function or customize the appearance of a web page.

3 Problem statement

The proposed system is aimed at overcoming various drawbacks plaguing conventional web translation by exploiting creativity of the general public. We provide below a brief list of drawbacks that have hampered the effectiveness of conventional web translation services.

- Sensitivity to malicious attacks. For example, when Google Translate detects a burst of high traffic from a frequent, it may display a CAPTCHA for the web user to solve, thus disrupting the translation workflow.

- Inaccurate translation. As machine translation is mainly based on grammatical rules, the translated text may fail to convey idiomatic meaning, and worst of all, it may lack context awareness. For example, “crane” on a zoo’s website may be misinterpreted as a machine instead of a bird.
- Limited user collaboration. Direct collaboration among the contributors is not allowed in general. For example, all user contributions to Google Translate have to be routed through its Google Translate Community [26].
- Restrictions on submitted translation. Translation is subject to approval in Google’s sole discretion and not available for immediate use.
- Lack of personalization and user preference.
- Inadequate support for multilingual web pages. Multiple languages mixed together on the same web page are not well handled as users can only specify one source language.
- Lack of support for minority languages and dialects. Google Translate supports some 100 languages. [27] However, given that there are over 5,000 languages spoken worldwide [28], Google Translate still lacks support for minority languages, let alone dialects.
- Overdependence on English as a pivot language. English is often used as an intermediate language between two other languages, leading to distortion in meaning.
- Poor support for ever-evolving Internet slang.
- Romanized text of languages such as Chinese is not well supported due to homophone.

There is a need to address these issues. Ultimately, it is desirable to propose a system that has a long-term goal of bridging cultural gaps in the realm of Internet, bringing together netizens with similar interests around the world.

4 Proposed solution: DHT-based collaborative website translation

Web users need to be assembled to perform translation on a website they all visit. The key issue is how to efficiently get a collaborative environment going in a distributed system.

4.1 Kademlia as the DHT backend

We have chosen Kademlia as the DHT backend because of benefits stemming from its XOR metric. Other contributing factors include agile routing and trust in long-standing nodes. The following list briefly summarizes advantages that Kademlia enjoys.

1. XOR metric is simple and fast to compute.
2. XOR metric is symmetric and hence load-balanced.
3. XOR operation along with all 160-bit binary numbers form an abelian group, thus allowing closed analysis.
4. Parallel and asynchronous binary-search routing yields high performance.
5. Preference for long-standing node in the routing table maximizes the probability that those nodes will remain online in the next hour. [8]

4.2 System architecture

The architecture of the proposed system consists of the following components: 1) user, 2) the Graphical User Interface for Translation, 3) volatile memory, 4) non-volatile memory, 5) an optional signaling server for managing sessions and metadata, 6) an optional STUN (Session Traversal Utilities for NAT) server for establishing connections [14], and 7) an optional TURN (Traversal Using Relays around NAT) server for relaying data [14]. Fig. 1 shows the diagram for the system architecture.

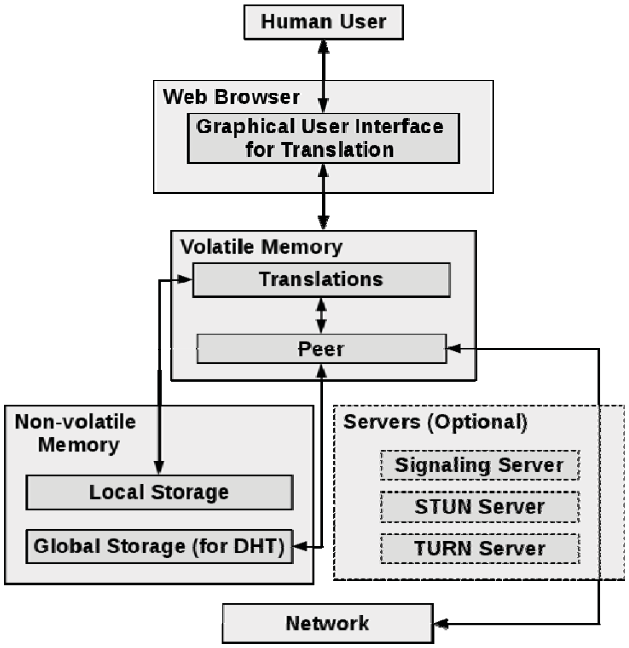


Fig. 1: Architecture for the Proposed System

A user is defined as a human user interacting with the proposed system. The user may contribute or retrieve translation through the web browser. There are two types of users, non-contributors and contributors. The former represents any general web user accessing websites displayed in foreign languages, whereas the latter are constituted of both amateur and professional translators.

“Graphical User Interface for Translation” (“Translation GUI” in short) in Fig. 1 is integrated into the web browser. The user can modify any web page in a WYSIWYG (i.e., What You See Is What You Get) manner. “Translation GUI” can be triggered simply by hovering mouse cursor over an eligible web element, such as a hyperlink. Once that happens, the element in question will be highlighted (e.g., surrounded by dashed line) and a dialog will pop up, allowing the user to perform translation-related operations.

“Volatile Memory” is typically made up of random-access memory (RAM) of computer systems and is chiefly used to accommodate “Translations” and “Peer” in Fig. 1. “Translations” is a data structure that stores all recently accessed data relevant to the current web page. It serves as a cache for all data read from or to be written to “Local Storage” and always holds the most up-to-date and hence authoritative information concerning a user’s interaction with the proposed system. No module in the proposed system other than “Translations” is allowed to interact with “Local Storage” directly. Data cached in translations need to be written to “Local Storage” from time to time.

In a BitTorrent network, a peer implements BitTorrent protocol and listens on a TCP (i.e., Transmission Control Protocol) port whereas a node implements DHT protocol and listens on a UDP port. In that context, a peer controls a node. As there is no entity that listens on a TCP port in the proposed system, however, a peer in the proposed system is equivalent to a node in the overlay network and the two terms can be used interchangeably.

Be default, either hard disk drive (HDD) or solid-state drives (SSD) is being used as “Non-volatile Memory”. “Non-volatile Memory” comprises two parts, namely, “Local Storage” and “Global Storage”. “Local Storage” stores user data (e.g., translations a user has chosen). Each hostname has its own allocated portion of “Local Storage”, since the same phrase might be interpreted quite differently on a different website. This helps resolve issues of polysemy and context-unawareness in conventional web translation. “Global Storage” is an integral part of DHT. It is the portion of storage “Peer” shares with its counterparts in the overlay network. “Global Storage” can also used to store routing tables for DHT.

In WebRTC, signaling is the process of exchanging media session description, which specifies the transport information along with media type, format and all associated media configuration parameters necessary to establish the media path. [29] Theoretically, WebRTC does not require a signaling server. Two users can simply exchange all forms of metadata by copying/pasting plain text from/to emails. But that is rather inconvenient as the signaling process is no longer transparent to the user. A signaling server that automates that process is therefore a rational choice. In the proposed system, we developed a custom signaling server to keep track of recently seen peers and hence facilitate bootstrapping for any new peer.

STUN and TURN servers can also be plugged into the proposed system for the purpose of NAT Traversal, if needed.

4.2.1 Peer interaction with DHT

The typical process of interactions between “Peer” and DHT in the proposed system can be described as follows.

1. Peer representing a monolingual user (i.e., a non-contributor) -- grabs published translations from DHT and uses one published translation (probably at random).
2. Peer representing a bilingual/multilingual user (i.e., a potential contributor) -- grabs published translations from DHT; screens the published translations (e.g., apply blacklist) and either uses one published translation if at least one of the published is satisfactory or creates a new translation (and uses it thereafter) if none of the published is satisfactory; then publishes the screened translations (possibly along with a new translation) to DHT.

4.2.2 Data format

User data are stored in JSON format [33] in both local storage and global storage.

1. Key format and value format for local storage

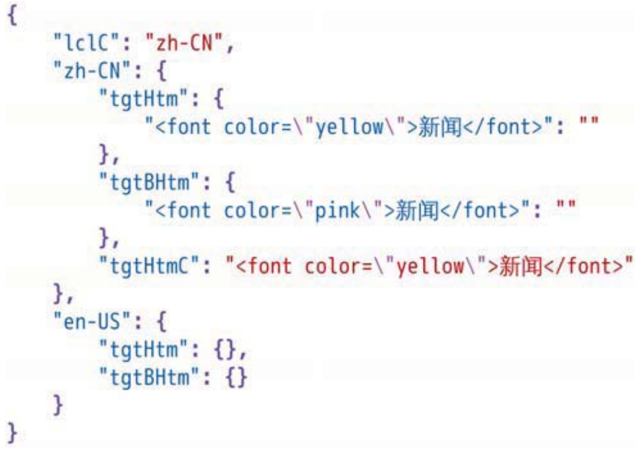
The key format for local storage is a fixed key prefix followed by source HTML (srcHtm in short). A key prefix is necessary since local storage items can reside in the same namespace as that of global storage. For example, the key for “http://www.cnn.com” with the key prefix being “prefixLS_” is “prefixLS_http://www.cnn.com”.

The value format for local storage is a collection of associative arrays that mingles a variety of data. An example of the value format for local storage is shown in Fig. 2. In Fig. 2, the chosen locale [31] (lclC in short) is “Simplified Chinese” whereas the chosen target HTML (tgtHtmC in short) is the “yellow” member. Note that the double quotes in member values have been escaped with the backslash (“\”). It is worth mentioning that a locale can contain empty fields, as is the case with tgtHtm and tgtBHtm (denoting blacklisted target HTML) of “en-US” in Fig. 2.

2. Key format and value format for global storage

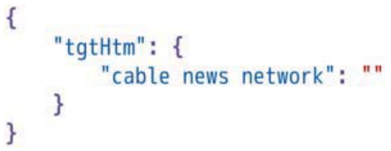
The key format for global storage is a fixed key prefix, followed by the concatenation of the locale to be synchronized with DHT and srcHtm. For example, if the key prefix is “prefixGS_” and we need to store American-English translations for the hyperlink “http://www.cnn.com/”, the key will be “prefixGS_en-UShttp://www.cnn.com”. When we need to query existing American-English translations for the hyperlink “http://www.cnn.com/” in DHT, however, we must strip the key prefix and then send a lookup request into the overlay network with the key being “en-UShttp://www.cnn.com”. It is worth pointing out that only the SHA-1 hash of the key is sent, thus concealing the content for which the user is seeking translation and hence protecting privacy.

The value format for global storage is what formulates the response from the overlay network. It is practically composed of the latest tgtHtm in DHT. The value contained in response may look like Fig. 3. The value in Fig. 3 has only one member, namely, “cable news network”, which is a proper noun that is not capitalized. Suppose we are dissatisfied with this all-lowercase proper noun and decides to replace it with “Cable News Network”. What we do is adding (i.e., saving) “Cable News Network” to tgtHtm and removing (i.e., blacklisting) “cable news network”. We subsequently send an insertion request to the overlay network, with the key being “en-UShttp://www.cnn.com” and the value being that in Fig. 4.



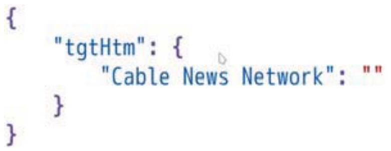
```
{
    "lclC": "zh-CN",
    "zh-CN": {
        "tgtHtm": {
            "<font color=\"yellow\">新闻</font>": ""
        },
        "tgtBHtm": {
            "<font color=\"pink\">新闻</font>": ""
        },
        "tgtHtmC": "<font color=\"yellow\">新闻</font>"
    },
    "en-US": {
        "tgtHtm": {},
        "tgtBHtm": {}
    }
}
```

Fig. 2: An Example of Value Format in Local Storage



```
{
    "tgtHtm": {
        "cable news network": ""
    }
}
```

Fig. 3: A Sample Response for Lookup of Key “en-UShttp://www.cnn.com”



```
{
    "tgtHtm": {
        "Cable News Network": ""
    }
}
```

Fig. 4: A Sample Request for Insertion of Value “Cable News Network”

4.2.3 Robustness

Peer is able to retrieve/store data from/to local storage even when it is disconnected from the overlay network (i.e., unable to reach any of its counterparts in the overlay network).

4.2.4 Customizable text style

Users are able to customize the look and feel of the translated text, usually via HTML and CSS, as exemplified in Fig. 2, where the attribute of color can be either yellow or pink.

4.2.5 Real-time translation update

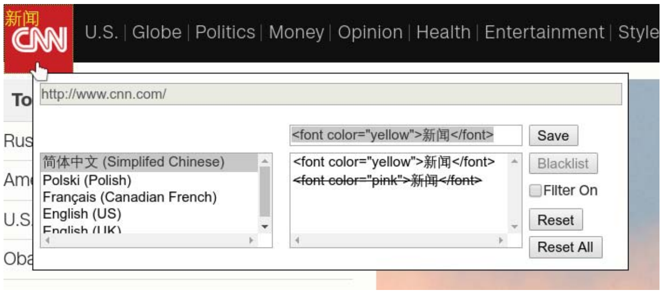
Peer checks periodically for translation updates from its counterparts with a view to populating them in the translation GUI in a timely fashion. It also publishes its own translations periodically to DHT. All updates are real-time as the period between two consecutive updates is merely a few seconds.

4.2.6 Offline mode

Although the overlay network is transparent to a novice user, connection to the overlay network can be intentionally disabled by an advanced user. This feature, what we call offline mode, is intended to further enhance data privacy.

5 A prototype of the proposed system

So far we have implemented a prototype that can function across browsers on different computers belonging to the same LAN. The prototype consists of a signaling server and a userscript converted from Node.js server-side code via Browserify. All code was written in JavaScript, including open-source third-party libraries such as KadTools [30]. Fig. 5 shows a screenshot of translation GUI in action. The translation GUI consists of several components, which are listed below.



http://www.cnn.com/

<font color="yellow">新闻</font>

简体中文 (Simplifed Chinese)
Polski (Polish)
Français (Canadian French)
English (US)
English (UK)

<font color="yellow">新闻</font>
~~<font color="pink">新闻</font>~~

新闻
CNN
U.S. | Globe | Politics | Money | Opinion | Health | Entertainment | Style
To
Rus
Am
U.S
Oba
Save
Blacklist
□Filter On
Reset
Reset All

Fig. 5: Screenshot of integrated translation GUI on a web page

1. Source HTML

In Fig. 5, the mouse cursor is placed over a hyperlink, i.e., “http://www.cnn.com/”. That hyperlink, which corresponds to the source HTML (srcHtm), indicates the source of translation.

2. Chosen locale

A chosen locale (lclC) specifies a user’s native or preferred language. It is composed of a language identifier and a region identifier, joined with a hyphen or underscore. For example, “en-US” denotes English spoken in the United States, or American English. On the Internet, however, it should be converted to some human-readable format such as “English (US)”. The lclC in Fig. 5 is “Simplified Chinese”. When the user resets translation of a web element (e.g., a hyperlink) via the reset button, lclC will be set to “-----”, signifying that the user intentionally opts not to translate that web element.

3. Chosen Target HTML

The chosen target HTML (tgtHtmC) is the translation a user has chosen. By “chosen”, we mean that the translation is either picked from a list of existing translations grabbed from DHT, which we refer to as target HTML (tgtHtm), or directly input by the user via keyboard and mouse. Translated text may not necessarily be plain text. It can be any valid HTML. The tgtHtmC in Fig. 5, i.e., “<font color="yellow">新闻</font>”[1], specifies that the translated text should be displayed in yellow. The tgtHtm in Fig. 5 used to have two valid members, namely, “<font color="yellow">新闻</font>” and “<font color="pink">新闻</font>”, displayed as a list below the input field (next to the save button) holding tgtHtmC. However, the “pink” member is shown in strikethrough format, implying that it has already been blacklisted via the blacklist button.

4. Save button

The save button saves tgtHtmC to translations (in volatile memory) and local storage (in non-volatile memory), and synchronizes changes, if any, with DHT. After the user presses the save button, the translation will be applied to the current page immediately, as can be seen in the upper-left corner of Fig. 5.

5. Blacklist button

The blacklist button is used to blacklist a member of tgtHtm that the user dislikes. It also moves that blacklisted member from tgtHtm to tgtBHtm stored in both translations data structure and local storage. Although the blacklisted member could be loaded again into translations data structure during next synchronization with DHT, that member would not be shown to the user because it is deprecated (i.e., blacklisted) according to tgtBHtm.

6. Filter-on checkbox

The filter-on checkbox can be unticked to show otherwise hidden tgtBHtm members so that blacklist decisions can be reversed via save button. In other words, it enables a user to move members of tgtBHtm back to tgtHtm (i.e., undo a blacklist operation). When the filter-on checkbox is unticked, tgtBHtm members are shown in strikethrough format.

7. Reset button

The reset button forces the proposed system to “forget” the fact that the current srcHtm has been translated. It restores any modified text of the current srcHtm back to its original state as if no translation had ever been performed.

8. Reset-all button

The reset-all button revokes all changes the proposed system has made to the current website, deleting all related data in local storage.

After collaborative translation is complete, the web page in question exhibits translated text to each collaborator, as is illustrated in Fig. 6.

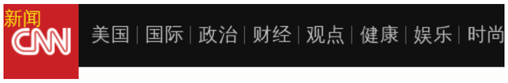


Fig. 6: Website Menu Bar in Fig. 5 after Being Collaboratively Translated into Simplified Chinese

6 Conclusions

Machine translation plays a vital role in modern web translation, saving huge amounts of human labor every day. Nevertheless refined translation entails collaboration between smart minds. Userscript lends itself to on-the-fly conversion of a foreign web page into one’s native language, and works even better when combined with the real-time inter-browser communication framework called WebRTC. We came up with an integrated solution fusing various state-of-the-art technologies to cater to the needs of translators as well as all other people curious about foreign cultures on the web.

The prototype we developed works as expected, yet it remains primitive and simplistic with respect to functionality and usability. For instance, it works with structured hyperlinks but does not work yet with unstructured plain text, although the proposed system does accommodate the latter case. We plan to further enhance the prototype and conduct a wide range of experiments on it until it reaches a full-fledged state.

7 References

[1] Miniwatts Marketing Group, “Top Ten Languages Used in the Web - November 30, 2015”, http://www.internetworldstats.com/stats7.htm.

[2] Google Inc., “Google Translate”, https://translate.google.com/.

[3] Ratnasamy, S., Francis, P., Handley, M., Karp, R. and Shenker, S., 2001, “A scalable content-addressable network”, Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '01), pp. 161-172.

[4] Wehrle, K., Götz, S. and Rieche, S., 2005, “7. distributed hash tables”, Springer Berlin Heidelberg.

[5] Stoica, I., Morris, R., Karger, D., Kaashoek, M.F. and Balakrishnan, H., 2001, “Chord: A scalable peer-to-peer

[1] “新闻” means “news” and hence is not an exact match for “CNN” but a personalized translation.

lookup service for internet applications”, ACM SIGCOMM Computer Communication Review, 31(4), pp. 149-160.

[6] Rowstron, A. and Druschel, P., 2001, “Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems”, Middleware 2001, pp. 329-350.

[7] Zhao, B.Y., Huang, L., Stribling, J., Rhea, S.C., Joseph, A.D. and Kubiatowicz, J.D., 2004, “Tapestry: A resilient global-scale overlay for service deployment”, IEEE Journal on Selected Areas in Communications, 22(1), pp. 41-53.

[8] Maymounko, P., Mazières, D., 2002, “Kademlia: A Peer-to-Peer Information System Based on the XOR Metric”, Peer-to-Peer Systems, pp. 53-65.

[9] BitTorrent.org, “DHT Protocol”, http://www.bittorrent.org/beps/bep_0005.html.

[10] Lua, E.K., Crowcroft, J., Pias, M., Sharma, R. and Lim, S., 2005, “A survey and comparison of peer-to-peer overlay network schemes”, Communications Surveys & Tutorials, 7(2), pp. 72-93.

[11] Information Technology Laboratory, “Secure Hash Standard (SHS)”, http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf.

[12] Johnston, A.B. and Burnett, D.C., 2012, “WebRTC: APIs and RTCWEB protocols of the HTML5 real-time web”, Digital Codex LLC.

[13] The WebRTC project authors, “WebRTC Home”, https://webrtc.org/.

[14] Dutton, S., “WebRTC in the real world: STUN, TURN and signaling”, http://www.html5rocks.com/en/tutorials/webrtc/infrastructure/.

[15] Node.js, “Node.js Foundation Members”, https://nodejs.org/en/foundation/members/.

[16] James Halliday, “Browserify”, http://browserify.org/.

[17] Feross Aboukhadijeh, “WebTorrent”, https://webtorrent.io/.

[18] Flanagan, D., 2011, “JavaScript: The definitive guide: Activate your web pages”, O'Reilly Media, Inc.

[19] Van Acker, S., Nikiforakis, N., Desmet, L., Piessens, F. and Joosen, W., 2014, “Monkey-in-the-browser: malware and vulnerabilities in augmented browsing script markets”, Proceedings of the 9th ACM symposium on Information, computer and communications security, pp. 525-530.

[20] Google Inc., “Website Translator”, https://translate.google.com/manager/website/.

[21] Google Inc., “Google Terms of Service”, https://www.google.com/policies/terms/.

[22] The Google Reader team, “Google Reader”, https://www.google.com/reader/about/.

[23] Dan Lewis, “Keep Google Reader Running”, https://www.change.org/p/google-keep-google-reader-running.

[24] Google Inc., “A second spring of cleaning”, https://googleblog.blogspot.com/2013/03/a-second-spring-of-cleaning.html.

[25] Google Inc., “View webpages cached in Google Search Results”, https://support.google.com/websearch/answer/1687222.

[26] Google inc., “Google Translate Community”, https://translate.google.com/community.

[27] Google Inc., “Language - Google Translate”, https://translate.google.com/about/intl/en_ALL/languages.html.

[28] SIL International, “Summary by world area | Ethnologue”, http://www.ethnologue.com/statistics.

[29] Loreto, S. and Romano, S.P., 2012, “Real-Time Communications in the Web: Issues, Achievements, and Ongoing Standardization Efforts”, IEEE Internet Computing, 16(5), pp. 68-73.

[30] KadTools, “KadTools”, http://kadtools.github.io/.

[31] ISO/IEC, “ISO/IEC 15897:2011(en), Information technology — User interfaces — Procedures for the registration of cultural elements”, https://www.iso.org/obp/ui/#iso:std:iso-iec:15897:ed-2:v1:en.

[32] Chakrabarti, C., “Google Cache 404 Error – Effect on Rankings”, https://ritoban.com/google-cache-404-error-effect-on-rankings/.

[33] Bray, T., “The JavaScript Object Notation (JSON) Data Interchange Format”, http://tools.ietf.org/html/rfc7159.html.

Towards Identity Management in Healthcare Systems

Silvino Neto[1], Felipe Silva Ferraz[1,2], Carlos André Guimarães Ferraz[2]

[1]CESAR
Recife Center for Advanced Studies and Systems
Recife, Brazil
silvino.neto@gmail.com, fsf@cesar.org.br

[2]Informatics Center
Federal University of Pernambuco
Recife, Brazil
{fsf3, cagf}@cin.ufpe.br

Abstract - *Information Systems are of key importance for efficient healthcare services. They improve patient care and administration, providing valuable support for medical diagnosis. To provide such services, healthcare information systems collect and store an extensive volume of patient data in digital format, referred as electronic health record. These records hold a significant amount of patient personal information that may be targeted by cybercriminals. Recently, appalling statistics concerning the exposure and theft of electronic health records have been reported. In this paper we examine the issues related to information privacy and security for healthcare systems and present a new approach for protecting patient data, using an Identity Management framework to preserve patient anonymity. To evaluate this approach, a case study experiment using a disease surveillance platform has been conducted, and its results are exposed in the remainder of this paper.*

Keywords: Security; Architecture; Healthcare Systems; Identification; Interoperability; Identity Management.

1 Introduction

Cyber-attacks are responsible for millions of data records stolen every year. Recently, cases of information theft have increased exponentially, in part due to an unprecedented volume of electronic personal information stored, processed or transmitted.

A recent report claims that 1-in-3 electronic medical records have been compromised in 2015 [1]. Health care providers and software vendors admit that almost a hundred million patient records have been exposed last year.

For a cyber criminal, health care data are usually up to 5, 10 or even 50 times more valuable than other forms of personal information. Health care records often include social security and credit card numbers, beyond exploiting the data they contain, criminals may use these records to fill and resell prescription medications and file fraudulent claims [2]. In spite of all the investments made to protect sensitive information, current technologies and solutions still lag behind the discovery of new vulnerabilities and threats by cybercriminals.

In one of our previous work, a disease surveillance middleware platform has been developed, in order to extract statistical information from electronic medical records for monitoring and prediction of outbreak occurrences [3]. Health records are transmitted using the Fast Healthcare Interoperability Resources (FHIR) protocol, which is an international standard for healthcare system-level data exchange. This middleware platform is designed to process thousands of patient records in a highly transactional environment. Patient records usually include both clinical and administrative data. During the course of this research project, we observed several weaknesses related to patient information privacy and identity exposure.

Silva Ferraz and collaborators proposed an experimental security architecture for smart cities, addressing the security flaws aforementioned. This architecture is composed by a software security layer that abstracts communication and cryptography, providing entities (individuals, services or systems) with a mechanism to interact with other systems using unique generated IDs for each corresponding system [4][5]. A significant contribution of this approach is to enable information consumers to manage their own identity across heterogeneous systems [5].

This strategy allows healthcare applications and platforms, which handle sensitive patient information, to prevent patient identity from being exposed, by using a security layer that overrides patient identification with a newly generated correlation ID. Therefore, the patient real identification is preserved, assuring the privacy and anonymity of individuals.

This paper analyzes security issues that may compromise sensitive information in the context of healthcare applications, discusses the strategies to mitigate the risks involved with data leakage, and presents a case study implemented as result of the development of the proposed security architecture, within a disease surveillance middleware platform. The remainder of this paper is organized as follows: Section II sets the scene by discussing the background topics and related research on healthcare systems security. Section III presents a disease surveillance middleware platform as a case study and describes the scenarios evaluated in Section IV. Section V presents our conclusions and future research.

2 Background

Information security and privacy in the healthcare industry is an issue of the utmost importance. The integrity and confidentiality of the information contained in a patient medical record must be ensured under all circumstances. In this section, we examine security-related aspects for healthcare information systems, addressing topics such as: privacy, systems vulnerabilities and threats.

2.1 Healthcare Privacy and Security

Appari, A. and Eric Johnson, M. noted that privacy is considered as a fundamental principle of the patient-physician relationship [6]. That same principle must be assured by healthcare information systems, in order to guarantee the confidentiality of the data contained in electronic health records. Eventually, a patient's medical record gathers a significant amount of personal information that may be used for malicious purposes, if stolen or violated by cybercriminals.

Given the emergence of many Internet-based healthcare information systems, the risk of exposure of the information contained in electronic medical records is considerably higher than the one observed for local network healthcare systems. Remote healthcare platforms are likely to accelerate patient care, public health incident reporting, and administrative procedures. However, the challenges presented by remote distribution are much greater in terms of guarantee of data integrity and confidentiality.

The evolution of the healthcare industry has been driven by developments in information technology and regulatory measures. Many countries have introduced legislation in order to assure the privacy of individuals, protecting against the disclosure of personal information [8]. Regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. and the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada, establish a set of rules to be followed by physicians, hospitals and other health care providers [8][17][18]. HIPAA/PIPEDA aims to ensure that all electronic medical records satisfy certain standards and requirements, regarding documentation, information handling and privacy [7].

Nowadays, electronic health records are constantly exchanged between several parties, being used for a range of purposes [9][12]. In spite of that, there are important questions left to be answered, concerning the ownership and the access privileges over the data collected [7]. As noted by Marci Meingast, Tanya Roosta, and Shankar Sastry, electronic patient records are usually accessible by physicians, insurance providers, and the patients themselves. But it is not clear who has the authority to maintain the information. HIPAA requires that all patients must be granted access to their own electronic medical records, allowing them to correct errors or omissions, and to monitor how their personal information is shared and used [7].

In light of this issue, this paper proposes the use of an Identity Management (IdM) mechanism, allowing patients to maintain and protect their identity, using a security layer that overrides sensitive information when exchanging electronic medical records [8]. This approach is meant to safeguard patient information against security issues like data breaches, identity theft, and unauthorized access. For instance, an insurance provider shall not be capable of visualizing information contained in a patient medical record, other than the reimbursement of medical expenses.

Additional alternatives may be used to improve information privacy and security. Anonymized electronic health records are suitable for data mining, disease surveillance and outbreak monitoring. Access control policies, including partial viewing of specific segments contained in a medical record, can be enforced by the use of a Role-Based Access Control (RBAC) framework [11].

2.2 Healthcare Vulnerabilities and Threats

There are several sources of threats to healthcare information privacy and security. Internet-based healthcare information systems are more likely to be exposed by cybercriminals compared to regular paper-based medical records. Sensor data, which consist of data captured and transmitted by wearable medical devices for remote patient monitoring, are also susceptible to eavesdropping and skimming attacks, when it's transmitted wirelessly. Data access, storage, and integrity are key challenges when implementing electronic health records and in-home sensor networks [7]. These are some examples of the threats that endanger the privacy and security of the patient information exchanged by healthcare systems. Past studies have broadly separated privacy threats and vulnerabilities into two main categories [6], presented as follows:

1) Organizational Threats: These are situations that occur as result of inappropriate access of patient information by either internal or external agents [6]. For example, a hospital employee that abuse their access priveleges to view patient data for no justifiable reason, or an outside attacker that infiltrates a hospital healthcare information systems in order to steal a patient medical record. Both cases are characterized as organizational threats, given that patient information has been violated due to the exploit of vulnerabilites in the organization itself. These attacks may be inflicted by an individual with modest financial backing, or very elaborated well-funded groups [6]. Studies suggest that organizational threats may be further categorized into five different levels, presented in increasing order of sophistication [6]:

Accidental disclosure: Healthcare personnel unintentionally disclose patient information to others.
Insider curiosity: An insider with data-access privilege pries upon a patient's records out of curiosity or for their own purpose.
Data breach by insider: Insiders access patient information and transmit it to outsiders for profit or revenge.
Data breach by outsider with physical intrusion: An outsider enters the physical facility either by coercion or forced entry and gains access to the system.
Unauthorized intrusion of network system: An outsider, including former employees, patients, or hackers, intrudes

into an organization's network from the outside to gain access to patient information or render the system inoperable.

2) Systemic Threats: These are risks related to the exploit of the information flow chain, perpetrated by legally authorized personnel that use the disclosed information beyond its original intent [6]. For instance, an employer that denies a promotion based on the information contained in the employee's medical records, or a health insurance company that denies coverage due to a preexisting medical condition.

3 Case Study

Previous studies brought into attention the need to make further improvements related to information security on healthcare information systems [5] [6] [7] [8]. Based on this need, this section describes the use of a Healthcare Security Layer (HSL) developed on top of the Platform for Real-Time Verification of Epidemic Notification (PREVENT), which is a disease surveillance middleware platform used to process statistics extracted from electronic medical records in order to anticipate and report outbreak occurrences [3].

3.1 Middleware Architecture

PREVENT is a Message-oriented Middleware (MOM) platform, built to collect and process a large volume of information, in a highly scalable fashion. It provides a set of RESTful interfaces to be used by healthcare information systems to exchange aggregate data reports held in electronic medical records [3].

PREVENT handles a significant amount of sensitive patient information. However, current support for data security mechanisms is limited to the use of the TLS/SSL data transport protocol [15][16]. Therefore, numerous security breaches have been observed in this platform, most of them related to privacy concerns. In order to address this situation, this case study examines the use of a Healthcare Security Layer to be integrated into PREVENT architecture.

The HSL aims to assure patient anonymity, protecting the privacy and confidentiality of individuals, by overriding sensitive information included in electronic medical records with new identifiers, generated from the combination of the patient personally identifiable information (e.g. Name, Social Security, and Credit Card Numbers, etc.) and the healthcare provider identification.

The diagram exhibited in Figure 1 is an overview of PREVENT multilayered architecture, including its most significant software components. PREVENT is composed by 5 different layers, each one of them with a distinct set of responsibilities, described as follows:

Presentation Layer: Hosts the RESTful Servlet endpoints used to receive messages and subscription requests dispatched by healthcare information systems.
Service Layer: Extracts and processes the data contained in the electronic medical records received from healthcare information systems, and coordinates communication with other layers in order to complete data analysis operations.
Persistence Layer: Stores and retrieves information persisted in a NoSQL database.
Communication Layer: Handles communication between PREVENT and the Google Cloud Publish/Subscribe Services API.
Security Layer: Encapsulates the access to PREVENT Communication Layer and provides a mechanism to ensure the privacy and confidentiality of patient data, using an Identity Manager component.

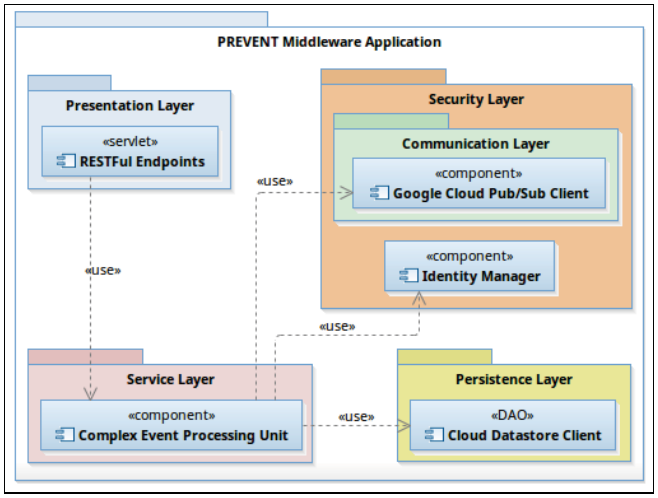


Figure 1. PREVENT Architecture Overview.

The physical architecture of this platform is presented in Figure 2. In this deployment diagram, it is possible to observe that communication between PREVENT and its subscribed healthcare information systems is performed using HTTP over TLS/SSL [15][16].

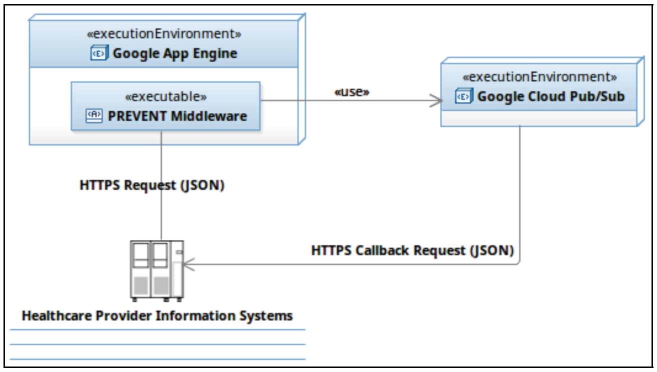


Figure 2. PREVENT Deployment Diagram.

In spite of the fact that the information exchanged between PREVENT and other healthcare systems is transmitted over a secure channel, when the data have reached the other end, they may be exposed and violated by a third party. In other terms, once the data is out of the scope of this platform, privacy and security concerns cannot be enforced, compromising the anonymity of patients. In order to address this issue, in the next section we discuss the use of an IdM framework [5] integrated into the PREVENT platform.

3.2 Healthcare Security Layer

Silva Ferraz et al. observed that, in the context of Smart Cities, citizen identity integration across multiple system platforms and services comprises the goal of allowing the citizen to manage their own identity. This also includes what type of information is released about them to whom or when, while anonymous, aggregate data are more widely available [5]. That same concept can be translated and applied here, in relation to the identity of a patient.

The HSL is responsible for overriding patient sensitive information contained in messages received from healthcare providers, prior to their delivery to subscribed healthcare information systems. Message information replacement occurs as part of a cryptographic procedure, enabling the protection of patient personally identifiable information. The new identifiers are generated by a combination of patient identifiers and healthcare provider ID, being unique within the whole platform [5].

In order to generate a new patient identifier, PREVENT concatenates the original identifier value with the healthcare provider ID, the resulting string value is then ciphered using a 256-bits cryptographic hash function (SHA-2) [20]. Finally, the obtained hash byte array is encoded back to string format using Base64 binary-to-text representation [19]. The new identifiers are then used to replace the original values in the messages received in FHIR/JSON format.

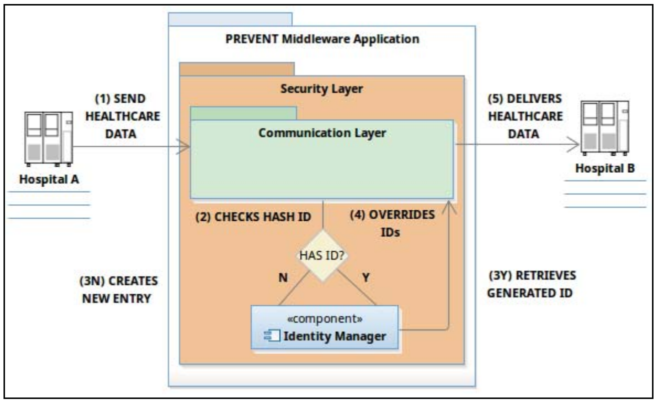


Figure 3. Healthcare Security Layer Basic Flow.

Also, a new entry is persisted in the platform NoSQL database, in order to correlate the original patient identifier, the healthcare provider ID, and the newly generated identifier. In order to illustrate the HSL data flow, a collaboration diagram is exhibited in Figure 3. This diagram presents a scenario where Hospital A wants to share electronic medical records with PREVENT middleware platform for aggregate data reporting. The patient records received are processed and delegated to PREVENT Communication Layer, which checks whether the received patient identifiers and the healthcare provider ID are related to previously generated hashes. According to the outcome of this test, hash IDs may be created or retrieved, in order to override patient information. At last, the updated messages are dispatched to Hospital B.

3.3 Electronic Health Records

As previously discussed in this paper, electronic health records are composed by a set of sensitive personal information. Recently, the numbers of incidents related to patient data leakage and theft have increased significantly. This situation exposes the need for more robust and sophisticated information security. PREVENT's HSL aims to improve information privacy by using two combined techniques: cryptography and anonymity. In this section, we demonstrate how it affects the contents of an electronic healthcare record, and what type of data should typically be anonymized.

A regular electronic patient record typically contain a significant amount of personal information, such as: full name, home address, phone number, social security number, and sometimes even a credit card number may be present. If data is exploited for malicious purposes, it may lead to a number of damages and consequences for the patient. Using a secure channel for message transmission ensures that the data within a patient record will not be compromised. However, when the data is delivered to another healthcare information system, it may be an easy target for violation or theft, as previously demonstrated in this paper. In the following sections, this paper presents an assessment of this platform, introducing a comparison between a secure and non-secure approach for healthcare data exchange, in terms of functional and performance-related results.

3.4 Non-Secure Approach for Data Exchange

In this approach, electronic health records are exchanged between this platform and the subscribed healthcare information systems, without any type of modification. Therefore, any personal information held in the patient record, is sent along with aggregate clinical data reports. Towards establishing a reference parameter for functional and performance-related results, in this section we analyze PREVENT functional behavior, and collect a few metrics in order to assess the performance of this middleware platform. The results gathered will set the boundaries for comparison between the secure and non-secure approaches.

This assessment is performed at the Google Cloud Platform, which is a modular cloud hosting service. The machine type used for this test is a standard Google Virtual Machine Instance, with a single virtual CPU that is equivalent to a 2.6GHz Intel Xeon E5 and 3.75GB of memory. The PREVENT middleware application is deployed at the Google App Engine, which is an application server for hosting cloud-based web applications [13][14]. The test cases used for this assessment are distributed amongst four different testing scenarios. The first and the second scenarios are composed by 300 message samples.

The third and the fourth scenarios are comprised by 500 messages each. A total of 1600 messages are expected to be sent to the middleware application using the Apache JMeter, which is a Java-based performance testing tool. Each message holds information related to a unique patient. The platform is configured to deliver messages to 50 registered HTTPS endpoints, each one of them acting as healthcare information systems. Each registered HTTPS endpoint is either a Java Servlet class deployed at the Google App Engine or a PHP file hosted at the Digital Ocean NGINX server that simply logs the current timestamps and the contents of each FHIR message received.

FHIR health records are usually represented in JSON format. They are composed by several key-value pairs that are used for data representation. According to the data structure of a FHIR message, sensitive patient information such as: home address, identifier value (social security number), and full name are stored in plain text in PREVENT NoSQL database. Later, the FHIR messages persisted are also delivered to the platform subscribers without any type of type of treatment for data anonymization. In the course of this assessment, this middleware platform presented a stable functional behavior. No messages were reported lost nor have any requests been rejected. All messages have been successfully delivered, in accordance with the expected outputs for this simulation. Table 1 illustrates a set of metrics that has been collected in order to analyze performance-related results. The results gathered during this assessment are an essential input, allowing us to identify potential performance bottlenecks related to the use of cryptographic algorithms for data anonymization.

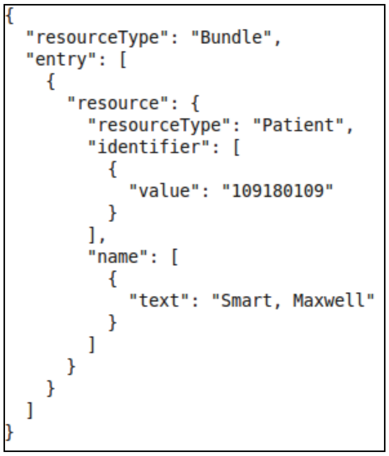
TABLE I. PERFORMANCE METRICS FOR NON-SECURE APPROACH

Number of Samples	Median (ms)	Throughput	Lost Messages	Failed Requests
300	231	211.942/min	0	0
300	218	218.125/min	0	0
500	209	257.889/min	0	0
500	211	250.077/min	0	0

3.5 Secure Approach for Data Exchange

This approach relies on the use of cryptography and encryption, in order to anonymize electronic health records exchanged between this platform and the registered healthcare information systems. Hence, all personal information included in the patient medical records must be ciphered and encoded prior to its storage and delivery. In the previous section, we have established reference parameters for functional and non-functional requirements, based on the results collected during the assessment of the non-secure approach. In the scope of this assessment, the same set of metrics is to be collected, for later comparison against the metrics previously obtained. The differences observed over the collected results are an accurate indicator of how significant is the impact for performance and functionality. The present assessment is to be conducted using precisely the same testing environment, as the one utilized for the non-secure approach. Furthermore, an updated version of the PREVENT middleware application has been deployed at the Google App Engine, including the security features provided by the Healthcare Security Layer [13][14]. The testing script to be used for this assessment is the same as the one presented for the previous experiment, including testing tools, methods, and datasets. Four rounds of execution are programmed for this evaluation, and the same amount of message samples will be used for processing.

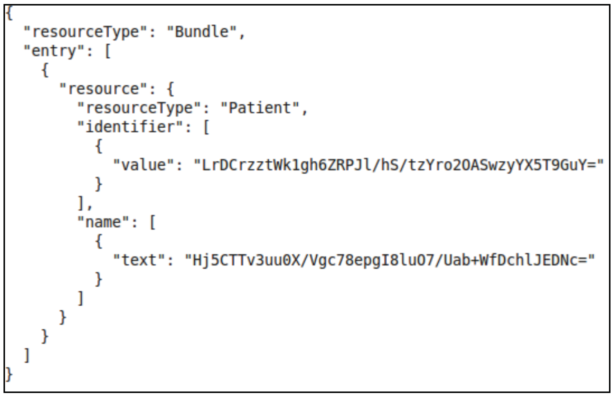
As previously discussed in this paper, FHIR health records are received by this platform over a secure channel. However, sensitive patient information held on these records is usually received in plain text format. That is, as long as the received information is kept within the boundaries of this disease surveillance platform, patient information privacy and security is preserved. The HSL aims to address this limitation by using an Identity Management framework that overrides sensitive patient information with ciphered values for outgoing messages [10]. The HSL allows system administrators to define a set of attributes contained in a patient medical record that are set to be encrypted for protection. Also, a new NoSQL entity kind named GeneratedHashValues was created in order to store the generated hash IDs, indexed by the combination of the patient attribute value and the healthcare provider ID. To demonstrate the HSL capabilities for this assessment, we have enabled the protection of two health records attributes: patient identifier (social security number) and full name. It is important to observe that this assessment will likely reproduce the worst-case scenario, given that an empty dataset for generated hash IDs will be used. Put simply, for every message received, both identifiers (social security number and full name) will be encrypted and stored, resulting in additional processing overhead.



```
{
  "resourceType": "Bundle",
  "entry": [
    {
      "resource": {
        "resourceType": "Patient",
        "identifier": [
          {
            "value": "109180109"
          }
        ],
        "name": [
          {
            "text": "Smart, Maxwell"
          }
        ]
      }
    }
  ]
}
```

Figure 4. FHIR Health Record Snippet Before Anonymization.

Figure 4 shows a FHIR health record snippet, obtained prior to data anonymization and protection. Several attributes were suppressed for the sake of simplicity and due to the size limitations of this paper. As previously discussed, under the non-secure approach, patient data is represented in human-readable format. After the encryption of sensitive patient information, a secure and privacy-assured patient record is obtained, as the one exhibited in Figure 5. As occurred for the non-secure approach, the middleware sustained a stable functional behavior during the whole experiment. No lost messages or functional errors have been observed. As presented in Table 2, the same set of metrics has been collected for further comparison.



```
{
  "resourceType": "Bundle",
  "entry": [
    {
      "resource": {
        "resourceType": "Patient",
        "identifier": [
          {
            "value": "LrDCrzztWk1gh6ZRPJl/hS/tzYro2OASwzyYX5T9GuY="
          }
        ],
        "name": [
          {
            "text": "Hj5CTTv3uu0X/Vgc78epgI8lu07/Uab+WfDchlJEDNc="
          }
        ]
      }
    }
  ]
}
```

Figure 5. FHIR Health Record Snippet After Anonymization.

In the next section, we compare and analyze the results obtained by both approaches, in order to establish the balance between the following QoS requirements: performance, privacy, scalability, and security.

TABLE II. PERFORMANCE METRICS FOR SECURE APPROACH

Number of Samples	Median (ms)	Throughput	Lost Messages	Failed Requests
300	230	205.590/min	0	0
300	232	235.331/min	0	0
500	229	248.063/min	0	0
500	233	238.226/min	0	0

4 Evaluation

This present study has analyzed privacy and security concerns for data exchange between healthcare information systems, aiming to compare the use of both secure and non-secure strategies. During this research, we have carried out the following simulation studies to validate the approach. Sections 3.4 and 3.5 compare functional behavior and performance-related results of both experiments, in order to assess the benefits and disadvantages of using a security layer for data exchange between healthcare information systems.

TABLE III. PERFORMANCE METRICS COMPARISON

Non-Secure Approach		Secure Approach		Results	
Median	Throughput	Median	Throughput	Diff. (%)	Diff. (%)
231 ms	211.942/min	230 ms	205.590/min	-0.43	2.56
218 ms	218.125/min	232 ms	235.331/min	6.42	-1.07
209 ms	257.889/min	229 ms	248.063/min	9,56	3.81
211 ms	250.077/min	233 ms	238.226/min	10.4	4.73

These experiments all used the PREVENT middleware platform for processing electronic health records received for outbreak detection and anticipation. PREVENT was running over the Google Cloud Platform, with fifty randomly chosen subscribers for message delivery simulation. Throughout both experiments, secure and non-secure approaches did not differ at all, in terms of functional behavior. All messages were successfully delivered after the completion of both experiments. This is a compelling evidence of the HSL transparent functional operation. Similarly, minor performance differences have been observed, when comparing the results of the secure and non-secure strategies. Tables I and II present the results obtained for each approach, in an isolated manner. Table III presents a comparison between the results obtained for both approaches, in a combined manner. In spite of the fact that our secure approach performs an additional cryptographic operation for patient information. The performance impact observed is a very insignificant decline, especially when measuring the benefits of using a security-based approach for patient data. After analyzing the measured numbers, we can conclude that the performance levels of our secure approach for healthcare data exchange was similar than the one observed for the common insecure strategy, given that in general the throughput and the median response time were slightly impacted. As security positive impacts we can point out the following:

Privacy increased: Now patient data are no longer associated with the patient real identity. Therefore, it is not possible to link a certain disease or condition to one specific individual.

Anonymity: In accordance with the previous bullet statement, the separation between patient identity and its data promotes a transparent and immediate mechanism for data anonymization, by simply removing information from the security layer.

Patient tracking: The papers [21] and [22] states that citizen tracking can be harmful, given the risks associated with sensitive data exposure. Applying that same principle to patient tracking, through the adoption of the proposed solution, patient tracking is no longer viable.

Authentication and Authorization enforcement: The presented solution is mainly focused on Identity Management, nevertheless it provides the means to re-enforce eventual authentication and authorization needs through its separated approach. Hence, in case a user or physician, for instance, has no appropriate permission to do so, they will not be granted access to the patient information.

It should be remembered that, this assessment was performed in simulation of the worst-case scenario for the secure approach, given that we used an empty dataset for generated hash IDs. Therefore all protected identifiers, contained in every message received, had to be encrypted and stored prior to message delivery.

5 Conclusions

This paper has presented a case study focused on the adoption of a security architecture designed for the improvement of information privacy and security. The proposed architecture has been originally envisioned as a response to privacy and security concerns observed in the context of Smart Cities. The security challenges encountered for smart-city environments are similar, to some extent, to the ones observed for healthcare information systems. In order to address this situation, the present research performed a case study based on the development of a Healthcare Security Layer (HSL) within an existing disease surveillance middleware platform, used to extract statistical information from electronic medical records for monitoring and prediction of outbreak occurrences. The HSL is comprised of an Identity Management framework implementation, enabling healthcare providers to protect sensitive patient information while exchanging patient records with other institutions. The HSL centralizes control over sensitive attributes or identifiers, preventing exposure or unauthorized access to restricted resources. Furthermore, HSL uses one-way cryptography for generating hash IDs for protected data. Thus, the obtained IDs are used as replacement of the original values prior to message delivery. Therefore, target systems never get access to the patient personal information, avoiding any possible violation. There is an increasing need for data protection, considering the escalation of threats related to data exposure and identity theft. In the last year, there have been widely reported cases of patient data leakage by major corporations. As consequence, security is currently perceived as a crucial concern of most scalable messaging systems, especially those where data is confidential and privacy must be guaranteed. Based on the facts exposed by this study, we believe that the following factors contribute to the use of an Identity Management framework for healthcare information systems: **Escalating threats of data exposure**, **Transparent functional mediation**, **Centralized control over digital identities**, **Insignificant performance impact**.

The results gathered in the course of this case study demonstrate the substantial benefits of using an Identity Management platform for enforcing privacy and security requirements for healthcare systems. It should be highlighted that transparent functional behavior has been observed during the whole experiment, and only a minor performance impact has been noted. As future work it is planned to conducted a more detailed performance and load validation in order to assess the impact of an extra layer in a performance based environment, also it is planned to connect the solution with different data bases and services, that works as data sources, validating the interoperable characteristic of the security layer.

6 References

[1] NBC TV News Report. [Online]. Available from: http://www.nbcnews.com/nightly-news/video/check-your-health-records-1-in-3-americans-info-compromised-in-2015-622124099777 2016.02.19

[2] Computerworld Online Magazine. [Online]. Available from: http://www.computerworld.com/article/3013013/healthcare-it/cyberattacks-will-compromise-1-in-3-healthcare-records-next-year.html 2016.02.19

[3] Silvino Neto, Márcia Valéria, Plínio Manoel, and Felipe Ferraz, "Publish/Subscribe Cloud Middleware for Real-Time Disease Surveillance," 10th International Conference on Software Engineering Advances, pp. 131-138, 2015.

[4] F. Silva Ferraz, C. Sampaio, and C. Ferraz, "Towards a Smart City Security Model Exploring Smart Cities Elements Based on Nowadays Solutions," ICSEA 2013, pp. 546–550, 2013.

[5] F. Silva Ferraz, C. Sampaio, and C. Ferraz, "Towards a Smart-City Security Architecture: Proposal and Analysis of Impact of Major Smart-City Security Issues," SOFTENG 2015, pp. 108–114, 2015.

[6] Appari, A. and Eric Johnson, M. "Information security and privacy in healthcare: current state of research," Int. J. Internet and Enterprise Management, v. 6, n. 4, 2010, pp. 279-314.

[7] M. Meingast, T. Roosta, and S. Sastry, "Security and Privacy Issues with Health Care Information Technology," Computer-Based Medical Systems (CBMS), IEEE Engineering in Medicine and Biology Society, 2006, pp. 5453-5458.

[8] Khin Than Win, "A review of security of electronic health records," Journal of Health Information Management, v. 34, n. 1, 2005, pp. 12-18.

[9] J. Singh, L. Vargas, J. Bacon, K. Moody, "Policy-based Information Sharing in Publish/Subscribe Middleware," IEEE Workshop on Policies for Distributed Systems and Networks, 2008, pp. 137-144.

[10] M. Adnan Tariq, B. Koldehofe, K. Rothermel, "Securing Broker-Less Publish/Subscribe Systems Using Identity-Based Encryption," IEEE Transactions on Parallel and Distributed Systems, 2014, pp. 518-528.

[11] J. Bacon, D. M. Eyers, J. Singh, P. R. Pietzuch, "Access Control in Publish/Subscribe Systems," 2th International Conference on Distributed Event-based Systems, 2008, pp. 23-34.

[12] J. Singh, J. Bacon, D. Eyers, "Policy Enforcement within Emerging Distributed, Event-Based Systems," 8th ACM International Conference on Distributed Event-Based Systems, 2014, pp. 246-255.

[13] Google Cloud Platform. [Online]. Available from: https://cloud.google.com/docs/. 2016.02.29

[14] Google App Engine. [Online]. Available from: https://cloud.google.com/appengine/docs/. 2016.02.29

[15] The Transport Layer Security (TLS) Protocol Version 1.2. [Online]. Available from: https://tools.ietf.org/html/rfc5246/. 2016.02.29

[16] Prohibiting Secure Sockets Layer (SSL) Version 2.0. [Online]. Available from: https://tools.ietf.org/html/rfc6176/. 2016.02.29

[17] Health Insurance Portability and Accountability Act. [Online]. Available from: http://www.hhs.gov/hipaa/. 2016.02.29

[18] Personal Information Protection and Electronic Documents Act. [Online]. Available from: http://laws-lois.justice.gc.ca/eng/acts/P-8.6/index.html/. 2016.02.29

[19] The Base16, Base32, and Base64 Data Encodings. [Online]. Available from: https://tools.ietf.org/html/rfc4648/. 2016.03.01

[20] US Secure Hash Algorithms (SHA and SHA-based HMAC and HKDF). [Online]. Available from: https://tools.ietf.org/html/rfc6234/. 2016.03.01

[21] F. Silva Ferraz and C. Ferraz, "More Than Meets the Eye In Smart City Information Security: Exploring security issues far beyond privacy concerns," 11th IEEE International Conference on Ubiquitous Intelligence & Computing, 2014, pp. 677-685.

[22] F. Silva Ferraz and C. Ferraz, "Smart City Security Issues: Depicting information security issues in the role of a urban environment," 7th International Conference on Utility and Cloud Computing, 2014, pp. 842-847.

Ethical concerns regarding the use of Intelligent User Interfaces

Suhair Amer
Department of Computer Science,
Southeast Missouri State University, One University plaza, Cape Girardeau, MO, USA 63701
samer@semo.edu

Abstract - *Intelligent user interfaces have progressed in many ways throughout the years. Interfaces are implemented to help with everyday tasks and applications. However, there are many ethical concerns regarding the use of intelligent user interfaces. People are concerned with privacy as in many cases such intelligence have to collect personal and non-personal information to complete an accurate profile of the user. In this paper we will discuss some of the uses of intelligent user interfaces, advantages and disadvantages and ethical concerns related to its use.*

Keywords: Intelligent user interface, artificial intelligence, ambient intelligence

1. Introduction

Intelligent user interfaces are interfaces that include some aspect of artificial intelligence. They are implemented to create a more cohesive communication between the computer and its user and adapt based on user's preferences. They can respond to a user's gestures, key strokes, and preferences through various online content. They gather this information using aspects of psychology, cognitive science or computer graphics to create a more interactive and personalized experience for the users [Tavani 2011].

Ambient intelligence describe technological environments that preform daily tasks by responding to the presence of people. They rely on intelligent user interfaces to reach its full potential. Ambient intelligence runs in the background and is not seen or detected by the user. Profiling becomes possible when intelligent user interfaces are implemented in an ambient intelligent environment. Profiling is the ability to personalize and automatically adapt to a person's behavior patterns making the more usable and useful. Most people are unaware that this technology is used in many things that they use on a daily basis [Tavani 2011].

Researchers dealing with the concept of pervasive computing are able to provide smart products that communicate unobtrusively. Ambient Intelligence refers to convergence of two factors that include ubiquitous and pervasive computing. On the other hand, bioinformatics refers to convergence of both information technology and biotechnology. Finally, Nano-computing refers to convergence of computing and nanotechnology [Seelman 2008].

Ubiquitous communication ensures communication through various interlinked computing equipment such as wireless local area networks, and radio frequency Identifications. The concept of pervasive computing is made possible with the use of enhanced intelligent users interfaces. Intelligent user interfaces ensure that there is enhanced interaction among people through better intuitive ways. Such interfaces can sense and relate to a person, situation, context, or the environment as compared to the traditional interfaces that only performed basic functions [Jutai et al. 2005].

Ambient intelligence helps people to live and work in environments that respond to their interactions in intelligent ways. Intelligent user interfaces utilizes artificial intelligence and concepts of being a personal assistant. The idea is to have interfaces that can be personalized and can adapt to users and their preferences. Usually users may not be even aware of the existence of intelligent user interfaces in their normal day to day interactions (which is what a natural user interface strives for). They are also may not be aware that we are surrounded with hundreds of intelligent networked computers that may sense the presence of the user and are aware of their personality, and needs [Tavani 2011].

An Intelligent User Interface connects between a personal and the system he/she is using. Such interface should make this interaction easier, and not be a burden placed on them especially if they need to learn to use this interface. Adapting to the user is done by using techniques from artificial intelligence to perform reasoning and learning. This can be done by performing user modeling and recognition. To be most effective, the interface allows the system's user interaction be adapted to different usages. It should utilize user modelling which allows a systems to retain information about a user. It should also utilize natural language technologies which lets a system either create or interpret text or speech in a system. Utilizing dialogue modelling allows a system to maintain a natural course of interaction using a user's primary language. Explanation generation allows a system to explain to the user its end results. Sometimes machines may have to learn and foresee future situations; therefore, the interface should be able to acquire dynamic information and use it with already acquired knowledge and provide seamlessly experience to the users. Current Intelligent interfaces replaces pointing and clicking with speaking or swiping. In addition, such interfaces should be developed in a way can be used by young and old, the tech whiz and not, and by the casual user or a person in an emergency [Laster 2001].

Currently, there are many examples of intelligent user interfaces, and some are more successful than others. For example, there are the adaptive and collaborative interfaces, affective interfaces, agent-based interfaces, model-based interfaces, and natural language interfaces [Leake 2004].

To summarize, not all intelligent systems use intelligent user interfaces. The intelligent interfaces usually use a set of techniques such as user adaptivity, user modelling, natural language technology, dialogue modelling, and explanation generation. Even if a technology uses these techniques, "an intelligent interface must utilize technology to make an improvement: the resulting interface should be better than any other solution, not just different and technically more advanced" [Ehlert 2003].

2. Examples of Uses

Currently, intelligent user interfaces can be found in and used by numerous devices. Traditionally interfaces were controlled by the computers keyboard, mouse, and monitor. Now, intelligent user interfaces are able to determine information about the user without receiving typed commands. Websites and search engines use this form of technology in order to create a better experience for its users. For example, advertisements are chosen depending on what the user views and what they click on while they are browsing through websites and using search engines such as Google. Cell phone companies started using intelligent user interfaces to design new and improved smart phones. Some smart phones now have the technology to make changes based on the user's eye movements. They are able to automatically scroll down as the user reads. Smart phones have many ways to monitor and make changes to the settings based on the surrounding environment. The phone's screen can become brighter or dimmer based on the brightness of the sun or lights. [Ehlert 2003]

Companies are trying to create tutoring systems that provide online learning content [Guerra et al. 2016] that utilizes intelligent user interfaces that would provide a better more personalized experience then if they were to hire a human tutor. An intelligent tutor is "a program that aims to give a personalized "education" to a user in a specific domain of knowledge" [Shute and Psotka 1994]. Tutoring programs are designed to detect strong and weak points in the user's subject. For example, if the program detects that the student is having trouble with a specific subject or area, the system will provide more practice problems and examples that involve that area. These tutoring programs are essential to online schools and degrees. Turing systems can also help with everyday life of people of all ages. The tutor program infer the user's understanding of the domain through analyzing the user's performance on test problems. The tutoring system provides active advice by intervening, and suggesting alternative courses of actions, or passively, by answering explicit users queries [Ehlert 2003].

Recommendation systems also use intelligent user interfaces. For example, Netflix looks at movies and television shows that the user has watched, and recommends new shows that the user may like based on their previously viewed programs. This recommendation list has helped many people find shows and movies they may not have considered without them being recommended to them [Ehlert 2003]. In general, to allow users to navigate content, an intelligent television uses intelligent interfaces [Sirpal, 2016].

Intelligent user interfaces are also used to filter information. Especially when surfing the internet. The user's search results are selected based on their search and browsing history. Other viewer's preferences are also taken into consideration when selecting information. Some systems consider similar users with similar interests and searches and recommend accordingly [Tavani 2011]. This is also important when dealing with data collection, as the system is logging what a user is reading or accessing and then provide websites or articles for the user to read based on their interests. Many companies are putting more time and effort into adaptive technologies.

There are many practical uses for Intelligent User Interfaces as seen with the refrigerator. This said there are three major overarching categories of IUI's that can be talked about. These are system functionality, user, and wants and needs. A IUI that focuses on system functionality "might have some knowledge of how to get around the system, or tasks a user would want to do. With this information, the system can present its interface in an intelligent manner, making navigation and operation more intuitive to the user."[Ehlert 2003] A great example of this is Apples Siri. This program allows users to ask the phone how to do things, to do things for them, or allows them to skip past the interface itself directly to the information they were wanting.

3. Advantages

Intelligent user interfaces are the key to have a more personalized experience when using technology. It simplifies the process of finding information and give users access to advertisements that would be of interest to them. For example, students may benefit from online tutoring programs [Ehlert 2003]. They can also help users find information about their surroundings and environment.

Newer smart phones may provide a street view of local shops or emergency services and give directions, distances and phone numbers, when one is lost in a city. Some of them are always connected and waiting for our voice commands and requests [Orland 2013].

Such systems also can be set up to take into consideration an individual's habits, preferences and ways of working. Using such information, the system can provide personalized interaction methods that are best for a user. It can also aid with filtering problems and information over-load since searching is usually a tedious and time consuming task. An intelligent interface can reduce the quantity of information to look at. Also, by filtering out irrelevant information, the interface can reduce the user's cognitive load. Sometimes , the system can also help find useful information one may not have been aware of [Alvarez-Cortes et al. 2009].

Other intelligent user interfaces can help a user learn a new software and can teach the user how to use its features. Such a feature is usually available with videogame introductions. In other examples, an intelligent user interface monitor a user's actions or tasks performed and try to understand the context and recognize his/her attempt, and finally deal with the execution of that task, allowing the user to focus its attention to others tasks [Alvarez-Cortes et al. 2009].

Other uses may involve informing the user of detailed and personal information about their environment. Such examples include informing the user about or even controlling climate that measures air quality, moisture in the air, and allergens especially those with medical problems [Tavani 2011].

4. Disadvantages

While intelligent user interfaces can be very helpful and beneficial, they can be a burden. When the user is given information that should be geared towards what they want, the results could be completely off. Users are usually are provided with information that has the same content or opinion. This will cause pigeon holes in the information. This is because the user is basing his/her beliefs and reference information on the information provided and is not aware of all aspects or sides of other information that is still available online [Ehlert 2003].

It is not easy to design Intelligent User Interfaces or programs that are able to act and give a personal, individual experience each time. Programmers have to avoid scripting scenarios, and instead set up programs that adapt and act on their own according to a given scenario. Programs that are scripted have to follow a direct flow chart or steps. However, if a program is set as a free agent, it is able to adapt to different situations and is able to 'converse' with a user, as opposed to following a set 'script' [Andre and Rist 2001].

Technology, is also, not advanced enough to be used efficiently and can end up causing more problems especially when movement of parts is involved. Sometimes, gathered information about the user is not correct and does not represent the user but a single instance of his/her behavior that may not occur again. This can lead to creating incorrect inferences about the user, his/her actions, or the situation. These incorrect inferences may require the user to perform or be involved in corrective actions [Tavani 2011].

Many users are naïve with regard to setting up their privacy options or are unaware of breaches. Many users don't update their software regularly and don't have current security checking software installed. This privacy related issues can be abused easily. If the intelligent user interface was not correctly set up, the system can get hijacked. In other instances, it may allow some parties to spread misinformation and control what people see and do. Some think that virtually anything said, done, and sometimes felt can be digitized, stored, and later on retrieved. Sometimes collected information can be used for denying some people from legal benefits such as health insurance [Tavani 2011].

Another concern is that although it is easy to build an intelligent user interface in theory it is actually quite hard to scale up and make it work at a larger scale. Artificially based intelligent systems have been typically developed by academia which develops and tests a limited number of devices that is addressing a specific problem and then those systems cannot be scaled up to the cover more functionalities or be used on a larger scale [Hook 2000].

Another disadvantage is that people can become technologically dependent on the intelligent user interfaces. This is a problem because even though they can relieve humans from worrying about performing many routine tasks, that are sometimes tedious and boring, it is mainly relieving us of cognitive effort that enables us to be fulfilled and flourish as humans. There are also worries that if humans depend heavily on such devices and they go down, we will not be able to perform our daily tasks. Similar to many businesses depending on grocery stores, if the grocery's store system goes down, all other businesses need to wait until it is back up [Tavani 2011].

5. Ethical Concerns

The use of intelligent user interfaces raises many ethical concerns. Many believe that the gathered information violates their privacy. For example, profiling is viewed as a major invasion of the user's privacy. This is because the information is collected and then used to create a profile of the person. This profile is then used to recommend what information is presented to the user. As these systems are working in the background without the knowledge of the user, it is hard to detect them. This is why some people are concerned that their personal information; such as, credit card numbers, bank information, and social security number may be also collected without their knowledge [Ehlert 2003].

Other ethical concerns include freedom and autonomy, privacy, and technological dependency. Many question whether human autonomy and freedom will be preserved with the use of Ambient intelligent. Some believe that humans will be able to have better control in the environment as they are able to interact more by the aid of technology that is responsive to their needs. However, other believe that it is not the humans who are gaining much control of the environment; rather such control is delegated to the machines. There is therefore the belief that the machines have robed humans the right and privilege to experience life first hand. Ambient intelligent can make human's life more controllable in three main dimensions: it can respond quickly to the needs of the users. It can also react quickly to the intentions of the users, as well as their actions. It can also provide users with personal information in a detailed manner [Jutai 2005].

In addition, there are other ways where Ambient intelligent can diminish the level of control that humans have enjoyed over their environment. For example, smart objects can make wrong inferences about a user's intentions, actions, or situation which can compromise the user's ability to remain in control over their environment [Jutai 2005]. There is also the possibility that in some instances, smart objectives may require user's corrective actions which denies the user's distinct role of making decisions on various circumstances in the environment. In rare occasions, the smart object may advance the needs and interests of other parties other than the user. In addition, there is the possibility that the smart object may lack the capacity to address human challenges over the environment.

With regard to privacy and surveillance, Ambient intelligent has four unique features that cannot be found on other famous computing applications. They include ubiquity, sensing, invisibility, and memory application. Since such devices are inserted invisibly in ambient intelligent surroundings, it increases the possibility of invading privacy. This is because there is a high possibility that the users will not realize their presence, as such, it is possible to disseminate and collect personal data [Nahrsted and Chu 2008]. Sensing devices, which are interlinked with the integrated user interface, are so sensitive that they are able to sense emotions emanating from human. Such emotions include stress, fear, and excitement and have the capacity to maintain records of collected data.

The greater the capabilities of technologies nowadays, as well as, the possibility of the presence of compromised privacy make individuals unsure about whether their presence is being recorded. The fact that people actions are being monitored and recorded means that they cannot be sure of what the future holds for them or in which context such private information will be used. Some believe that when personal information is being accessed, it should be with their informed consent because it is a violation of their fundamental rights and privileges as an individual. There is a need for the stakeholders in the information sector to come up with regulations that safeguards the interest of the society. For example, there should be place measures that ensure that privacy and confidentiality of the public. This is especially when human nowadays rely heavily on cyber technology that relies on the convergence of wireless technologies, the internet, and advanced electronics [Nahrsted and Chu 2008].

Other ethical issues include Privacy and surveillance threats. For example, the interface of Xbox One can be used to spy on its users. It comes with a Kinect (a camera and microphone array) that when launched, it is supposed to be always on. This feature was supposed to help the end user sign their apps, control their system with voice or gesture, and enable them to launch or accept a Skype call. Users became concerned that they could be hacked or spied on and that their personal and biometric information that is stored for the Xbox fitness game or recognition system that is stored on Microsoft's cloud servers would be attacked [Orland 2013].

Edward Snowden NSA leaks are another example where users are concerned about how much of their personal data and browsing is being monitored by the government. There is also the idea that researchers are planning to go beyond taps and keystrokes, and to use the accelerometers and gyroscopes in smartphones to determine a user's gait, and analyze which apps are opened and at what times of day and at which locations [Stromberg 2013]. Although such information may be gathered for good reasons by the intelligent user interface, others may use them for other reasons such as sending you annoying add to stalking or blackmailing. That is why some believe that these ethical concerns can be addressed by having total transparency. Meaning that, the user should be informed about any information that will not be kept private and might be shared.

Many systems that utilize a user model forces users to accept that the system will keep a representation or patterns of their behavior. Some intelligent user interface systems require that users share their preferences with a user community. For example, Netflix gets its recommendations by checking other users that have watched a particular film that the user has watched, and recommends other films that the other users have also watched. Another example is Doppelgänger which allowed people to create personalized news papers so that they could view the news that interested them. This system allowed people to copy the personalized newspapers of others. The difference between both systems is that Netflix was anonymous, however, Doppelgänger, displayed actual names instead of being anonymous [Hook 2000].

Another concern with the use of intelligent user interfaces is trust. This occurs whenever systems start doing things automatically for users which may not always be what the user wants. This will cause a problem because if the users do not trust a system, they will not use it. This would be a problem if, for example, the system starts sorting mail, filtering news, retrieving information from the web, selling and buying goods, etc., [Hook 2000].

6. Conclusion

While the use of intelligent user interfaces will always raise ethical concerns, many will continue to enjoy the ease that they create. Tutoring programs, intelligent search filtering, and cell phone technology are just a few of the ways that intelligent user interfaces are being implemented.

Many people are connected and are using computer systems or smart-phones and as these systems become more complex and feature-rich, it is important that the way we interface with them keeps up with these changes. It not convenient if our interaction with devices is complicated as the tasks they perform. Interfaces need to be able to deal with huge amount of information, be helpful and be personalized to the end user while trying to complete a task [Virvou and Kabassi 2002].

It is important to understand that humans are relying heavily on the tasks performed by electronics, machines and computers. It is important for the user to have the tools that would help him/her decide when and what is considered safe in relation to what is being controlled or accessed via the interfaces. The users should feel safe while having such interfaces controlling or accessing their sensitive information.

7. References

[Alvarez-Cortes et al. 2009] Alvarez-Cortes ,Victor, Zarate , Victor H., Ramirez Uresti , Jorge A. and Zayas , Benjamin E. (2009). Current Challenges and Applications for Adaptive User Interfaces, Human-Computer Interaction, Inaki Maurtua (Ed.), ISBN: 978-953-307-022-3, InTech, DOI: 10.5772/7745. Available from: http://www.intechopen.com/books/human-computer-interaction/current-challenges-and-applications-for-adaptive-user-interfaces

[Andre and Rist 2001] Andre, Elizabeth, and Thomas Rist. "Controlling the Behavior of Animated Presentation Agents in the Internet." AI MAgazine 22.4 (2001): 53-66. Web.

[Ehlert 2003] Ehlert, Patrick. Intelligent User Interfaces : Introduction and Survey. Mediamatics/Data and Knowledge Systems Group. Delft University of Technology, Feb. 2003.

[Guerra et al. 2016] Guerra, Julio, et al. "An Intelligent Interface for Learning Content: Combining an Open Learner Model and Social Comparison to Support Self-Regulated Learning and Engagement." Proceedings of the 21st International Conference on Intelligent User Interfaces. ACM, 2016.

[Hook 2000] Höök, K. "Steps to Take before Intelligent User Interfaces Become Real."Interacting with Computers(2000): n. pag. Web Ree Source 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).

[Jutai et al. 2005] Jutai, J. W., Fuhrer, M. J and DeRuyter, F. (2005). Toward a taxonomy of assistive technology device outcomes. American Journal of Physical Medicine & Rehabilitation, 84, 294–302.

[Laster 2001] Lester, James. "Introduction To The Special Issue On Intelligent User Interfaces." AI Magazine 22.4, 13-107. Winter 2001.

[Leake 2004] Leake, David. "The Seventh International Conference on Intelligent User Interfaces (IUI-2003)." AI Magazine 24.4 (2004): 131-32. Web. 2 May 2015.

[Nahrsted and Chu 2008] Nahrsted, K., Chu, H., (2008). QoS-aware Resource Management for Distributed Multimedia Applications. Journal on High-Speed Networking 7(3/4), 1998.

[Orland 2013] Oreland, Kyle. "Hands-on With the Xbox One: Kinecttct, Interface, and OS Impressions. » arstechnica.com, Nov. 2013.

[Seelman 2008] Seelman, Katherine D. (2008). Converging, Pervasive Technologies: Chronic and Emerging Issues and Policy Adequacy Assistive Technology: The Official Journal of RESN. 20(3), 152-156

[Sirpal, 2016] Sirpal, Sanjiv. "Systems and methods for providing user interfaces in an intelligent television." U.S. Patent No. 9,232,168. 5 Jan. 2016.

[Stromberg 2013] Stromberg, Joseph. " How You Type Could Become Your New Password. " Smithsonian.com., Jul. 2013.

[Tavani 2011] Tavani, Herman T. Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing. John Wiley & Sons, Inc. 2011.

[Virvou and Kabassi 2002] Virvou, Maria, and Kabassi, Katerina. "Reasoning About Users' Actions In A Graphical User Interface." Human-Computer Interaction 17.4, 369-398. 2002.

Round-Robin Staggered-Imputation (RRSI) Algorithm for Enhanced Real-Time Prognostics for Dense-Sensor IoT Applications

Kenny C. Gross, Kalyan Vaidyanathan, Anton Bougaev, and Aleksey Urmanov

Oracle Corporation
{kenny.gross,kalyan.vaidyanathan,anton.bougaev,aleksey.urmanov}@oracle.com

***Abstract* -** New real-time prognostic algorithms are being developed for large-scale internet-of-things (IoT) applications for high-throughput ingestion of time-series sensor signals for manufacturing, transportation, and utilities IoT applications. As the numbers of sensors grow for IoT prognostics, and the sampling rates for modern data-acquisition (DAQ) become ever higher, the available bandwidth of legacy industrial networks often becomes saturated, which limits the accuracy and fidelity of real-time empirical prognostic models. We introduce here a new "round-robin staggered-imputation" (RRSI) algorithm that increases the available bandwidth for legacy networks, with no hardware modifications/upgrades. An example proof-of-concept demonstration is provided with million-sensor prognostic monitoring of data center assets (servers, storage systems, integrated "engineered systems"). The concept is readily extensible to any industrial IoT applications involving large numbers of sensors with high sampling rates for high-accuracy prognostic fault monitoring and with low false-alarm and missed-alarm probabilities, thereby achieving higher reliability, availability, and serviceability (RAS) for IoT customer business-critical and mission-critical assets.

Keywords: Prognostic Pattern Recognition, IOT Prognostics

1 Introduction

Enterprise computer systems designed for high reliability, availability and serviceability (RAS) application environments now contain hundreds to thousands of physical sensors to monitor the condition of individual components such as power supplies, dc/dc converters, memory and CPU modules, ASIC's, hard disk drives (HDDs), solid state drives (SSDs), fan motors, and other components. For example a relatively small two-rack-unit (2U) typically now has up to 150 physical transducers measuring temperatures, voltages, currents, fan speeds, power levels, and vibrations. Currently-shipping 4U enterprise servers have over 600 sensors. A rack-sized engineered system such as Oracle's M6 has 3400 physical sensors, which happens to be the same number of sensors as in an 800 MWe nuclear reactor. Medium sized enterprise data centers these days have over 1 million sensors. Extensive dense-sensor-prognostics experience gained over the last decade by Oracle's "Electronic Prognostics" (EP) applications [1-4] in million-sensor data centers [5-6] is now being extended to other industrial IoT application domains.

For enterprise data center assets, as computers get hotter and denser with Moore's law, thermal and electrical margin assurance for systems present ever increasing challenges. This has motivated the increase in density of sensors in systems. The challenge this has presented for prognostic machine-learning algorithmics is that IO bandwidth for system bus architectures has not kept pace with Moore's law. In the future, the industry may eventually move to a new system bus standard, but no one server vendor wants to "go it alone" and implement a non-standard system bus architecture. Hence, the computing industry is stuck for the foreseeable future having to live with a very limited bandwidth for a bus architecture designed more than 15 years ago, back when typical servers contained 5-6 sensors, in sharp contrast to today's servers that are provisioned with hundreds or thousands of sensors, and in a smaller form factor.

The specific challenge for real-time prognostics that has been created by limited bus bandwidth is that as the number of sensors in servers has grown, the sampling rate for those sensors has necessarily been throttled back. Thus, for new rack-sized systems containing over 3,000 sensors, the sensors can be sampled only once a minute or longer. This creates a significant challenge for real-time prognostic fault monitoring [7-9]. Depending on a particular degradation mechanism one wants to monitor for using prognostic anomaly detection, higher sampling rates are desirable and lead to higher prognostic accuracy, with lower false-alarm and missed-alarm probabilities, to unambiguously detect the incipience or onset of signatures of impending failures.

This paper overcomes the above challenges by providing a novel analytical innovation for systematically increasing the sampling rate for groups of important individual sensors so that all diagnostic and prognostic functionality can be met within the hardware IO bandwidth constraints of typical system bus architectures and even for very large numbers of sensors.

It is important to point out that we do not just pre-assign fast sampling rates for some sensors while reducing sampling rates for other sensors. The simplest way to introduce this radically new approach is by way of a very trivial 3-sensor analogy:

Suppose we have a piston that is compressing air in a simple cylinder. Suppose we have 3 sensors, a Pressure sensor (P), Temperature sensor (T), and a gauge that measures the Volume of the cylinder (V).

It is desired to measure the 3 sensors P, V, and T with a high sampling rate, but suppose the digitized observations must come through a system bus with very limited IO bandwidth.

One way we could analytically increase the effective sampling rate for these 3 sensors, but without increasing the IO bandwidth, would be if we just sampled 2 signals at a time and computed the 3rd signal. In this trivial example, we happen to know from the Ideal Gas Law a relationship between the 3 signals [P*V = N*R*T]. Knowing this relationship, for this simple hypothetical use case, we could:

for the 1st polling interval:

1) Sample P and V, use those measurements to compute T

for the 2nd polling interval:

2) Sample P and T, use those measurements to compute V

for the 3rd polling interval:

3) Sample V and T, use those measurements to compute P

In effect, even though we have a limited bandwidth, we can now sample 33% faster and use the above “staggered sampling” algorithm to compute the unsampled signal at each time interval. Granted, we are using some CPU cycles in the service processor...but those are in abundance. We're trading off very limited system bus bandwidth for very cheap CPU cycles in the Service Processor (SP). The result is that we have complete time series traces for P, V, and T with a 33% greater sampling rate, but meeting the same constrained IO bandwidth.

For the multitudes of telemetry metrics inside complex computer servers, we rarely if ever have nice first-physics equations relating any of the telemetry signals to any other signals as in the trivial example above. However, we do have highly accurate empirical cross correlation relationships as learned through advanced statistical pattern recognition. Specifically, we use an advanced pattern recognition approach called the Multivariate State Estimation Technique (MSET), originally developed by the US Dept of Energy for prognostic applications in nuclear plants, Nasa’s Space Shuttle, and safety-critical applications [10-16].

The approach developed herein is very analogous to the trivial 3-variable motivational illustration above, where we use MSET to “impute” one or more staggered values in a systematic round-robin approach. [Disclaimer: we use the term MSET in the generic sense to represent a form of nonlinear, nonparametric (NLNP) regression, but are not endorsing a specific commercial implementation of MSET.] It is important to point out here that we are not using conventional “missing value interpolation” algorithms that “fill in” missing values in a univariate time series using conventional interpolation schemes. Conventional forms of univariate “missing value interpolation” algorithms suffer from the fact that they are inherently a “lossy” computation (in other words, no matter how cleverly one interpolates to replace a missing value in a univariate time series, the “true” value could be significantly different). The reader will see that our novel round-robin “staggered imputation” algorithmic structure developed herein is not “lossy” at all … in fact our novel algorithm introduced in this paper for enhanced prognostics and enhanced cyber security is “gainey” insofar as the “imputed” values are actually more accurate than if the values had been sampled by a hardware or software sensor. (See technical details below on how the round-robin staggered-imputation [RRSI] algorithm introduced in this paper achieves significantly higher bandwidth for prognostic IoT and cyber security applications, and actually *gains* in accuracy for the digitized time-series telemetry signals being monitored.

The approach introduced below is fundamentally analogous to the trivial 3-variable model described above: i.e. we stagger the sampling rate in a systematic fashion and use a high-accuracy functional relationship [via nonlinear nonparametric regression via MSET] to compute the unsampled values for signals from the sampled values for other signals, such that all sensors can be sampled under given constraints on the service bus bandwidth and computational resources of the service processor, and at the same time all desired anomaly mechanisms can be reliably detected.

2 RRSI Algorithmic Implementation

We develop herein a technique for assigning sampling policies to individual sensors and clusters of sensors in a business-critical IoT customer assets based on the cross-imputability property that allows sampling of all sensors under given bandwidth and computational constraints and, at the same time, to assure reliable detection of degradation mechanisms and impending failures via prognostic anomaly detection. The cross-imputability property is computed for each sensor. The cross-imputability of a sensor tells how well the values of this sensor can be predicted from the values of the other, correlated sensors in the system. All sensors are then grouped or clustered according to their cross-imputability property. [Note: the term “cluster” in this document is used in the statistical sense and does not imply that physical sensors are spatially clustered into close proximity to one another. Sensors that are well correlated with one another and are clustered according to their cross-predictibility may or may not be in close physical-spatial proximity.] Each cluster is assigned an appropriate cluster sampling policy. The number of sensor clusters and types of cluster sampling policies can vary depending on specifics of each monitored enterprise computer system.

From extensive proof-of-concept implementation investigations, we have found that three sensor clusters work very well. For a 3-cluster approach, the first sensor cluster contains all sensors with low cross-imputability. The second sensor cluster contains all sensors with moderate cross-imputability. And the third sensor cluster contains all sensor with high cross-imputability. Each of the three clusters is assigned a sampling policy. For example, in the first cluster, which contains digitized sensor sequences that are difficult to predict using the other sensors (because of low cross-correlations with the other transducers), each sensor is sampled with a fixed high sampling rate. In the second cluster all sensors are sampled with a moderate sampling rate. And in the third cluster sensors are sampled in a staggerd round-robin fashion with appropriate sampling rates. All sampling rates across the clusters are adjusted to satisfy bandwidth and computation constraints and assure reliable detection of impending failures.

2.1 Training phase

During the training phase we are capturing all sensor values with highest possible sampling rates, ignoring the bandwidth and computational limitations applicable to the operating environment. During the design stage, some additional hardware/software can be employed to assure capturing all sensors at sampling rates adequate to assure reliable detection of impending failures, and use that sampling rate that is higher than during normal operation for the purposes of training MSET.

During this phase we build non-parametric models that predict the value of a sensor using the values of all the other sensors. A preferred nonparametric technique is a multivariate state estimation technique (MSET). For the demonstration of the concept feasibility a nonlinear kernel regression technique was used. To build a reliable non-parametric model, a data set is split into a training and validation subsets. A model is built on the training subset and is validated on the validation subset. Figures 1-3 demonstrate the process of model building. The top subplot shows the training subset with dots connected by lines and prediction made by the machine learning model with circles. One model is built for each sensor.

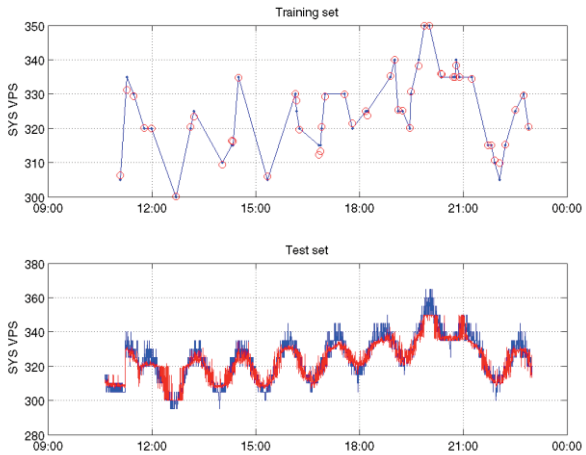


Fig. 1: Example of a sensor with good cross-imputability. Predictions of the values of VPS (virtual power sensor) using all 147 variables but with VPS as the predicted variable. The top subplot shows the training data set consisting of 50 values picked at random times during a 12 hour time period. The dots connected by lines represent the actual VPS values; the circles represent the predictions by the model built on these 50 values. Outstanding prediction accuracy is evident by the close alignment of the predictions with the measured samples. The bottom subplot represents the predictions (in red) of the VPS values (in blue) on the full data set. The predicted values follow very well the actual measured values, which reflects good cross-imputability of the VPS sensor.

2.2 Computing the cross-imputability property value for each sensor

The cross-imputability is computed on the entire data set. In the preferred embodiment the cross-imputability is assessed by using the prediction error on the entire data set. If the prediction error is small, the cross-imputability is high. The values of sensors with high cross-imputability are well predicted by non-parametric models. If the prediction error is large, the cross-imputability is low. The values of sensors with low cross-imputability are hard to predict by the models and are therefore sampled with a uniform sampling rate.

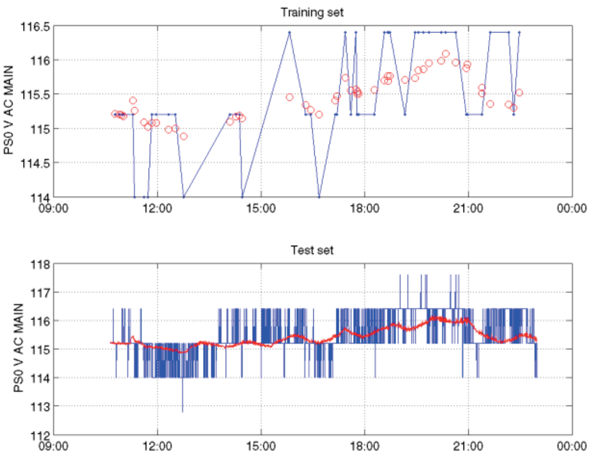


Fig. 2: Example of a sensor with poor cross-imputability. This sensor is input AC voltage to one of the redundant server power supplies.

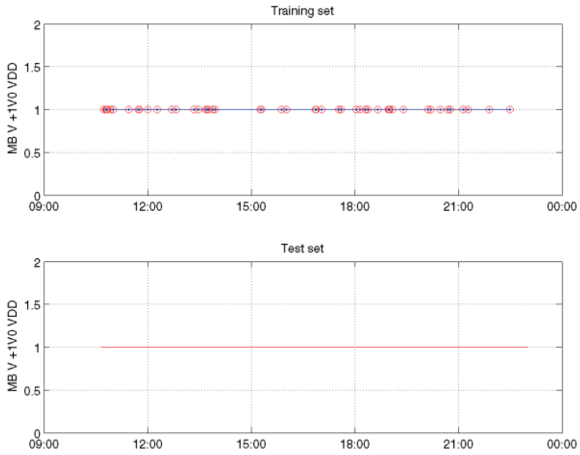


Fig. 3: Example of a flat-line sensor. There's no variation in the sensor's values to be able to build a cross-imputability model.

All sensors are clustered according to the cross-imputability property. Each cluster comprises sensors with a specified range of cross-imputability. For example, one cluster may comprise sensors with low cross-imputability. All sensors in this cluster must be sampled with the highest sampling rates appropriate to catch impending failures. Another cluster may comprise of sensors with high cross-imputability. Sensors in this cluster may be sampled using a staggered round-robin approach that dramatically reduces the required bandwidth and computational resources.

2.2 Monitoring phase

Sensors are sampled according to the corresponding cluster sensor policies.

A demonstration has been performed on a single socket 1U enterprise computer server that has 148 internal sensors. A nonparametric pattern-recognition model was built for all 148 sensors and each sensor's cross-imputability was computed. All sensors are first sorted according to the prediction error. The sensors with high prediction error are the ones with low cross-imputability. An example of a high imputability sensor is given in Figure 1. This is a virtual power sensor. Its value can be reliably predicted using other sensors and a pattern recognition model, hence its imputability is high. An example of a lower-imputability sensor is shown in Figure 2. This is an AC input voltage to one of the power supplies.

3 Analytical Bandwidth Enhancement:

The RRSI algorithm provides a systematic way of clustering sensors in sensor-rich IoT systems to devise realizable sensor sampling policies such that all sensors are sampled under given constraints on service bus bandwidth and computation resources of the system processor and assuring the reliable detection of impending failures or other anomalous conditions required by failure detection and control applications.

There is no loss of accuracy (in fact there is a gain in accuracy from the new RRSI algorithm). This is the case because the imputed value is being computed with multiple other correlated signals that are measured at the same instant in time as the imputed value, so that if there is a disturbance in the system being monitored, it will very likely be reflected through the multiple correlated values being leveraged in our round-robin staggered-imputation algorithm. (As opposed to a conventional univariate missing-value imputation approach via interpolation that is basically filling in a “blind spot” in a univariate time series with an interpolated value).

For effective utilization of sensors in sensor-rich computer systems, this RRSI approach allows system designers to intelligently devise sensor sampling policies using trained pattern recognition models. For design, the method empowers thermal and mechanical engineers with a means of specifying more sensors when desirable under limited bandwidth and computational resources constraints and dropping sensors not important for reliable detection of impending failures and for control.

4 Conclusions

The analytical bandwidth-enhancement technique introduced in this paper consumes as inputs raw time-series digitized telemetry signals from all types of IoT physical transducers, but then processes and acts upon that information with a novel “round-robin staggered-imputation” (RRSI) innovation reported here, to in effect achieve higher sampling rates, with more accurate signals, than was ever heretofore possible in enterprise computing servers. The end result is higher sensitivity for Prognostic Health Monitoring applications that enhance the reliability, availability, and serviceability for IoT assets and for enterprise servers, storage, engineered systems, and networks. While the RRSI concept was presented in this paper in the context of business-critical IT assets in the Cloud Computing industry where a single medium-sized data center contains over 1-million sensors that are “always on”, there are nevertheless important use cases for Internet-of-Things customers in the Manufacturing, Utilities, and Transportation market sectors where data-acquisition bandwidth constraints may be fixed by hardware and firmware designed years earlier, but there is a desire to implement highly-reliable predictive anomaly detection algorithms with low false-alarm and missed-alarm probabilities, such as MSET. The new RRSI innovation presented in this paper is helping to achieve higher reliability margins while reducing penalizing challenges to overall asset-availability goals for business-critical and mission-critical assets monitored by advanced pattern-recognition for proactive anomaly detection.

5 References

[1] “Prognostics of Electronic Components: Health Monitoring, Failure Prediction, Time To Failure,” K. G. Gross, K. W. Whisnant and A. M. Urmanov, Proc. New Challenges in Aerospace Technology and Maintenance Conf. 2006, Suntec City, Singapore (Feb 2006).

[2] "Electronic Prognostics Techniques for Mission Critical Electronic Components and Subsystems," K. C. Gross, K. W. Whisnant and A. M. Urmanov, Proc. 2006 Components for Military and Space Electronics Symposium, Los Angeles, CA, (Feb 2006).

[3] “Proactive Detection of Software Aging Mechanisms in Performance-Critical Computers,” K. C. Gross, V. Bhardwaj, and R. L. Bickford, Proc. 27th Annual IEEE/NASA Software Engineering Symposium, Greenbelt, MD (Dec 4-6, 2002).

[4] "Incipient Fault Detection in Storage Systems using On-Line Pattern Recognition" K. Vaidyanathan, K. C. Gross and R. Dhanekula, Proc. 60th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA (April 2006).

[5] “Integration of Electronic Prognostics with Software Aging and Rejuvenation for Business-Critical Enterprise Servers,” K. C. Gross, 4th IEEE Intn'l Workshop on Software Aging and Rejuvenation, Dallas, TX (Dec 2012).

[6] “Proactive Fault Monitoring in Enterprise Servers,” K. Whisnant, K. C. Gross and N. Lingurovska, Proc. 2005 IEEE Intn'l Multiconference in Computer Science & Computer Eng., Las Vegas, NV (June 2005).

[7] “Failure Avoidance in Computer Systems,” A. Urmanov and K. C. Gross, Proc. 59th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA (Apr 18-21, 2005).

[8] “Proactive Detection of Software Anomalies through MSET,” K. Vaidyanathan and K. C. Gross, Proc. IEEE Workshop on Predictive Software Models (PSM-2004), Chicago (Sept 17-19, 2004).

[9] "Electronic Prognostics Through Continuous System Telemetry," K. C. Gross, K. W. Whisnant and A. Urmanov, Proc. 60th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA (April 2006).

[10] “MSET Modeling of Crystal River-3 Venturi Flow Meters,” J. P. Herzog, S. W. Wegerich, K. C. Gross, and F. K. Bockhorst, ***Proc. ASME/JSME/SFEN 6th Intnl. Conf. on Nuclear Eng.***, San Diego, CA (May 10-15, 1998).

[11] “Multivariate State Estimation Technique (MSET) Surveillance System,” J. P. Herzog, K. C. Gross, S. W. Wegerich, and R. M. Singer, Appendix H of ***On-Line Monitoring of Instrument Channel Performance***, **TR-104965**, EPRI (Oct 1998).

[12] “Fault-Tolerance Improvement for a MSET Model of the Crystal River-3 Feedwater Flow System,” A. Miron, S. Wegerich, F. Yue, K. C. Gross, and J. Christenson, ***Proc. IEEE Nuclear Science Symp. and Medical Imaging Conf.,*** Toronto (Nov 1998).

[13] “Regularization Methods for the Multivariate State Estimation Technique (MSET),” N. Zavaljevski, K. C. Gross and S. W. Wegerich, ***Proc. Intn’l Conf. On Mathematics and Computations***, Madrid, Spain (Sept 27-30, 1999).

[14] “A Pattern-Recognition-Based Fault-Tolerant Monitoring and Diagnostic Technique,” R. M. Singer, K. C. Gross, R. W. King, S. Wegerich, ***NASA STI/Recon Technical Rept. N 04/1995; 96:10057***, Argonne National Lab, Idaho Falls, ID (Apr 1995).

[15] “Sensor Fault Detection in Nuclear Power Plants Using the Multivariate State Estimation Technique and Support Vector Machines,” N. Zavaljevski and K. C. Gross, ***Proc. 3rd Intn'l Conf. of the Yugoslav Nuclear Society, YUNSC2000***, October 2-5, 2000, Belgrade, Yugoslavia

[16] “Using Support Vector Machines in the Multivariate State Estimation Technique,” N. Zavaljevski and K. C. Gross, ***Trans. American Nuclear Soc.,*** Long Beach, Ca (Nov 14-18, 1999).

SESSION

POSTER PAPERS

Chair(s)

TBA

An evaluation of social media use in a golf club

Jonathan Bishop
Centre for Research into Online Communities and E-Learning Systems, Swansea, Wales, GB.

Abstract - *This article looks at the social media strategy used in a golf club, namely Pontypridd Golf Club. It compares what it was like prior to the advent of social media and afterwards. It does this through interviewing one of the club's former golf captains, who was involved on both occasions. The study finds that one of the factors most affecting whether the golf club took up social media was the skill of the officers that ran the club. It was expected that a technology office would exist in order to update the website. It was not expected that officers with a particular portfolio would update the parts of the website within their own remit. Understandably, systems like WordPress were deemed complex, but even Facebook was updated by an individual rather than the officers concerned. The study concludes that increasing digital literacy will be essential to making social media use common in golf clubs and potentially any social or recreational group.*

Keywords: Social media, golf clubs, Facebook, WordPress

1 Introduction

Social media is a term used to refer to social networking services that are based primarily around user-generated content. Comparisons can be drawn between current social networking services that use a single-tree structure and those that use a two-tree set-up. Single-tree websites are those that are either content-based or community based, whereas two-tree websites are those which mix content with community [1]. The earliest social networking websites, namely A Guide to Robin Hood and Northern England (1999) and Llantristant Online (2002) were two-tree, but were not popular [2]. The buddylists on the first were manually populated and on the second were user-generated, which was the same with Facemash and Facebook respectively [3].

Facebook became popular when it opened for non-student use in 2006, which was the same year as Twitter. At this point ADSL was taking off as the main means to connect to the Internet. Social media could thus be considered to consist of 'post-modem social networks,' on the basis that services like Twitter and Facebook did nothing that the social networking services from the 1990s and 2000s hadn't done previously, but which were popular in the era of ADSL and optic fibre-based routers, as opposed to dial-up modems.

2 Evaluation and Analysis

The evaluation and analysis consists of an investigation that analyses and evaluates the use of social media practices within a golf club.

2.1 The organisation and interviewee

The organisation chosen was Pontypridd Golf Club. The gold club was chosen because the researcher designed and built a website for it prior to the advent of post-modem social networking services, such as Twitter and Facebook. It was therefore possible to have a convenience sample in the form of an interviewee that was involved with the club at the time and still active within it.

2.2 Methodology

To investigate Pontypridd Golf Club's use of social media and interview methodology was used. This involved designing a semi-structured interview around the criteria known to assist with the effective marketing of websites, namely 'online servicescapes' [4, 5].

2.3 Results

The investigation showed how the use of social media by the golf club has been used – or not used – to aid the organisation and its communication with members of the public and members.

2.3.1 Practices

Asked why the content management system adopted prior to Facebook was not successful it was suggested it was a human resource issue. *"I think at the time, they had a secretary didn't they?,"* the interviewee suggested. *"And she wasn't very technically minded, she couldn't use the Internet or anything."* For a long time, it was the norm that anything to do with IT was done by a *'technology officer,'* rather than the officer responsible for a portfolio. A membership officer might expect an IT expert to post a membership newsletter online rather than do it as part of their role. This was confirmed by the interviewee, who said that the secretary, *"didn't have the capability and eventually even though the website gave them some presence, they weren't able to do it* *themselves"* and that he thought *"that was the only reason really, that they needed something that was simpler."*

The interviewee made clear this made a big impact on the adoption of the earlier website. *"[T]he reason was the club didn't have the digital capability it has and that's why the site failed*."

2.3.2 Effectiveness

The effectiveness of those social media practices adopted by the club was an issue. After the original content management system was abandoned, the club then turned to another web designer. The former captain interviewed said that they *"came along and offered to do one again for them, and that was when [they were] gone from there and that's when they had nothing but trouble from that company."* This shows how the problems of low digital literacy can impact on an organisation's success, especially when it comes to their use of social media. The former captain said how even now he was still responsible for social media, rather than each club officer. *"They don't bother to update it (Facebook) and I'll do it for them, but they've got a website now haven't they?"*

The interviewee was asked why the club is not using a system like WordPress. *"Well I don't know if they, uh, the website is WordPress, but a lot of us don't understand, I don't even know how to use WordPress,"* he said. *"I've not looked into it; you know they say it's simple, it's only simple if somebody shows you how."*

3 Discussion

This study has looked at the social media policies of a golf club between 2004 and 2007, and what has happened since then. It found that a big problem in the adoption of social media was the technical skills of the office holders. Even today, with platforms like Facebook and WordPress, the person with IT skills is doing the job that a dedicated office holder should do as part of their role. If they can do something offline one might think they should learn to do that online. Without digital literacy being something everyone has, this study has shown that the burden of using social media will not be shared, but continue to lie with one person responsible for IT, rather than the person who would do the offline equivalent of the task.

4 References

[1] Derek M. Powazek. "Design for Community: The Art of Connecting Real People in Virtual Places". New Riders, 2002.

[2] Jonathan Bishop. "Evaluation-Centred Design of E-Learning Communities: A Case Study and Review". The 2nd International Conference on Internet Technologies and Applications (ITA'07), Wrexham, GB. 6-9 September 2007, V. Grout, D. Oram & R. Picking, Eds. University of Wales Press, Wrexham, GB, 2007. , 1-9.

[3] José Van Dijck. "Facebook as a tool for producing sociality and connectivity"; *Television & New Media,* 1527476411415291, 2011.

[4] L. C. Harris & Mark M. H. Goode. "Online servicescapes, trust, and purchase intentions"; *Journal of Services Marketing,* 24., 3, 230-243, 2010.

[5] Antje Cockrill, Mark M. H. Goode & Daniel Emberson. "Servicescape matters—Or does it? The special case of betting shops"; *Marketing Intelligence & Planning,* 26., 2, 189-206, 2008.

Adaptive Relay Scheme Based on Post SNR in Cooperative Communication System

Seung-Jun Yu, Chang-Bin Ha and Hyoung-Kyu Song*
uT Communication Research Institute, Sejong University, Seoul, Republic of Korea
*Corresponding Author: songhk@sejong.ac.kr

Abstract – *In this paper, an adaptive relay scheme based on post SNR in cooperative communication system is proposed. conventional cooperative scheme has used the same modulation at a relay. The proposed scheme uses hierarchical modulation at the source and adaptive transmission based on the post signal to noise ratio (SNR) at a relay. Accordingly, the symbol error rate (SER) and throughput performances are improved by the proposed scheme.*

Keywords: Cooperative Communication, Adaptive Transmission, Channel State, Hierarchical Modulation

1 Introduction

A multiple-input multiple-output (MIMO) relay network [1] is composed of source, relay, and destination nodes, each of which is equipped with multiple antennas. Network information theory has shown that the use of multiple relay nodes in source and destination (S-D) communication makes the capacity of the S-D system logarithmically increase with the number of relays. A simple description of the dual-hop relaying protocol [2] is following. In the first hop, the source broadcasts to the relays. In the second hop, the relays simply transmit sources signals in separated channels or transmit them using a space-time code to the destination. Consequently, the reliability of the communications is improved whereas the throughput might go down since the transmission is performed by two times. Conventional dual-hop scheme has used the same modulation method at a relay. When the quality of the source-relays link is good, the usage of higher order modulation at the source improves total system throughput. However, in opposite case, bit error probability is increased through using higher order modulation.

2 System Model

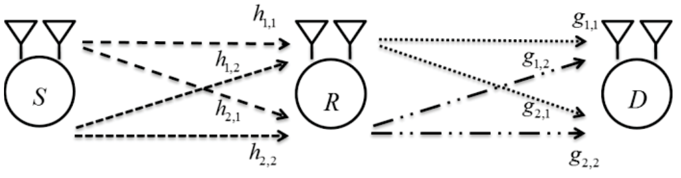


Fig. 1: System model

Consider a dual-hop relay system with 3 nodes which is made up of a source *S* with two antennas, a relay *R* with two antennas and a destination *D* with two antennas. A source *S* and a relay *R* have a single antenna. The coefficient of the link between *i*-th antenna at *S* and *j*-th antenna at *R* is $h_{i,j}$ and $g_{i,j}$ is the coefficient of the link between *i*-th antenna at R and *j*-th antenna at *D*, where *i, j* = 1, 2. It is assumed that each channel goes through Rayleigh fading and the link coefficients $h_{i,j}$ and $g_{i,j}$ are independent and identically distributed (i.i.d). Also, the channel state information (CSI) is known to R and D.

3 Proposed Scheme

In this section, an adaptive relay scheme based on post SNR in cooperative communication system. When the post SNR is low, the proposed scheme maintains SER performance. In order to transmit adaptive signals in a relay, the post SNR is calculated. To know the channel state for each subcarrier, post-signal-to-noise ratio (Post SNR) which can be determined by the channel state is used. The post SNR value for *k*-th subcarrier (ρ_i) is represented as follows,

$$\rho_i = \frac{\left\|X(k)\right\|^2}{\sigma_n^2 \left\|\mathbf{G}_i(k)\right\|^2}, \tag{1}$$

where $\|X(k)\|^2$means transmit power of subcarrier, σ_n^2 denotes noise power and $\mathbf{G}_i(k)$ is the *i*-th row of the Moore-Penrose pseudo-inverse matrix. The noise considered by the zero-forcing (ZF) method which can be denoted as,

$$\mathbf{G}_{ZF} = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H. \tag{2}$$

In Eq. (1), because the transmit signal power is 1, transmit power little impacts on the post SNR. Most impact on the value of the post SNR power is obtained by a Moore-Penrose pseudo-inverse matrix. If the $\|\mathbf{G}(k)\|^2$ value is large, the post SNR is to be small and the channel state is determined to poor state. And vice versa, if the $\|\mathbf{G}(k)\|^2$ value is small, the channel state is determined to good state. The proposed scheme uses hierarchical 16-QAM. If the $\|\mathbf{G}(k)\|^2$ value is large, a relay discards the latter 2 bits. In hierarchical 16-QAM, groups of 2 MSBs are mapped to HM class 1 and groups of 2 LSBs are mapped to HM class 2, respectively. The proposed scheme may be operated using following steps. A source broadcasts hierarchical 16-QAM symbols X_{12} and X_{34} to a relay. The received symbols at a relay in frequency domain are denoted as

$$\begin{aligned} Y_{R_1} &= H_{1,1}X_{12} + H_{2,1}X_{34} + N_{R_1}, \\ Y_{R_2} &= H_{1,2}X_{12} + H_{2,2}X_{34} + N_{R_2}, \end{aligned} \tag{3}$$

where the subscript R_{i} stands at the *i*-th antenna in a relay, $H_{\text{i,j}}$ represents the frequency responses of link between *i*-th antenna at *S* and *j*-th antenna at *R* and *N* is a complex Gaussian random variable with zero mean and variance σ^2. At second time slot, the destination receives adaptive modulated symbols from a relay. When the value of the post SNR is large, the received symbols in the frequency domain can be denoted as

$$\begin{aligned} Y_{D_1} &= G_{1,1}X_{12} + G_{2,1}X_{34} + N_{R_1}, \\ Y_{D_2} &= G_{1,2}X_{12} + G_{2,2}X_{34} + N_{R_2}, \end{aligned} \tag{4}$$

where D_{n} is an index of the *n*-th destination antenna, $G_{\text{i,j}}$ is the channel frequency response of link between *i*-th antenna at *R* and *j*-th antenna at *D*, and *N* is a complex Gaussian random variable with zero mean and variance σ^2. In cased of a small post SNR, the received symbols in the frequency domain can be denoted as

$$\begin{aligned} Y_{D_1} &= G_{1,1}X_{1} + G_{2,1}X_{3} + N_{R_1}, \\ Y_{D_2} &= G_{1,2}X_{1} + G_{2,2}X_{3} + N_{R_2}. \end{aligned} \tag{5}$$

Finally, the original signals are reconstructed with ZF detection algorithm in a destination.

4 Simulation Result

In a simulation, the parameters are as follows. The fast Fourier transform (FFT) size is 128 and the cyclic prefix (CP) length is 32. A source, a relay and a destination node are used over 8-path Rayleigh fading channel. Fig. 2 shows the SER performance of the proposed scheme. According to the value of post SNR, a relay transmits adaptive modulated symbols. Consequently, the proposed scheme shows higher SER performance than that of the conventional scheme.

5 Conclusions

In this paper, an adaptive scheme based on post SNR in cooperative communication system is proposed. According to the value of post SNR, the SER performance of the proposed scheme is improved by using an adaptive modulation. In particular, the proposed scheme guarantees reliable transmission for low SNR.

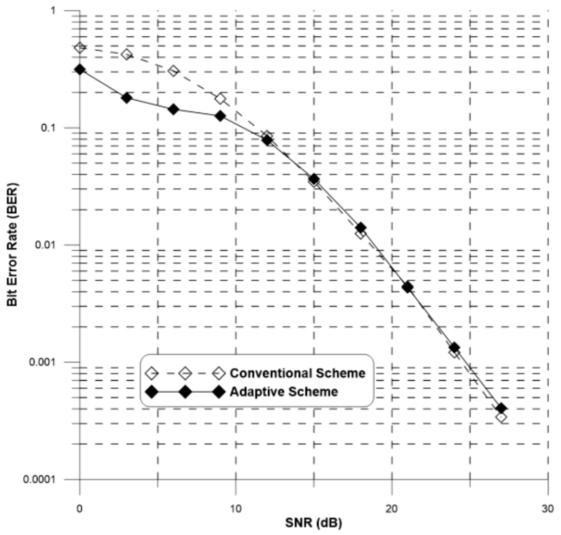


Fig. 2: The SER performance of the proposed scheme

Acknowledgement

This work was supported by the IT R&D program of MOTIE/KEIT [10054819, Development of modular wearable platform technology for the disaster and industrial site].

References

[1] T. Abe, S. Hui, T. Asai, and H. Yoshino, "A Relaying Scheme for MIMO Wireless Networks with Multiple Source and Destination Pairs,". In *Proc. Asia-Pacific Conference on Communications*, 2005, p. 1099.

[2] J.-C. Shin, J.-H. Song, J.-H. Kim, and H.-K. Song, "Dual-Hop Transmission Scheme Based on Hierarchical Modulation in Wireless Networks," *IEICE Trans. Commun*., vol. E93-B, pp. 1645—1648, June 2010.

The Batteryless Wi-Fi Backscatter System and Method for Improving the Transmission Range

Young-Min Ko, Seung-Jun Yu, Seongjoo Lee and Hyoung-Kyu Song*
uT communication Research Institute, Sejong-University, Seoul, Republic of Korea
*Corresponding Author: songhk@sejong.ac.kr

Abstract – *Recently, The Internet of things (IoT) system has attracted attention. IoT is a technology to connect all the objects to the internet as well as computer. Among the IoT technology, the research of devices so that they can communicate without power supply has been actively conducted. Batteryless system permits us to communicate without power supply devices. In this paper, batteryless backscatter system is used as a tag. And mobile devices which are embedded wireless fidelity (Wi-Fi) chipset are used as a reader. The backscatter tag can be obtained Internet connectivity from the reader. Conventional Wi-Fi backscatter system has limitation in the transmission range. In this paper, the proposed algorithm can be obtained improved reliability as well as overcoming the limitation about transmission range.*

Keywords: Backscatter, Batteryless, IoT, Tag, Wi-Fi

1 Introduction

Nowadays, the batteryless communication has attracted attention with great interest for ‘Internet of things (IoT)’. The backscatter technique transmits the information by using the radio-frequency (RF) powered device. Therefore, various approaches for energy harvesting have been researched. Among many energy sources, the ambient RF energy can be used for harvesting on the backscatter system. A large amount of energy can be regenerated from the RF energy waste according to the increasing usage of the wireless communication [1].

When the energy-harvesting is applied to the backscatter system, it can be used in various IoT fields. Because of the energy-harvesting can achieve simple and cost-effective communication, it can contribute greatly to IoT industrialization and commercialization. By using precoding scheme which is used on the MIMO system, the backscatter system can transmit the signal to the desired place. Also, mobile phone which basically embeds Wi-Fi chipset can be used as a reader device on the backscatter system [2].

This paper proposes an improved batteryless backscatter communication system using the ambient Wi-Fi signal as an energy source. Furthermore, the proposed scheme employs the precoding scheme in order to obtained improved reliability as well as overcoming the limitation about transmission rage in accordance with the low-power environments.

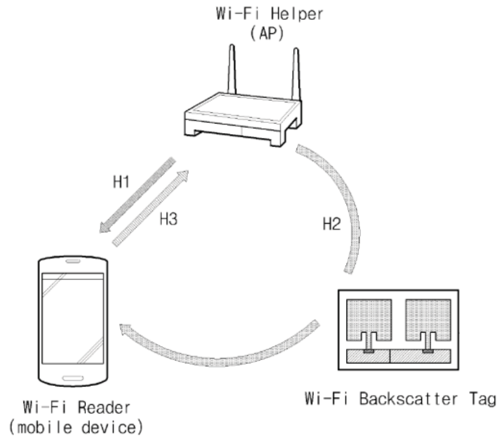


Fig. 1: Wi-Fi backscatter system model

2 Wi-Fi Backscatter System Model

The Wi-Fi backscatter system is represented in Fig. 1. Wi-Fi backscatter system consists of Wi-Fi helper, Wi-Fi reader and several Wi-Fi backscatter Tags. Each tag has its own ID. If the Wi-Fi reader requests the information about one of the several tags to the Wi-Fi helper, Wi-Fi helper wakes up the specific tag by using the Tag ID. The tag transmits its ID and information by backscattering the signal of Wi-Fi helper. The information of tag can include product-related information which the Wi-Fi backscatter user wants to know. The tag is a passive tag. It is implemented without a battery. Therefore, tag can be smaller and the production is possible at low cost. As a result, the range of use is limitless.

3 Proposed Wi-Fi Backscatter System

Conventional Wi-Fi backscatter system has disadvantages about transmission range, reliability and throughput. The proposed Wi-Fi backscatter system improves the reliability by using precoding scheme. The proposed Wi-Fi backscatter system is divided into four steps. The first step is a step that the reader received the broadcasted Wi-Fi signals from helper. The second step is a step that the broadcasted Wi-Fi signals are backscattered from the tag and the backscattered signals are received by the reader. The third step is to identify information about the signals sent by the Tag at the reader. In the fourth step, channel estimation is processed through the received signal at the reader. And the reader transmits the information for precoding to the helper. After that, the helper broadcasts precoded Wi-Fi signals by using precoding information received from the reader. The received signals applying precoding in the reader are as follows:

$$Y = PXH_1 + PXH_2 + N, \quad \left(P = \frac{1}{H_2} \right). \qquad (1)$$

Y means the received signals at the reader. P is precoding factor. X refers to the response signals. H_1 and H_2 is the independent and identically distributed (i.i.d.) zero-mean Gaussian channel. N is the additive white Gaussian noise (AWGN). P , X and H_1 are already known at the reader. Therefore, the only incoming signal component through the tag can be extracted. By sending feedback for precoding on the every time slot, the channel fading impact is reduced in advance. As a result, the proposed Wi-Fi backscatter system can obtain improved reliability as well as overcoming the limitation about transmission rage.

4 Simulation Results

This section shows the simulation results for proposed scheme. The parameters for the simulation are defined as follows: The standard of the IEEE 802.11n is used for the simulation. The fast Fourier transform (FFT) size is 64 and the band-width is 20MHz. the modulation type is BPSK and coding rate for the convolutional coding is 1/2. The number of total transmitted packets is 1000. Fig. 2 shows the bit error rate (BER) performance of the proposed scheme and conventional scheme. It indicates that the proposed scheme improves BER performance according to the signal to noise ratio (SNR). The proposed scheme using precoding obtains about 22dB more than the conventional scheme in BER of 10^{-3} from 3m distance. In Fig. 3, the throughput performance of the proposed scheme and conventional scheme are shown. Because of feedback processing, it is confirmed that the throughput reduces to half. However, the throughput problem can solve through the high-order modulation scheme.

5 Conclusions

This paper proposes a precoding scheme in the batteryless backscatter system to offer the reliable communication. Therefore, the proposed scheme can solve the disadvantages of conventional system about transmission range, reliability and throughput. Also, it can be exploited for various IoT fields.

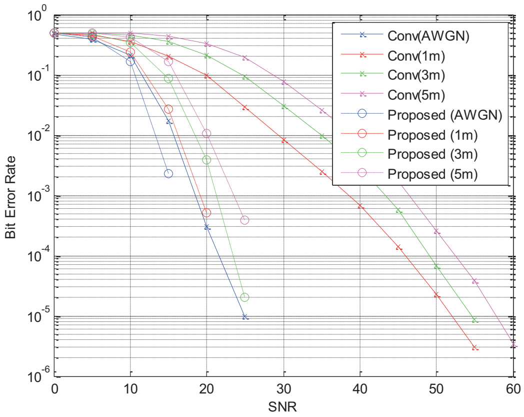


Fig. 2: BER performance for proposed and conventional scheme

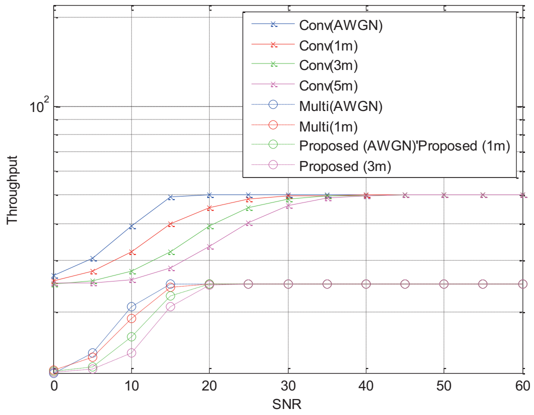


Fig. 3: Throughput performance for proposed and conventional scheme

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.B0126-15-1076, Development of non-powered technology combined with ambient RF energy harvesting and Backscatter data transfer)

References

[1] B. Debasis, S. Jaydip, “Internet of things: applications and challenges in technology and stadardization,” Wireless Personal Comminication, Vol. 58, No. 1, pp.4969-4974, May 2011.

[2] W. Saad, X. Zhou, Z. Han, H. V. Poor, “On the Physical Layer Security of Backscatter Wireless Systems,” IEEE Trans. Wireless Commun., Vol. 13, No. 6, pp.3442-3451, June 2014.

[3] D.M. Dobkin, “The RF in RFID. Burlington,” Elsebier, 2008.

Signaling Scheme Using Phase Shifting in Wi-Fi Backscatter System

Chang-Bin Ha, Young-Min Ko, Seongjoo Lee and Hyoung-Kyu Song*
uT communication Research Institute, Sejong-University, Seoul, Republic of Korea
*Corresponding Author: songhk@sejong.ac.kr

Abstract – *In this paper, the signaling scheme using phase shifting is proposed for the improved performance of the Wi-Fi backscatter system. Because the Wi-Fi backscatter system is based on the RF-powered device, the achievement of high reliability is difficult. In order to solve the problem, the proposed scheme shifts the phase of signal in according to the transmitting information. The simulation result shows that the proposed scheme has the improved reliability.*

Keywords: Wi-Fi backscatter system, Phsae shifting, RF-powered device

1 Introduction

Recently, internet of things (IoT) is rapidly emerged [1]. IoT integrates all things with device to device communication (D2D) and the special-purposed sensor. However, mainly used wireless communication systems such as wireless fidelity (Wi-Fi) and Bluetooth do not support D2D or the sensor.

In order to solve the problem, the Wi-Fi backscatter system is proposed [2]. For the battery-free sensor, the radio frequency (RF)-powered devices in the Wi-Fi backscatter system harvest energy by collecting ambient RF signal. Also, because of the infrastructure of the Wi-Fi, the Wi-Fi backscatter system is suitable for the characteristic of IoT. However, because the harvested energy is limited, the Wi-Fi backscatter system uses the communication by unit of packet. Therefore, the Wi-Fi backscatter system has very low data rate and reliability compared to the conventional communication system.

2 System Model

Fig. 1 shows the Wi-Fi backscatter system model. The system consists of Wi-Fi AP, Wi-Fi reader, and Wi-Fi tag. Wi-Fi tag is RF-powered device. In the uplink, information is conveyed to the reader by using the impedance modulation that adjusts reflectivity for ambient RF signal from the AP in according to the information bit. In the downlink, information is conveyed to the tag by using the on-off modulation that transmits Wi-Fi packet for bit ‘1’ and doesn’t transmit Wi-Fi packet for bit ‘0’.

3 Proposed Wi-Fi Backscatter System

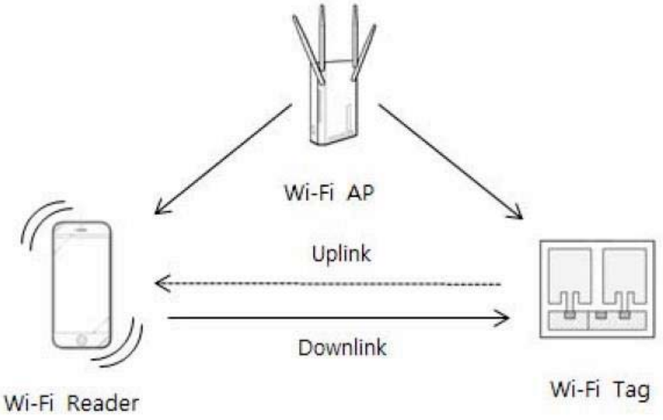


Fig. 1: Wi-Fi backscatter model

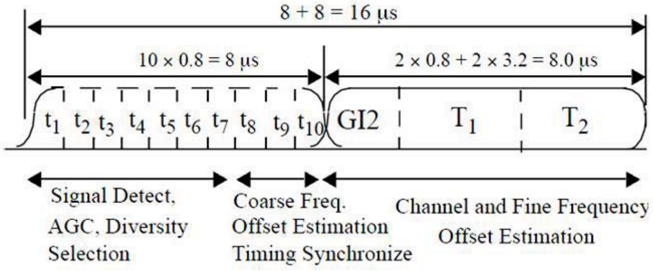

8 + 8 = 16 μs
10 × 0.8 = 8 μs
2 × 0.8 + 2 × 3.2 = 8.0 μs
t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
GI2
T1
T2
Signal Detect, AGC, Diversity Selection
Coarse Freq. Offset Estimation Timing Synchronize
Channel and Fine Frequency Offset Estimation

Fig. 2: Wi-Fi structure of IEEE 802.11a

Fig. 2 shows the considered Wi-Fi packet in the Wi-Fi backscatter system [3]. In this paper, T_2 is the considered signaling part. T_2 is as follows:

$$\mathrm{X} = [X^0, X^1, \cdots, X^{N_C-1}], \tag{1}$$

where N_C is the number of sub-carriers (64). Because of complexity problem, the possible phase shifting values are limited as follows:

$$\mathrm{V} = [1, -1, j, -j]. \tag{2}$$

Before searching T_2 for phase shifting values, the data of T_2 is converted from frequency domain to time domain as follows:

$$\mathrm{x} = IFFT\{\mathrm{X}\}. \tag{3}$$

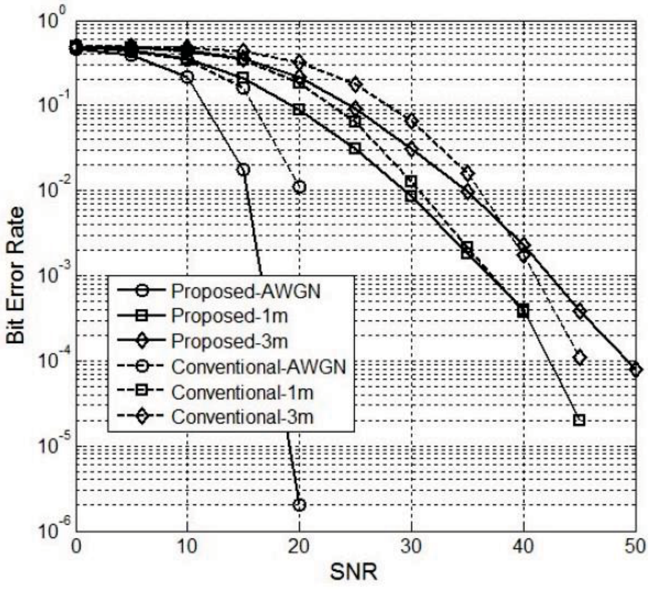
The proposed scheme shifts phase of T_2 in according to the transmitting information. If transmitting information bit is ‘1’, the phase offset among sub-carriers is adjusted to a minimum and the transmitting power is increased. Contrary, if transmitting information bit is ‘0’, the phase offset among sub-carriers is adjusted to a maximum and the transmitting power is decreased. The equation for the searching step of phase shifting value is as follows:

$$S_1 = [S_1^{\,0}, S_1^{\,1}, \cdots, S_1^{\,N_C-1}] = \underset{[S_1^{\,0}, S_1^{\,1}, \cdots, S_1^{\,N_C-1}]}{\arg\max} \left(\sum_{k=0}^{N_C-1} | S^k \cdot \mathrm{x}^k | \right), S^k \in V \qquad (4)$$

$$S_0 = [S_0^{\,0}, S_0^{\,1}, \cdots, S_0^{\,N_C-1}] = \underset{[S_0^{\,0}, S_0^{\,1}, \cdots, S_0^{\,N_C-1}]}{\arg\max} \left(\sum_{k=0}^{N_C-1} | S^k \cdot \mathrm{x}^k | \right), S^k \in V \qquad (5)$$

4 Simulation Results

Fig. 3 shows the bit error rate (BER) performance of the proposed and conventional Wi-Fi backscatter system. Additive white Gaussian noise (AWGN) is considered as the noise at the receiver. Also, ultra-wideband (UWB) channel is considered as the channel between transmitter and receiver. In according to the distance between transmitter and receiver, the power density is inversely proportional and the performance is degraded in according to the distance. In the simulation results, the proposed scheme has about 2.3 ~ 3dB SNR gain up to 30dB.



.

Fig. 3: Evaluation of BER performance

5 Conclusions

This paper proposes the phase shifting scheme in the Wi-Fi backscatter system. The conventional Wi-Fi backscatter system has very low reliability due to the limited power. In order to solve the problem, the proposed scheme adjusts phase offset among sub-carriers in the Wi-Fi packet. From the simulation results, it is confirmed that the proposed scheme has improved BER performance.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.B0126-15-1076, Development of non-powered technology combined with ambient RF energy harvesting and Backscatter data transfer)

References

[1] A.M. Oritz, D. Hussein, P. Soochang, et al, “The cluster betweeninternet of things and social networks: review and reserach challenges,” IEEE Internet of Things Journal., vol. 1, pp. 206–215, April, 2014.

[2] B. Kellogg, A. Parks, S. Gollakota, et al, “Wi-Fi backscatter: internet connectivity for RF-powered devices,” in ACM SIGCOMM, 2014, p. 607-618.

[3] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, High-speed Physical Layer in the 5GHZ Band, IEEE Std. 802.11a, 1999

Pineapple House: A User-Centered and Location-Aware Mobile House Rental App

Shu-Hao Chuang[1], **and Yung-Ting Chuang** [2*]

[1] Dept. of Occupational Safety and Health, Chung Hwa University of Medical Technology, Tainan, Taiwan
[2] Department of Information Management, National Chung Cheng University, Chia-Yi County, Taiwan

Abstract - *In Taiwan, when students need to look for available rental houses near their campus, there is no such platforms or websites that they can rely on and give them a very complete list of all vacant rental houses. Some of the reasons are due to the lack of computer skills of landlord, extra charges of listing fees, or difficulties of management. Meanwhile, since the emergence of digital technologies has pushed everyone to have their mobile devices, we have thus develop an user-centered and location-aware mobile housing management system, called Pineapple House, where the goal is to provide an user-friendly interfaces that assist both landlords and students to use their mobile phones to post and view house rental advertisements. We hope that such infrastructure will ultimately allow students to assist other students by increasing their social level of contact with others in their geographic areas.*

Keywords: Cloud Computing, Housing Management; Internet-Of-Things; Android Applications Development; Database Management; Web Applications

1 Introduction

Craigslist, one of the biggest classified advertisement websites in United States, helps people to post advertisement without any extra service fee in their community, such as posting jobs, housing, sale items, wanted item, donations, services, etc. However, these kinds of communal classifieds are not that popular in Taiwan. Furthermore, some of the communal classifieds require posting fees, but are not user-friendly (e.g. hard to manage or post/edit/remove advertisements). Therefore, when students need to look for available rental houses near their campus, they cannot only rely on the current available classified websites since these websites cannot give them a very complete list of all vacant rental houses. On the other hands, most of the students use PTT terminal-based Bulletin Board System [4] to gather more opinions or suggestions about the rental houses. However, since most of the landlords do not know how to use PTT to post their housing information, students still cannot obtain all of the rental houses from PTT. Meanwhile, with the rise of digital technologies, mobile phones have become ubiquitous in day-to-day life. Since the emergence of digital technologies has pushed users to learn how to use mobile devices, we have thus develop an user-centered and location-aware mobile housing management system, called Pineapple House, where it allows landlords and students to use their mobile phones to post and view house rental advertisements. Our Pineapple House app has the following features: 1) allows landlord to post their house rental advertisements from their mobile phones; 2) enables students to build their profile and customize their house rental preferences, as well as filtering search results and display the only rentals that they like; 3) keeps track of students' browsing histories; 4) allows students to put notes to particular house rental advertisement; 5) having contact feature that helps students to contact landlord via their mobile phones; 6) allows both landlords and students to send and receive messages with each other from the application; 7) enable users to compare between different house rentals; 8) enables landlord to upload multiple house images from their mobile phones; 9) provides location-aware and navigates students to find the designated house; and 10) reminds both students and landlord about their housing appointments' times and locations. Our goal is to have a system which is similar to Craigslist, but with more user-friendly features. We hope that such infrastructure will ultimately allow students to assist other students by increasing their social level of contact with others in their geographic areas.

2 Related Work

Some of the current house rental websites in Taiwan includes: 1) 591 [1], 2) Cloud Rental Housing Life[2], 3) and KiJiJi [3]. In these rental websites, they have bookmarks, browsing history, notes, contact landloards, send/receive messages, and share features. However, our Pineapple House does not only include the features of other rental websites, but also has comparison, location and navigation, offline access, appointment(s) management, housing preferences, and upload multiple pictures.

3 Design of Pineapple House mobile app

An overview of Pineapple House is shown in Fig. 1. Pineapple House consists of three distinct groups: (1) the web server that stores web pages, (2) the client with mobile application, and (3) the database which stores all the rentals. Arrows on the connecting lines indicate the direction of the communication flow. In our server side, we chooses Apache HTTP server application to store and deliver our web pages to the network, and our main website are all written in PHP

*Corresponding author: ytchuang@mis.ccu.edu.tw scripting language. In our client side, we uses: 1) HTML5 and Javascript languages to develop our main functions of the applications, 2) CSS3 and jQuery for addressing UI design and different sizes of mobile screens, and 3) uses Apache Cordova to integrate all of resources into the browser environment, thus creating a cross-platform app. Lastly, in database side, we uses MariaDB to maintain all of our house rental data.

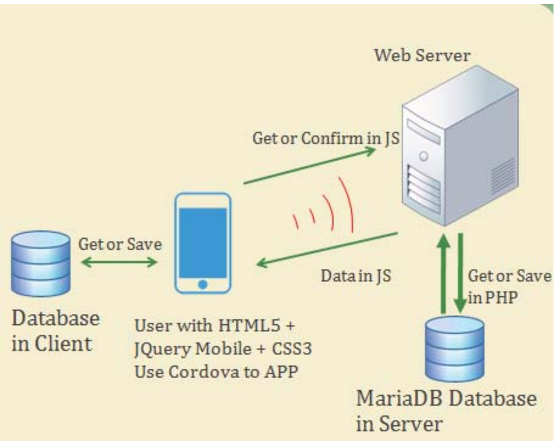


Figure 1. Pineapple House that consists server, database, and client.

4 User Interface of Pineapple House

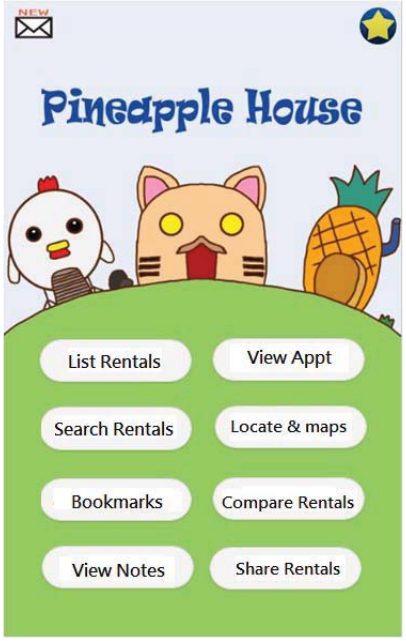


Figure 2. Screenshots of Home page

In this section, we describe some user interfaces and screenshots of the Pineapple House mobile app, where our system is suitable for any operating system. When a user starts the app, our system determines whether this mobile device has been logged in previously. If the user has not been logged in before, they can only view the house rental advertisements. If users wishes to further customize their rental advertisements, view their browsing history/bookmarks, create notes, contact landlords, share housing information, use reminder and calendar futures, they would need to login to the system. In addition, a new user can click on the "Create Account" button, in this case the system will navigate to the "Register" page to register for a new account.

Once the user successfully logs into the system, it will display the following eight options, as shown in Figure 1: 1) List rentals, 2) Search Rentals, 3) Bookmarks, 4) View Notes, 5) View Appointments, 6) Locate & Maps, 7) Compare rentals, and 8) Share Rentals. The explanations of the eight options are discussed below.

When the user clicks on the "List Rentals" button, the application would display all the non-expired rentals. Once user clicks a particular rental, they would be able to see the detailed view of a particular rental, detailed description, the ratings from other people, and the true location of the house. In addition, users can also use applications to filter out the house rentals, so that it would only apply users' profiles and only show the rentals that best matches users' rental preferences. When the user clicks on the "Search Rentals" button, the system allows users to search and retrieve rentals that matches users' search keywords. In "Bookmarks" page, the system shows the rentals that the users have previously bookmarked. In "View Notes" page, the user can create and view their notes which are associated with some house rentals. We believe that this is an useful feature since users may use these notes as some important resources for the future rental decisions. When the user clicks on "View appointment", our system allows them to set up an appointment time with the landlord, as well as reminding them about this appointment beforehand. In "Locate and Maps" feature, the system would detect users' current location, and would navigate them to the designated rental location. Moreover, the user can use "Compare Rentals" to compare different house rentals, as well as using "Share Rentals" to share their favorite house rentals with other people.

Acknowledgment

We would like to thank three undergraduate students, Po-Chia Chen, Cian-Yu Lin, and Ching-Hsuang Hsu, for developing the user interfaces of the Pineapple House app.

References

[1] 591 Housing Rental Bargain Platform. https://rent.591.com.tw/ Accessed April 26, 2016

[2] Cloud Rental Housing Life Network. . http://house.nfu.edu.tw/CCU/viewnews.html Accessed April 26, 2016

[3] KiJiJi Classified Website. http://kijibuy.com/. Accessed April 26, 2016.

[4] PTT Bulletin Board System, telnet://ptt.cc, http://www.ptt.cc/bbs/index.html. Access April 26, 2016

SESSION

LATE BREAKING PAPERS

Chair(s)

TBA

A Study on Privacy in the Internet of Things

George Santos, Lamarck Rocha, and Adriano Albuquerque
University of Fortaleza (UNIFOR), Av. Washington Soares, 1321 - Edson Queiroz, CEP 60811-905
Fortaleza, Ceará, Brazil, *Email: george.santos@edu.unifor.br,lamarckrocha@hotmail.com, adrianoba@unifor.br*

Abstract— *The Internet of Things (IoT), presents a vision of the future in which Internet users, computer systems and everyday objects that have action and sensing will be able to cooperate with unprecedented and economic benefits. With these enhancements, the concern about safety arises, since we are not only manipulating computers, but also objects of our everyday life. This paper describes a bibliographic study on a systematic review format in order to show security methods that ensure data privacy in environment the Internet of Things.*

Keywords: Internet of things, systematic review, security, privacy, software quality

1. Introduction

Since the beginning of communication networks, protection of the data is a very important question, but as the Internet gets more and more modern and improved the security concerns grew gradually, the factor that helps this growing is the privacy of data.

After the evolution of the conventional networks to the ”Internet of things” security became inseparable from life, that means that any interference, accidental or malicious, in the system of a car, pacemaker, or a nuclear reactor can be a risk for a person’s life. With this, we have to guarantee the IoT system quality referenced to security. According to ISO/IEC 25010:2011, this level of system quality protects information and data in a way that only allowed people can access, guaranteeing then its confidentiality[1] and automatically guaranteeing data privacy.

That being said, security must be debated during the whole device’s life cycle, from the initial project to the operational environment in all the levels[2]. This way, the factor of privacy security inside the ”Internet of things” is a problem that is in every environment, like smart homes, smart grids, smart city e healthcare[3].

This paper brings a systematic review about works that includes studies about privacy inside Internet of things, making an important report about which methods and techniques there are and are used to guarantee privacy on the environments inserted in the Internet of things.

This paper is organized this way, in Section 2 says what is a systematic review, in Section 3 we discuss the methodology of the systematic review, in Section 4 the chosen papers are categorized, in the Section 5 the chosen papers about privacy are analyzed, and in Section 6 the conclusion about the analysis is presented.

2. Systematic Review

A systematic reviews is one research way that uses literature as font of data about some theme and are very used to integrate informations of a pack of separated studies about some knowledge area, and it can present opposing or coincident results, also identify themes that need evidences[4].

According to [5] a systematic review has the point to reach a high level of scientific stringency. Unlike the usual reviews of literature, a systematic review is executed in a formal way, obeying one predetermined protocol. Then, its result becomes more believable, because systematic review uses a harsh methodology, that is also able to auditing and repetition.

So, the main reason to use a systematic review of literature is that it is particular and far. That being said, they are made according to a predetermined strategy, making the research uprightly, so it can be well evaluated and present great results.

3. Methodology

Like being said in this paper, the systematic review of literature that presents a process to identify, evaluate and interpret the available data to the question in research was used [6]. To this paper, the used models are the reported in Kitchenham and Charters that include three main stage: the definition of evaluation protocol, doing the review and reporting the evaluation[6].

The evaluation protocol that we use in this paper is compost for the following steps: (a) research question; (b) research strategy; (c) including and excluding criterion; (d) studying selection. Those steps are discussed in the following subsections.

3.1 Research Question

This systematic review has the point to show a general view about which techniques are used to guarantee privacy of data that is provided into the Internet of things. Inside this scene, research questions (RQ) were made to be answered for the systematic review of literature. Next, in the Schedule 1, it shows the research questions and the motivation of why they were made. And the answers for the related to the research questions like which techniques can guarantee

Table 1: Research questions to the systematic review

Research Questions	Motivation
RQ1 There are which methods in Internet of things to guarantee the data privacy?	To have a general view on the kinds of techniques that there are in the Internet of things to guarantee the data privacy.
RQ2 Which methods are used in the Internet of things?	To know which the most used technique there is in the Internet of things.
RQ3 How positive are the results of those methods?	Quantify and classify the methods that were already used to guarantee the data in the Internet of things.

the security and the privacy of data (RQ1), which of them are more used in different environments (RQ2) and how they can be more efficient and which can not solve exactly the guarantee of data privacy (RQ3). Those questions will help the researchers decide which questions must be the focus of their future researches.

3.2 Research Strategy

In a systematic review is very important to define the strategy of the research, because this may help the researchers to get a bigger number of relevant studies[6]. The research was made in digital libraries as it shows the schedule 2, through its research system. The research criterion used in this selection were:

a. To have a research system where logical expressions can be used or others ways to search.
b. To include database of the papers referencing to the security area.
c. The research systems must make searches in the whole text of the paper.
d. The research will only be evaluated if the searches from papers were made in digital libraries.
e. The paper only will be accepted if it's written in English.

Table 2: Digital Libraries

#	Digital Libraries	Research Terms	Web Address
DL1	ACM Digital Library	Keywords,Title of the paper	portal.acm.org
DL2	IEEE Explore	Keywords,Title of the paper,Abstract	www.ieee.org
DL3	ScienceDirect	Keywords,Title of the paper,Abstract	www.elsevier.com
DL3	Researchgate	Keywords,Title of the paper,Abstract	www.researchgate.net

During the search for publications, it is used the search terms that match the title of the paper, the keywords and the abstracts that are on the digital libraries. As[6] strategies were used to model the research terms. Those terms are:

a. To identify the main research terms;
b. To identify synonymous for big terms;
c. To use booleans ”AND” and ”OR” to gather the specific terms.

This way, the used terms to make the searches in the digital libraries were ”Privacy” and ”Internet of things”. After that, we used some synonymous and other terms to bring more relevance to the search:

(*privacy* ***OR*** *privateness* ***OR*** *confidentiality* **OR** secrecy) **AND** (*Internet of Things* **OR** IoT **OR** Healthcare **OR** Smart)

3.3 Including and Excluding Criterion

According to[6] the including and excluisng criterion must be well defined, because it's after that we can get right evidences that may help to get results from the studies. So, as it's presented on the Figure 1, we created three kinds of selections:
a. Selection and discontinuance of the papers: this preliminary section were made among the papers presented by the research systems in the digital libraries trough the defined expressions that were reported previously. The papers that passed for this selection were filled on a database.
b. Selection of usable papers (first selection): Even if the preliminary selection of the papers were made, we have to make another selection using new criterion to exclude papers that are not so relevant for the search, because in the first selection only the search terms are used. The criterion that were used are: which methods can guarantee the privacy in the Internet of things and which papers that make a analysis and search about how to guarantee the privacy of the informations and the users.
c. Selection of usable papers (second selection): Right after the first selection, there still may be papers that don't have information to be used that are relevance to the search that is going to be studied. So, the criterion to be used only the papers that have technique that is usable in the guarantee of privacy data.
After the previously discussed selections being done, we'll construct one schedule with the papers that have methods to guarantee privacy. Because when they get excluded, only the papers that have coherency to the criterion descried during the selection.

3.4 Selection of the Studies

According to the Figure 1, selections were made in the papers of the research, until we reach the number of relevant papers for the research studying. During the search for the papers in the digital libraries using the terms that match each other in the subsection 3.2, 300 papers that have attributes discussed in the subsection 3.3, then only 169 papers get to stay.

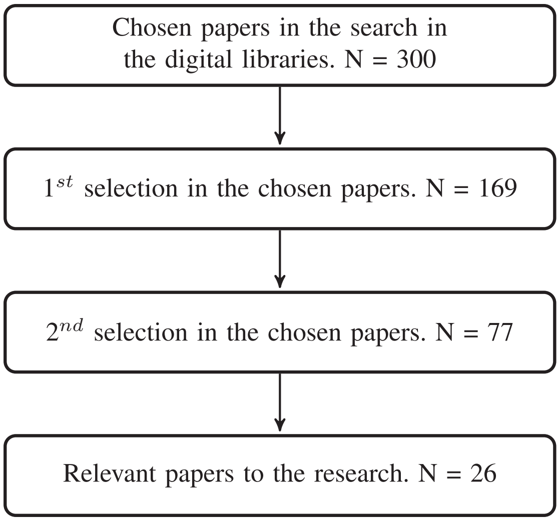

Chosen papers in the search in the digital libraries. N = 300
1^{st} selection in the chosen papers. N = 169
2^{nd} selection in the chosen papers. N = 77
Relevant papers to the research. N = 26

Fig. 1: Selection Studies

Even after done the first selection, there may be some papers that don't have informations to be used[7]. So, the second selection using new criterion more relevant, reviewing again the 169 papers that we got in the first selection until the number of relevant papers to our studying reached 77. Even if selections were done before, we made another selection using the snowballing technique[8] that consists in take the references from the 77 papers hoping we find potential studies to the search. That being done, we got 26 relevant papers to the systematic review of literature.

4. Categorizing the Papers

We categorize the papers in the following way wishing improve the studies referenced to the chosen papers.

- Really Relevant Paper (RRP)
- Relevant Paper (RP)
- Few Relevant Paper (FRP)

According to the Schedule 3, really relevant papers(RRP) are those that have very well-done analysis and search, showing which technique used to preserve the privacy of the informations on the environments in the Internet of things. The relevant papers(RP) are those that their techniques didn't get a very good outcome to guarantee the data privacy. And the few relevant papers(FRP) are those that make an general analysis of the security in the Internet of things and say in their sections the methods to guarantee privacy.

Table 3: Categorizing the chosen papers

#	Title	Authors	Font	Year	Category
AS1	Negotiation-based privacy preservation scheme in Internet of Things platform	Ukil A Bandyopadhyay S Joseph J Banahatti V Lodha S	ACM	2012	RRP
AS2	Anomaly detection and privacy preservation in Cloud-Centric Internet of Things	Butun I Kantarci B Erol-kantarci M	IEEE	2015	RP
AS3	The Internet of Things for Health Care: A Comprehensive Survey	Kwak K	IEEE	2015	FRP
AS4	Negotiation-based privacy preservation scheme in Internet of Things platform	Ukil A Bandyopadhyay S Joseph J Banahatti V Lodha S	ACM	2012	RRP
AS5	Anomaly detection and privacy preservation in Cloud-Centric Internet of Things	Butun I Kantarci B Erol-kantarci M	IEEE	2015	RP
AS6	The Internet of Things for Health Care: A Comprehensive Survey	Kwak K	IEEE	2015	FRP
AS7	Privacy Preserving Data Analysis in Mental Health Research	Li J Li X	IEEE	2015	RRP
AS8	Security and privacy challenges in industrial internet of things	Darmstadt T	IEEE	2015	FRP
AS9	Security Analysis and proposal of new Access Control model in the Internet of Thing	Ouaddah A Anas I Elkalam A Ouahman A	IEEE	2015	RRP

#	Title	Authors	Font	Year	Category
AS10	Big Data Privacy in the Internet of Things Era	Perera C Zomaya A	IEEE	2015	RP
AS11	Location Privacy in the Era of the I nternet of Things and Big Data Analytics	Minch R	IEEE	2015	RP
AS12	Smart Meter Privacy for Multiple Users in the Presence of an Alternative Energy Source	Gomez-Vilardebo J Gunduz D	IEEE	2015	RRP
AS13	SeCoMan: A Semantic-Aware Policy Framework for Developing Privacy-Preserving and Context-Aware Smart Applications	Huertas Celdran A Garcia Clemente F Gil Perez M Martinez Perez G	IEEE	2014	RRP
AS14	A Security Information Transmission Scheme for Internet of Things	Li Z Fu J Fan W Long R	IEEE	2013	FRP
AS15	Trustworthy Infrastructure Services for a Secure and Privacy-respecting Internet of Things	Gessner D Olivereau A Segura A Serbanati A	IEEE	2012	RRP
AS16	Preference-based Privacy Protection Mechanism for the Internet of Things	Tao H Peiran W	IEEE	2010	RRP
AS17	The dark side of the interconnection: security and privacy in the Web of Things	Catuogno L Turchi S	IEEE	2015	RP
AS18	On Security and Privacy Issues of Fog Computing supported Internet of Things Environment	Lee K Kim D Ha D Rajput U Oh H	IEEE	2014	RP
AS19	Privacy and Security in Internet of Things and Wearable Devices	Das M	IEEE	2015	RRP
AS20	Review of Internet of Things in Development of Smart Cities with Data Management and Privacy	Burange A Misalkar H	IEEE	2015	RRP
AS21	Security, privacy and trust in Internet of Things: The road ahead	Sicari S Rizzardi A Grieco L Coen-Porisini A	Science Direct	2014	FRP
AS22	Internet of things: Privacy issues revisited	Weber R	Science Direct	2015	RP
AS23	Privacy in Internet of Things: A Model and Protection Framework	Samani A Ghenniwa H Wahaishi A	Science Direct	2015	Rrp
AS24	Privacy protection in pervasive systems:State of the art and technical challenges	Bettini C Riboni D	Science Direct	2015	RP
AS25	Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications	Al Ameen M Liu J Kwak K	Science Direct	2012	FRP
AS26	A multi-agent approach: To preserve user information privacy for a pervasive and ubiquitous environment	Chandramohan D Sathian D Rajaguru D Vengattaraman T Dhavachelvan P	Science Direct	2015	RRP

5. Result Analysis

Based upon the schedule 3, we'll start analyzing the papers taking into accounts their categorizing. The really relevant papers show techniques that had positive acceptation in simulations, according what was said previously, they are techniques that have positive impact in the security e guarantee of the information's security. One of those papers shows one framework of preservation of the privacy in the environment in the Internet of things, just with one hiding tool to the informations [3].Another really relevant paper talks about a privacy analyst for smart measurers through informations and statistics got from real measurers[9].

Still analyzing the RRP, we have papers that discuss techniques of how to guarantee privacy for e-health as said[10] where it's developed one preservation analyst of data privacy that is used to guarantee informations of mental-health patients. In the Smart Cities' environment, we have a paper that comments about a data manager and privacy, protecting informations kept in the smart city[11]. In the others really relevant papers, it's discussed techniques that are used in the environment of the Internet of things in a general term.

The relevant papers show preservation techniques of the privacy referenced to the big data of the Internet of things, one of those papers discuss that there is in the Internet of things one retirement of data in big grade, but in the extraction time of that data how to guarantee the privacy of the users[12] and in the other paper, it's discussed how to preserve user's information, because the localization of the information in the Internet of things is developed in big quantities, making it reviewed questions research about privacy [13].

And in others two relevant papers, they're surveys that talk in their papers about privacy in a general way, presenting important questions about privacy. One of them says that in the environment of Internet of things there are unique questions about the privacy protection[14] and the need of an action that can minimize even more this problem of guarantee the informations privacy.

The last analysis that has been done is referenced to the few relevant papers. Those papers discuss in a general way the question about security and privacy of the informations in the Internet of things or show little impact techniques to guarantee privacy. Among them, we have one paper that talks about questions of Internet of things in the industrial area and the security problems that can be prejudicial[15]. In the other paper comments about general questions security in the Internet of things, because in all the security requirements that's always mentioned the privacy[16].

6. Conclusion

Following the current growing of the services of the Internet of things it's need to guarantee beyond the security, the privacy of the users and the informations [16]. Then, privacy has been a important worry in e-health, for treating patients' informations[17], also in smart city, smart home, and others.

That said, the main point of this paper was to make a systematic review of literature referenced to the subject of how to guarantee the privacy of data in the environment of the Internet of things. Right after, the analysis of the papers was noticed that they have techniques and methods to guarantee the privacy of people's important data. But after the ending of the studying it's noticed that there's a large camp of studies opened to be explored for researchers in the environments that are localized inside the Internet of things.

Although, there are many directions to future research works in this area, one of them would be the creation of a model that guarantees privacy since the user's data to be sent gets inserted, during the dispatch, until the addressee received it. This way, after this report there are a lot of studies to be done for we have privacy to our information in this new technology area.

References

[1] ISO/IEC, "ISO/IEC 25010 - Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models," Tech. Rep., 2011.

[2] R. Billure, V. M. Tayur, and V. Mahesh, "Internet of things-a study on the security challenges," in *Advance Computing Conference (IACC), 2015 IEEE International*. IEEE, 2015, pp. 247–252.

[3] A. Ukil, S. Bandyopadhyay, J. Joseph, V. Banahatti, and S. Lodha, "Negotiation-based privacy preservation scheme in internet of things platform," in *Proceedings of the First International Conference on Security of Internet of Things*. ACM, 2012, pp. 75–84.

[4] R. F. Sampaio and M. C. Mancini, "Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica," *Braz. J. Phys. Ther.(Impr.)*, vol. 11, no. 1, pp. 83–89, 2007.

[5] B. Kitchenham, "Procedures for performing systematic reviews," *Keele, UK, Keele University*, vol. 33, no. 2004, pp. 1–26, 2004.

[6] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering. 2007," *URL http://www. dur. ac. uk/ebse/resources/Systematic-reviews-5-8. pdf*, 2007.

[7] E. E. E. de Arruda Paiva, "Uma abordagem de apoio à avaliação e melhoria da produtividade de desenvolvedores de software," *UNIFOR*, 2011.

[8] D. Budgen, M. Turner, P. Brereton, and B. Kitchenham, "Using mapping studies in software engineering," in *Proceedings of PPIG*, vol. 8, 2008, pp. 195–204.

[9] A. Ukil, S. Bandyopadhyay, and A. Pal, "Privacy for iot: Involuntary privacy enablement for smart energy systems," in *Communications (ICC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 536–541.

[10] J. Li and X. Li, "Privacy preserving data analysis in mental health research," in *Big Data (BigData Congress), 2015 IEEE International Congress on*. IEEE, 2015, pp. 95–101.

[11] A. W. Burange and H. D. Misalkar, "Review of internet of things in development of smart cities with data management & privacy," in *Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in*. IEEE, 2015, pp. 189–195.

[12] C. Perera, R. Ranjan, L. Wang, S. U. Khan, and A. Y. Zomaya, "Big data privacy in the internet of things era," *IT Professional*, vol. 17, no. 3, pp. 32–39, 2015.

[13] R. P. Minch, "Location privacy in the era of the internet of things and big data analytics," in *System Sciences (HICSS), 2015 48th Hawaii International Conference on*. IEEE, 2015, pp. 1521–1530.

[14] R. H. Weber, "Internet of things: Privacy issues revisited," *Computer Law & Security Review*, vol. 31, no. 5, pp. 618–627, 2015.

[15] A.-R. Sadeghi, C. Wachsmann, and M. Waidner, "Security and privacy challenges in industrial internet of things," in *Proceedings of the 52nd Annual Design Automation Conference*. ACM, 2015, p. 54.

[16] S. Sicari, A. Rizzardi, L. Grieco, and A. Coen-Porisini, "Security, privacy and trust in internet of things: The road ahead," *Computer Networks*, vol. 76, pp. 146–164, 2015.

[17] M. Anwar, J. Joshi, and J. Tan, "Anytime, anywhere access to secure, privacy-aware healthcare services: Issues, approaches and challenges," *Health Policy and Technology*, vol. 4, no. 4, pp. 299–311, 2015.

Defending Against Zero-day Polymorphic Worms Using Hidden Markov Models (HMMs)

Mohssen M. Z. E. Mohammed, Neco Ventura
University of Garden City , Sudan
Department of Electrical Engineering, University of Cape Town, South Africa

Emails: m_zin44@hotmail.com; neco.ventura@uct.ac.za

Abstract - *Internet worms pose a major threat to the Internet infrastructure and their destruction causes loss of millions of dollars. Therefore, the network must be protected as much as possible to avoid losses. In this paper we propose automatic and accurate system for Zero-day polymorphic worms. We have designed a novel double-honeynet system, which is able to detect new worms that have not been seen before. We are using Substring Extraction Algorithm (SEA) and Hidden Markov Models (HMMs) for detecting Zero-day polymorphic worm.*

Keywords: Polymorphic Worms, Honeynet, IDSs.

1 Introduction

A computer worm is a malicious program that spreads automatically among hosts on a network by exploiting various vulnerabilities present on those hosts. A computer worm differs from a computer virus in that a computer worm can run itself. A virus needs a host program to run, and the virus code runs as part of the host program. A polymorphic worm is a worm that changes its appearance with every instance [1]. It has been shown that multiple invariant substrings must often be present in all variants of worm payload. These substrings typically correspond to protocol framing, return addresses, and in some cases, poorly obfuscated code [8].

Intrusion detection is the process of monitoring computers or networks for unauthorized entrance, activity, or file modification. IDS can also be used to monitor network traffic, thereby detecting if a system is being targeted by a network attack such as a denial of service attack. There are two basic types of intrusion detection: host-based and network-based. Each has a distinct approach to monitoring and securing data. Host-based IDSs examine data held on individual computers that serve as hosts, while network-based IDSs examine data exchanged between computers. There are two basic techniques used to detect intruders: Anomaly Detection and Misuse Detection (Signature Detection). Anomaly Detection is designed to uncover abnormal patterns of behavior, the IDS establishes a baseline of normal usage patterns, and anything that widely deviates from it gets flagged as a possible intrusion. Misuse detection (signature detection), commonly called Signature Detection, uses specifically known patterns of unauthorized behavior to predict and detect subsequent similar attempts. These specific patterns are called signatures [15, 16].

Our research is based on Honeypot technique. Developed in recent years, honeypot is a monitored system on the Internet serving the purpose of attracting and trapping attackers who attempt to penetrate the protected servers on a network. Honeypots fall into two categories. A high-interaction honeypot such as (Honeynet) operates a real operating system and one or multiple applications. A low-interaction honeypot such as (Honyed) simulates one or multiple real systems. In general, any network activities observed at honeypots are considered suspicious [1, 10].

This paper is organized as follows: section 2 discusses the related work regarding automated detection for Zero-day polymorphic worms. Section 3 reviews anatomy of polymorphic worms. Section 4 introduces the proposed system architecture to address the problems faced by current automated signature systems. Section 5 discusses signature generation using Hidden Markov Model (HMM). Section 6 concludes the paper.

2 Related work

Honeypots are an excellent source of data for intrusion and attack analysis. Levin et al. described how honeypot extracts details of worm, exploits that can be analyzed to generate detection signatures [3]. The signatures are generated manually.

One of the first systems proposed was Honeycomb developed by Kreibich and Crowcroft. Honeycomb generates signatures from traffic observed at a honeypot via its implementation as a Honeyd [5] plugin. The longest common substring (LCS) algorithm, which looks for the longest shared byte sequences across pairs of connections, is at the heart of Honeycomb. Honeycomb generates signatures consisting of a single, contiguous substring of a worm's payload to match all worm instances. These signatures, however, fail to match all polymorphic worm instances with low false positives and low false negatives.

Kim and Karp [6] described the Autograph system for automated generation of signatures to detect worms. Unlike Honeycomb, Autograph's inputs are packet traces from a DMZ that includes benign traffic. Content blocks that match "enough" suspicious flows are used as input to COPP, an algorithm based on Rabin fingerprints that searches for repeated byte sequences by partitioning the payload into content blocks. Similar to Honeycomb, Autograph generates signatures consisting of a single, contiguous substring of a worm's payload to match all worm instances. These signatures, unfortunately, fail to match all polymorphic worm instances with low false positives and low false negatives.

S. Singh, C. Estan, G. Varghese, and S. Savage [7] described the Earlybird system for generating signatures to detect worms. This system measures packet-content prevalence at a single monitoring point such as a network DMZ. By counting the number of distinct sources and destinations associated with strings that repeat often in the payload, Earlybird distinguishes benign repetitions from epidemic content. Earlybird, also like Honeycomb and Autograph, generates signatures consisting of a single, contiguous substring of a worm's payload to match all worm instances. These signatures, however, fail to match all polymorphic worm instances with low false positives and low false negatives.

New content-based systems like Polygraph, Hamsa and LISABETH [8, 11 and 12] have been deployed. All these systems, similar to our system, generate automated signatures for polymorphic worms based on the following fact: there are multiple invariant substrings that must often be present in all variants of polymorphic worm payloads, even if the payload changes in every infection. All these systems capture the packet payloads from a router, so in the worst case, these systems may find multiple polymorphic worms but each of them exploits a different vulnerability from each other. So, in this case, it may be difficult for the above systems to find invariant contents shared between these polymorphic worms because they exploit different vulnerabilities. The attacker sends one instance of a polymorphic worm to a network, and this worm in every infection automatically attempts to change its payload to generate other instances. So, if we need to capture all polymorphic worm instances, we need to give a polymorphic worm chance to interact with hosts without affecting their performance. So, we propose a new detection method "Double-honeynet" to interact with polymorphic worms and collect all their instances. The proposed method makes it possible to capture all polymorphic worm instances and then forward these instances to the Signature Generator which generates signatures, using a particular algorithm.

An Architecture for Generating Semantics-Aware Signatures by Yegneswaran, J. Giffin, P. Barford, and S. Jha [9] described Nemean, Nemean's incorporates protocol semantics into the signature generation algorithm. By doing so, it is able to handle a broader class of attacks. The coverage of Nemean is wide which makes us believe that our system is better in dealing with polymorphic worms specially.

An Automated Signature-Based Approach against Polymorphic Internet Worms by Yong Tang and Shigang Chen[10] described a system to detect new worms and generate signatures automatically. This system implemented a double-honeypots (inbound honeypot and outbound honeypot) to capture worms payloads. The inbound honeypot is implemented as a high-interaction honeypot, whereas the outbound honeypot is implemented as a low-interaction honeypot. This system has limitation. The outbound honeypot is not able to make outbound connections because it is implemented as low-interaction honeypot which is not able to capture all polymorphic worm instances. Our system overcomes this disadvantage by using double-honeynet (high-interaction honeypot), which enables us to make unlimited outbound connections between them, so we can capture all polymorphic worm instances.

3 Anatomy of Polymorphic worms

As stated in [4] a polymorphic attack is an attack that is able to change its appearance with every instance. Thus, there may be no fixed or predictable signature for the attack. As a result, it may evade detection because most current intrusion detection systems and antivirus systems are signature-based. Exploit mutation and shellcode polymorphism are two common ways to generate polymorphic attacks. In general, there are three components in a polymorphic attack:

3.1 Attack vector

An attack vector is used for exploiting the vulnerability of the target host. Certain parts of the attack vector can be modified to create mutated but still valid exploits. There might still be certain parts, called the invariant, of the attack vector that have to be present in every mutant for the attack to work. If the attack invariant is very small and exists in the normal traffic, then an IDS may not be able to use it as a signature because it will result in a high number of false positives.

3.2 Attack body

The code that performs the intended malicious actions after the vulnerability is exploited. Common techniques to achieve attack body (shellcode) polymorphism include register shuffling, equivalent instruction substitution, instruction reordering, garbage insertions, and encryption. Different keys can be used in encryption for different instances of the attack to ensure that the byte sequence is different every time.

3.3 Polymorphic Decryptor

This section contains the part of the code that decrypts the shellcode. It decrypts the encrypted attack body and transfers control to it. Polymorphism of the decryptor can be achieved using various code obfuscation techniques.

4 Double-honeynet system

The purpose of Double-honeynet system is to detect unknown (i.e., previously unreported) worms automatically. A key contribution of this system is the ability of distinguishing worm activities from normal activities without any involvement of experts in the field.

Figure 1 shows the main components of the Double-honeynet system. Firstly, the incoming traffic goes through the Local Router which samples the unwanted inbound connections and redirects the samples' connections to Honeynet 1. As the redirected packets pass through the Local Router, Packet Capture (PCAP) library is used to capture the packets and then to analyze their payloads to contribute to the signature generation process.

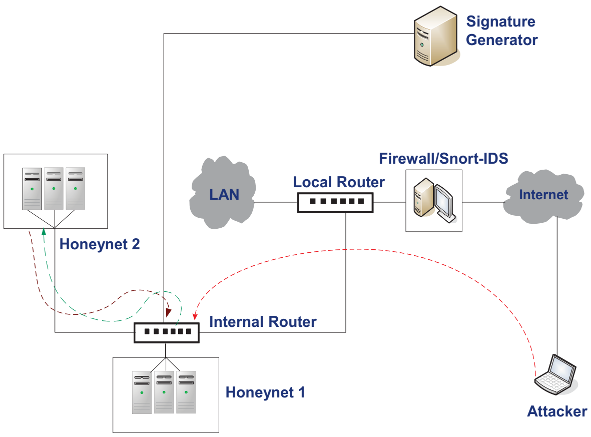


Figure 1: System Architecture

The Local Router is configured with publicly-accessible addresses, which represent wanted services. Connections made to other addresses are considered unwanted and redirected to Honeynet 1 through the Internal Router. Once Honeynet 1 is compromised, the worm will attempt to make outbound connections to attack another network. The Internal Router is implemented to separate the Double-honeynet from the Local Area Network (LAN). This Router intercepts all outbound connections from Honeynet 1 and redirects those to Honeynet 2, which perfroms the same task forming a loop. The looping mechanism allows us to capture different instances of the polymorphic worm as it mutates on each loop-iteration. We stop the loop after a considerable amount of time in order to collect polymorphic worms.

5 Signature Generation Using Hidden Markov Model (HMM)

In this section we show how we could use the Hidden Markov Model (HMM) to generate signatures for Zero-day polymorphic worms. In this paper we will combine the Hidden Markov Model (HMM) with the Substring Extraction Algorithm (SEA) to generate signatures for Zero-day polymorphic worms. The goal of the SEA is to extract substrings from one of the polymorphic worm instances, then use the HMM to find a signature from the remaining instances using the extracted substrings by using SEA. Before discussing the HMM, below we will discuss the SEA.

5.1 Substring Extraction Algorithm (SEA)

SEA is a string matching algorithm. String matching [2] is an important subject in the wider domain of text processing. String matching algorithms are basic components used in implementations of practical software used in most of the available operating systems. Moreover, they emphasize programming methods that serve as paradigms in other fields of computer science (e.g., system or software design). Finally, they also play an important role in theoretical computer science by providing challenging problems.

String matching generally consists of finding a substring (called a pattern) within another string (called the text). The pattern is generally denoted as,

$$x= x[0..m-1]$$

whose length is m and the text is generally denoted as

$$y=y[0..n-1]$$

whose length is n. Both the strings-pattern and text are built over a finite set of characters which is called the alphabet and denoted by Σ, whose size is denoted by σ.

The string matching algorithm plays an important role in network intrusion detection systems (IDS), which can detect malicious attacks and protect the network systems. In fact, at the heart of almost every modern intrusion detection system, there is a string matching algorithm. This is a very crucial technique because it allows detection systems to base their actions on the content that is actually flowing to a machine. From a vast number of packets, the string identifies those packets that contain data, matching the fingerprint of a known attack. Essentially, the string matching algorithm compares the set of strings in the rule-set with the data seen in the packets, which flow across the network.

Our work uses Substring Extraction Algorithm (SEA) and Hidden Markov Model (HMM) to generate signatures for polymorphic worm attacks. The SEA aims at extracting substrings from a polymorphic worm, whereas HMM aims at finding out multiple invariant substrings that are shared between polymorphic worm instances to use them as signatures.

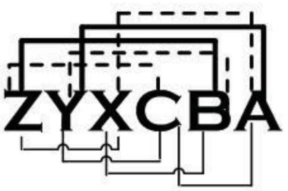
5.1.1 Substring Extraction Algorithm (SEA)

In this subsection, we show how our proposed Substring Extraction Algorithm (SEA) is used to extract substrings from one of the polymorphic worm variants that are collected by the Double-honeynet system.

Let us assume that we have a polymorphic worm A, that has n instances (A1,..., An) and Ai has length Mi for i=1,…,n. Assume that A1 selected to be the instance from which we extract substrings and the A1 string contains a1 a2 a3... am1. Let, X be the minimum length of a substring to be extracted from A1. The first substring from A1 with length X, is (a1 a2 ... ax). Then, we shift one position to the right to extract a new substring, which will be (a2 a3... ax+1). Continuing this way, the last substring from A1 will be (am1-X+1... am1). In general, if instance Ai has length equal to M, and let a minimum length of the substring that we are going to extract from A1 equals to X, then the Total Number of Substrings (TNS) that will be extracted from Ai could be obtained by this equation:

$$\text{TNS } (A_i) = \text{M-X+1}$$

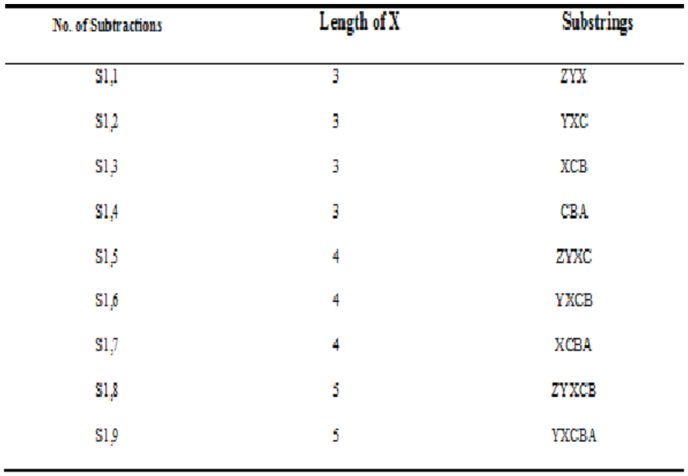
The next step is to increase X by one and start new substrings extraction from the beginning of A1. The first substring will be (a1 a2 ... ax+1). The substrings extraction will continue satisfying this condition, X< M.
Figure 2 and Table 1 show all substrings extraction possibilities using the proposed Substring Extraction Algorithm (SEA) from the string ZYXCBA assuming the minimum length of X is equal to three.



Thin solid line X=3, The substrings are ZYX, YXC, XCB, CBA
Dashed line X=4, The substrings are ZYXC, YXCB, XCBA
Thick solid line X=5, The substrings are ZYXCB, YXCBA

Figure 2: Extraction Substrings

Table 1: Substrings Extraction



No. of Subtractions	Length of X	Substrings
S1,1	3	ZYX
S1,2	3	YXC
S1,3	3	XCB
S1,4	3	CBA
S1,5	4	ZYXC
S1,6	4	YXCB
S1,7	4	XCBA
S1,8	5	ZYXCB
S1,9	5	YXCBA

The output of the SEA will be used by HMM.

5.2 Hidden Markov Model (HMM)

In this subsection we will discuss the HMM. The Hidden Markov Model (HMM) is a powerful statistical tool for modeling generative sequences that can be characterized by an underlying process generating an observable sequence. HMMs have found applications in many areas of interest in signal processing, and in particular speech processing, but have also been applied with success to low level NLP (Natural language processing) tasks such as part-of-speech tagging, phrase chunking, and extracting target information from documents. Andrei Markov gave his name to the mathematical theory of Markov processes in the early twentieth century [2], but it was Baum and his colleagues that developed the theory of HMMs in the 1960s [2].

Figure 3 shows an example of a Markov process. The model presented describes a simple model for a stock market index. The model has three states, Bull, Bear and Even, and three index observations up, down, unchanged. This model is a finite state automaton, with probabilistic transitions between states. Given a sequence of observations, example: up-down-down we can easily verify that the state sequence that produced those observations was: Bull-Bear-Bear, and the probability of the sequence is simply the product of the transitions, in this case 0.2 × 0.3 × 0.3.

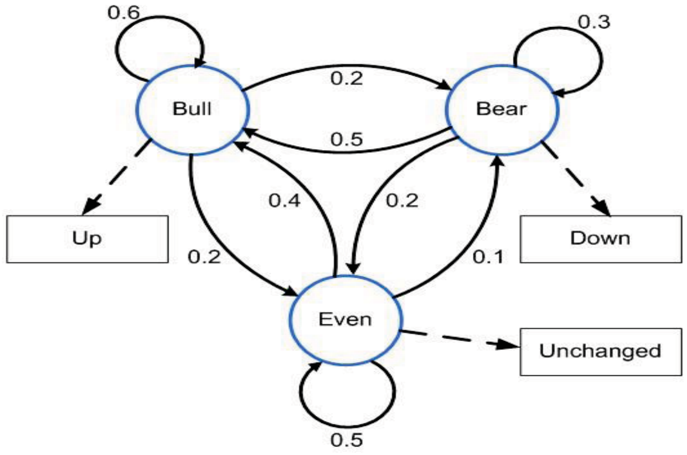


Figure 3: Markov process example.

Figure 4 shows an example of how the previous model can be extended into a HMM. The new model now allows all observation symbols to be emitted from each state with a finite probability. This change makes the model much more expressive and able to better represent our intuition, in this case, that a bull market would have both good days and bad days, but there would be more good ones. The key difference is that now if we have the observation sequence up-down-down then we cannot say exactly what state sequence produced these observations and thus the state sequence is '*hidden*'. We can however calculate the probability that the model produced the sequence, as well as which state sequence was most likely to have produced the observations [2].

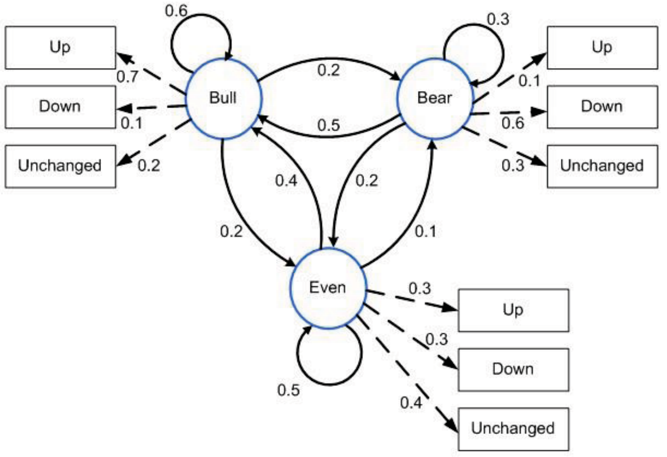


Figure 4: Hidden Markov model example.

Let us find a bit more details about Hidden Markov Models (HMM).

HMMs are similar to state-space models in that the sequence of observations is assumed to have been generated from a sequence of underlying hidden states. The main difference is that in HMMs, the state is assumed to be discrete rather than a continuous random vector. Let s_t denote the hidden state of an HMM at time *t*. We assume that s_t can take discrete values in {1, . . . , K }. The model can again be written as in:

$$P(s_{1:T}, \mathbf{y}_{1:T}|\theta) = \prod_{t=1}^{T} P(s_t|s_{t-1},\theta)P(\mathbf{y}_t|s_t,\theta)$$

where, $P(s_1|s_0,\theta)$ is simply some initial distribution over the K settings of the first hidden state; we can call this discrete distribution π, represented by a K × 1 vector. The state-transition probabilities $P(s_t|s_{t-1},\theta)$ are captured by a K × K transition matrix A, with elements, $A_{ij} = P(s_t = i|s_{t-1} = j,\theta)$. The observations in an HMM can be either continuous or discrete. For continuous observations y_t one can for example choose a Gaussian density; thus $P = (y_t|s_t = i,\theta)$ would be a different Gaussian for each choice of i ∈ {1, … , K }. This model is the dynamical generalization of a mixture of Gaussians. The marginal probability at each point in time is exactly a mixture of K Gaussians—the difference is that which component generates data point y_t and which component generated y_{t-1} are not independent random variables, but certain combinations are more and less probable depending on the entries in *A*. For y_t a discrete observation, let us assume that it can take on values {1, . . . , L}. In that case the output probabilities $P(y_t|s_t,\theta)$ can be captured by an L × K emission matrix, E.

The model parameters for a discrete-observation HMM are θ = (π, A, E). Maximum likelihood learning of the model parameters can be approached using the EM algorithm, which in the case of HMMs is known as the Baum-Welch algorithm. The E step involves computing $Q(s_t)$ and $Q(s_t, s_{t+1})$ which are marginals of $Q(s_{1:T}) = P(s_{1:T}|y_{1:T},\theta)$. These marginals are computed as part of the forward–backward algorithm which, as the name suggests, sweeps forward and backward through the time series, and applies Bayes rule efficiently using the Markov conditional independence properties of the HMM, to compute the required marginals. The M steps of HMM learning involve re-estimating π, A, and E by adding up and normalizing expected counts for transitions and emissions that were computed in the E step [2].

6 Conclusion

We have proposed automated detection for Zero-day polymorphic worms using Hidden Markov Models (HMMs). We are using the following steps to generate signature for Zero-day polymorphic worms: First: collecting Zero-day polymorphic worms samples by using the Double-honeynet. Second, we used the SEA to extract substrings from one instance of a polymorphic worm. Third, we used the HMM to find invariant substring that are shared between the instances. The main objectives of this research are to reduce false alarm rates and generate high quality signatures for polymorphic worms.

7 References

[1] L. Spitzner. “Honeypots: Tracking Hackers”. Addison Wesley Pearson Education: Boston, 2002.

[2] D. Gusfield. “Algorithms on Strings, Trees and Sequences”. Cambridge University Press: Cambridge, 1997.

[3] J. Levine, R. La Bella, H. Owen, D. Contis ,and B. Culver. “The use of honeynets to detect exploited systems across large enterprise networks“; Proc. of 2003 IEEE Workshops on Information Assurance, New York, Jun. 2003, pp. 92- 99.

[4] P. Fogla M. Sharif R. Perdisci O. Kolesnikov W. Lee. “Polymorphic Blending Attacks”; Proc. of the 15th conference on USENIX Security Symposium, Vancouver, B.C., Canada, 2006.

[5] C. Kreibich and J. Crowcroft. “Honeycomb–creating intrusion detection signatures using honeypots“; Workshop on Hot Topics in Networks (Hotnets-II), Cambridge, Massachusetts, Nov. 2003.

[6] H.-A. Kim and B. Karp. “Autograph: Toward automated, distributed worm signature detection“; Proc. of 13 USENIX Security Symposium, San Diego, CA, Aug., 2004.

[7] S. Singh, C. Estan, G. Varghese, and S. Savage. “Automated worm fingerprinting“; Proc. Of the 6th conference on Symposium on Operating Systems Design and Implementation (OSDI), Dec. 2004.

[8] James Newsome, Brad Karp, and Dawn Song.“ Polygraph: Automatically generating signatures for polymorphic worms“; Proc. of the 2005 IEEE Symposium on Security and Privacy, pp. 226 – 241, May 2005.

[9] V. Yegneswaran, J. Giffin, P. Barford, and S. Jha. “An architecture for generating semantics-aware signatures“; Proc. of the 14th conference on USENIX Security Symposium, 2005.

[10] Yong Tang, Shigang Chen. “ An Automated Signature-Based Approach against Polymorphic Internet Worms“; IEEE Transaction on Parallel and Distributed Systems, pp. 879-892 July 2007.

[11] Zhichun Li, Manan Sanghi, Yan Chen, Ming-Yang Kao and Brian Chavez. Hamsa. “Fast Signature Generation for Zero-day Polymorphic Worms with Provable Attack Resilience“;Proc. of the IEEE Symposium on Security and Privacy, Oakland, CA, May 2006.

[12] Lorenzo Cavallaro, Andrea Lanzi, Luca Mayer, and Mattia Monga. “LISABETH: Automated Content-Based Signature Generator for Zero-day Polymorphic Worms”; Proc. of the fourth international workshop on Software engineering for secure systems, Leipzig, Germany, May 2008.

[13] Mohssen M. Z. E. Mohammed, H. Anthony Chan, Neco Ventura. “Honeycyber: Automated signature generation for zero-day polymorphic worms“; Proc. of the IEEE Military Communications Conference, MILCOM, 2008.

[15] Snort – The de facto Standard for Intrusion Detection/Prevention. Available: http://www.snort.org, March 2016.

[16] Bro Intrusion Detection System. Available: http://www.bro-ids.org/, March, 2016.

[17] Al-Sakib Khan Pathan, Mohssen M. Z. E. Mohammed," Building Customer Trust in Cloud Computing with an ICT-Enabled Global Regulatory Body", Wireless Pers Commun, DOI 10.1007/s11277-015-2729-z.

No Place to Hide: A Review of Privacy Toward a Safer Internet of Things

T. Ray Campbell [1] Felix Akinaladejo [2]

[1] College of Engineering & Information Sciences, DeVry University, Charlotte, NC, USA

[2] Faculty of Engineering & Computing, University of Technology, Kingston, Jamaica

Abstract - *The Internet of Things (IoT) has been described as 'everything' connecting in real-time; a connectedness among devices anytime, anywhere and in real-time. Many technologies for the protection of privacy have been proposed but there still exist many flaws; flaws based on the design of the IoT. This paper introduces a discussion on the privacy and threats surrounding security and privacy of IoT platforms and the handling of the associated data generated from activities on the platforms. This work is a review toward a vision paper that captures some of the inherent security flaws in the Internet of Things.*

Keywords: Internet of Things, Security, Privacy, Cloud Computing

1 Introduction

The Internet of Things (IoT) has been described as 'everything' connecting in real-time, a connectedness among devices anytime, anywhere and in real-time. It has also been touted in [1] as "an augmentation of almost all electronic devices connecting together in an Internet-based architecture where the information derived from this augmentation will improve [their] basic operation, assist in becoming smarter devices, being more efficient, and cost effective".

This new Internet technology is gaining momentum in both design and implementation; it is no longer something to embrace in the future; it is here and now. Gartner, a technology research firm, predicts that the IoT will grow at a faster rate than that of current smartphones and tablets to the tune of 26 billion installed base by 2020 [2].

The current iteration of the Internet has proved valuable for almost all endeavors; it has allowed interaction in a global environment for information exchange and related services accessed through various applications, diverse protocols, and multiple communications tools [3]. According to the views espoused in [4], the IoT is being considered as the next 'wave' in the evolution of the Internet providing extensive interconnectedness of [small] devices that are networked and equipped with sensors and radio-frequency identifiers that are programmable, more intelligent, and enjoying its own Internet presence.

The Internet of Things by design is ubiquitous; it will radically change how information and services are delivered; it is constantly changing the dynamics in varying areas to create a competitive environment and framework that is intended to promote quality, reliability, flexibility, and adaptability [5]. This ubiquity of IoT operations, although competitive and providing plenty of opportunities for business, will present various challenges of which security and privacy are paramount, and could open a haven for nefarious activities by cyber-criminals. Other challenges exist for engineers and administrators who must build, administer, and manage many small yet complex systems while figuring out how to deal with and analyze the deluge of data that the myriad of devices will generate from IoT activities.

Underlying the IoT infrastructure are programmable intelligent smart devices. These devices are being used in multiple activities such as measuring moisture in the atmosphere, checking and monitoring soil moisture in agriculture, and their products; controlling carbon emission in the environment from factories and motor vehicles; alerting drivers and traffic controllers of traffic jams; monitoring patients blood pressure in the health care arena; opening doors; monitoring your home for many different activities including temperature changes; and even unlocking your car doors once the owner's presence is nearby. These activities and more are done without actual human intervention [4].

Some of the security challenges posed by the IoT paradigm comprise unique and susceptible ways of capturing data. The systems are designed with ubiquity and the physical distribution affords attackers great opportunities to gain logical or physical proximity to targets [15]. The ability to collect personal private information is growing and will be one of the single largest components and challenges for user privacy. As indicated in [6], while IoT technologies are currently used for the common good in industries like defense, military, healthcare, air traffic and ground transportation, manufacturing and other areas of national interests, there could be great commercial value and political interests for hackers through widespread cyber criminal activities.

The review for this vision paper is intended to bring awareness to the security ramifications of IoT, and to highlight the way forward or how to address the challenges that can inhibit the promises of IoT services.

2 IoT Related Work

Lu, et al., [6] posit that IoT systems have the ability to collect personal and private information and as such are becoming an important challenge to IoT itself for secured access generally, and specifically for user privacy. Their definition of privacy is based on the psychological consciousness of ‘human shame’ that is manifested when personal and private information is divulged. Individual privacy is subjective; the concept is different for each individual with each having its own boundary, threshold, and share attributes.

Confidentiality is another factor that encircles privacy; this is the degree to which secrecy is embedded in the information that may be hidden or disclosed in the normal course of an encounter.

Lu, et al., studied four primary levels of privacy relating to security goals for IoT systems and these are:

(1) Low security privacy which relates to allowance for unencrypted with authorized access;
(2) Basic security privacy deals with data in storage and transmission;
(3) Medium security privacy addresses validity of access through authorization covering one-time read / write access;
(4) High security privacy is a higher level of encryption encompassing different person / keys.

As these goals are different for each person, it therefore means that individuals must adopt different individual protection strategies and technologies to achieve their corresponding goals.

Roman et al., [7] posit that security is one of the biggest challenges in Internet of Things and this hurdle must be overcome in order for the push in IoT to be fully realized. IoT is “the interacting of billions of devices” with or without human interventions resulting in virtual entities. Protecting the data generated from these interacting devices, and the service provisioning of all the relevant factors must be secured with limiting incidents to affect the entire system. Roman and his colleagues further espoused that the task of protecting the Internet of Things can be very complex and difficult. This is due in part to the number of attack vectors arising from global access, and always-on connectivity that is the bedrock of the Internet; all of which are available to nefarious attackers. Roman et al., [7] (para. *3.3.4.1 pp 10*), proposed an approach based on “privacy-by-design” where all entities would directly manage their own data, including user-centric access policies and mechanisms to control the type and distribution of the data. Since users can theoretically be tracked anywhere and anytime, entities should be available to help users become even more aware of how their surroundings can capture and use their information.

In [19], Hoepman pointed out that the full ramifications of IoT will unfold a confluence of cyberspace and physical space that will present challenges that will require wide research to find IoT-specific solutions for security, privacy, and trust.

Although IoT is new, it is growing at a fast rate. In keeping with the growth many researchers are doing work to analyze the various challenges that are becoming more evident as more ‘things’ are identified, communicating, and interacting with each other [8]. Furthering the dialogue in [8], Miorandi et al., state that “security represents a critical component for enabling widespread adoption of IoT technologies and applications” (*pp 1505*). They suggest that stakeholders might be reluctant to adopt IoT if there are no built-in system-level assurances of confidentiality, authenticity, privacy, and trust. In this vein, the security challenges spread across communications and networking, platform and data management, and applications and services.

2.1 IoT Related Threats

Ubiquitous feature-rich mobile devices have promoted the fast emergence and continuing growth of a ‘people-device-centric’ sensing paradigm that is a central component in the evolution of the Internet of Things as posited by Yao et al., in [9].

Security in the Internet of Things paradigm forces the user to trust what he has no control over and limits what he can do. Once the ‘things’ start communicating and interacting, the data generated would invariably be part of some cloud storage system from where access will be open to data management, analytics, application, and services sub-systems. The device owner or end-user is integral to the data generation process and may not have much control or any control at all over the accessibility of the data.

The security concerns in IoT are taken to mean that devices are protected against unauthorized access, unauthorized modification of any software or firmware and control and exporting of any operational information including user sensory data [18]. These security parameters span communication, networking, data, platform, applications, and services. For the purposes of this review, the concerns are centered around data privacy and trust.

2.1.1 Data Threats and Mitigations

Threats to data represent confidentiality; this is a fundamental issue that must guarantee that only the authorized entities can have access to and modify data [8].
The fact that data storage will primarily be in the cloud and communication and access will be via an Internet connection means that there will always be the probability for data leakages [10], and this therefore means that enhanced methods of authentication and encryption should be designed in order to minimize the prospects of leaked data being accessible to the wrong people.

Threats to the data in the cloud may comprise run-time execution, malicious threats such as denial of service (DoS), natural system threats such as accidents and the adversarial types such as insider attacks. Natural system type threats like accidents can be predicted through estimation and probability theories, and with redundancies put in place for failover, an acceptable threshold can be developed. However, in the case of the adversarial threats and attacks such as DoS, it becomes highly unlikely that a predictable point of acceptance can be determined [11]. Malicious attacks on the data can result in devastating impacts.

Miorandi et al. [8] posit that the everyday approach to ensuring data confidentiality does not apply in a straightforward manner in the IoT. Firstly, this is because of the amount of data that is generated which invariably will impact scalability. Secondly, because of the controls over access in an on-line mode with all the access rights that would be applicable in a dynamic and ever-changing environment where data streams are constantly changing. The sheer dynamics of monitoring the data streams and applying appropriate access controls is a phenomenal task. What therefore are the solutions? These rest to a large degree on trust that the system, the data managers, service providers and other stakeholders will apply appropriate operational measures to assure that vulnerabilities are minimized and, when present, they are easily and quickly contained [12].

Trust models in the Internet of Things arguably have different meanings based on the context; there is no consensus although the importance is widely recognized. This absence of consensus is undeniably an issue that will grow as IoT architecture and cloud integration scalability grows. As device deployment continues to grow, and applications and services, and the interactions of these connected devices that result in a deluge of data continue to grow, trust elements will conceivably follow.

2.1.2 Privacy Threats

Privacy with emphasis on confidentiality is a fundamental requirement that needs to be embedded in IoT; this is so because it must define the rules under which individual's personal data is accessed based on the application domains and the technologies used to do so. In the application domains, in particular healthcare, IoT technologies are being harnessed widely in wireless mobility and ubiquitous data exchanges. This of course poses new issues for privacy, since wireless channels can increase the risks of violation through eavesdropping, spoofing, masking attacks, and launching of denial of service (DoS) attacks.

Miorandi et al., [8] depicted in their graphical model *(Fig. 1)* a definition that covers privacy and suggest that it be promulgated and accepted across the board for IoT. Development of innovative techniques for enforcement, support of the varying heterogeneous devices, and a balance of anonymity must be presented in IoT applications in order to specify, and in all cases, deter the explicit identification of the smart devices connected to the platforms.

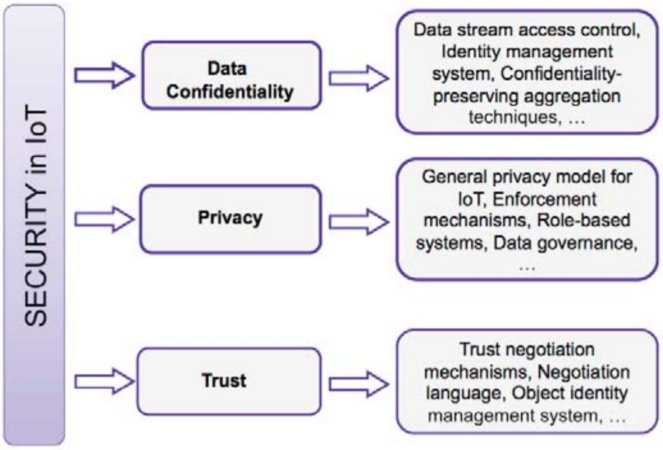

SECURITY in IoT
Data Confidentiality
Data stream access control, Identity management system, Confidentiality-preserving aggregation techniques, ...
Privacy
General privacy model for IoT, Enforcement mechanisms, Role-based systems, Data governance, ...
Trust
Trust negotiation mechanisms, Negotiation language, Object identity management system, ...

Fig. 1. Source: Miorandi, et al., (2012)

Two important questions were raised by Corcoran in [17] that surround connected devices "how will we be able to manage and preserve privacy when every device equipped with wide ranging sensing technologies are connected? And with cameras getting smaller and wearables are being made available for more practical use, how will we know when surreptitious recordings are being done?" The granularity of information from location-aware IoT systems have enormous impact for sensitivity and privacy of users, and for attackers this is an increasing opportunity for a broader spectrum of targets from which to choose [15]. The answers to these questions, and more, are relevant to the overarching question of how privacy will be preserved.

Data collected through IoT-related devices such as smart wearables, smart phones, and smart homes provide contextually rich information that could pose even more significant challenges for privacy and anonymity. Current technology that traces communication through interfaces such as MAC addresses and geo-positioning systems (GPS) and the combination of multiple devices will be fodder that provide unique footprints that tie into data analytics to extract information and knowledge that the user is not aware of or has any control over. Although current technologies do not preserve anonymity, with the full implementation of IoT technologies this will be far more difficult to curtail. In [13], Perera et al. said that the solution must involve the development of elements of anonymity within IoT platforms that cover data collection, storage, transmission and routing, analytics, and aggregation of meta data

3 Way Forward

A clear and holistic approach to privacy as it relates to the evolving IoT platforms must be agreed upon and put into operation. In keeping with the statements in Ziegeldorf, Morchon, and Wehrle [14], this definition should offer elements of a guarantee on the "awareness of the risks associated with smart things and services; individual control of the data and processing of personal information; and control over the subsequent use and dissemination of any other information that may have been collected outside of the personal purview of the user" (*pp 2 para 2.1*).

With this in mind, and with the new ways of interaction that the IoT platforms present, cognizance must be given to privacy threats arising from the ability to identify individuals through the data collected about them. This identification is related to associating an identity to persons, and thus enabling and aggravating other threats such as profiling and surveillance tracking through localization. Localization in this case is not only geared to outdoor activities but through the evolving and significant requirements for the full functioning of IoT platforms, that is, the infinitesimal use of sensors that are designed to 'track' activities and report them for a higher level of analysis.

Significant body of works have explored privacy issues in the ubiquitous computing parameters that is the cornerstone of the IoT, and as alluded to in [15], the establishment of trusted communication channels and the protection of contextual information are very important and integral factors to the protection of user privacy. The ability to control access to device management and information is a vital component for preservation of privacy.

4 Future Work

New approaches to privacy needs to be developed for the uniqueness of IoT; these approaches must prevent abuse and/ or leakage of personal data once collected. Approaches that collect and maintain user profiles on personal devices held by the user are suggested as areas for future development. This means that the security surrounding use and access are held within the purview of the user with the devices that store the profiles acting as a type of 'personal firewall' [19].

As a continuation to this position paper, the author intends to pursue further work in the area of the IoT and to develop a security and privacy model that is applicable to a safe IoT platform. This model would promulgate the principles that underlie a secured infrastructure to maintain personal privacy, trust and governance, while minimizing threats as service attributes in an IoT platform.

5 Conclusions

The Internet of Things is intended to introduce and bring many advances and opportunities to the ways in which devices and people interact. However, with those opportunities come many challenges for information security and privacy.

This paper sought to introduce a discussion on the privacy and threats of IoT platforms and the handling of the associated data generated from activities on the platforms. This review looked at available literature related to user privacy and the data than can be stored in IoT systems that are available and accessible through ubiquitous mobile devices and applications. It is intended to motivate the need for detailed analysis of some of the challenges surrounding privacy threats in the IoT. Since data from IoT activities are invariably stored in a cloud repository, security in the cloud environment is an important topic that creates ongoing dialogue and research to find the best combination for security: confidentiality, availability, access, privacy and trust with proper defense strategies. As cloud computing systems continue to grow perversely, secure preservation and safe access to information through hardware and software solutions such as detection and prevention systems, encryption, and strong authentication mechanisms are needed [16]. Securing the systems and data are ongoing activities that require tools that will be able to perform early detection and mitigation of malicious activities.

In addition, profiling is one of the threats that could be considered most dangerous. This is because the wrong person could be targeted based on meta data and data analytics that could pinpoint persons for surveillance and tracking based on data that are irrelevant to the individual.

Protecting the user data and the privacy of their information is paramount to a successful IoT platform. Consideration must be given to the information and the security and privacy surrounding the collection, storage, sharing and embedded systems that use and control the devices that make up the Internet of Things.

6 References

[1] R. Want, & S. Dustdar, (2015). Activating the internet of things. *Computer* – IEEE Computer Society, Sept. 2015. Available: www.computer.org/computer-sept

[2] J. Rivera, & R. van der Meulen, (2013) Gartner says the Internet of Things installed based will grow to 26 billion units by 2020. Available http://www.gartner.com/newsroom/id/2636073

[3] R. Want, B. Schilit, & S. Jenson, (2015). Enabling the internet of things. *Computer* IEEE Computer Society, Jan. 2015.

[4] K. Pretz, (2013). Exploring the impact of the internet of things. *IEEE - The Institute*, October 2013.

[5] F. Anon, V. Vavarathinarasah, H. Minh, & L. Chung-Horng, (2014). Building a framework for internet of things and cloud computing. In *Proceedings of 2014 IEEE International Conference on Internet of Things (iThings, 2014)*: DOI 10.1109/iThings.2014.28.

[6] X. Lu, Z. Qu, Q. Li, & P. Hui, (2015). Privacy information security classification for internet of things on Internet data. *International Journal of Distributed Sensor Networks,* 2015, Article ID 932941. DOI 10.1155/2015/932941

[7] R. Roman, J. Zhou, & J. Lopez, (2013). On the features and challenges of security and privacy in distributed

internet of things. *Computer Networks*, 57(10):2266-2279, July 2013

[8] D. Miorandi, S. Sicari, F. De Pellegrini, & I. Chlamtac, (2012). Internet of things: Vision, applications, and research challenges. *Ad Hoc Networks* 10(2012) 1497-1516, September 2012.

[9] Y. Yao, L. Yang, & N. Xiong, (2015). Anonymity-based privacy-preserving data reporting for participatory sensing. *IEEE Internet of Things Journal*, 2 (5), October 2015.

[10] K. Lourida, A. Mouhtaropoulos, & A. Vakaloudis, (2013). Assessing database and network threats in traditional and cloud computing. *International Journal of Cyber-Security and Digital Forensics (IJCSDF)* 2 (3), 2013

[11] T. R. Campbell, (2014). Disambiguating cloud security. In *Proceedings of WorldCom'14 International Conference on Internet Computing and Big Data*, Las Vegas, NV July 21 – 24, 2014

[12] M. Burmester, (2013), Trusted clouds. In *Pre-Proceedings of International Workshop on Trustworthiness, Accountability, and Forensics in the Cloud* (TAFC), Malaga, Spain, June 6-7, 2013 ISSN: 2078-2247

[13] C. Perera, R. Ranjan, L. Wang, S. Khan, & A. Zomaya, (2014). Privacy of big data in the internet of things era. *arXiv preprint arXiv:1412.8339*.

[14] J. Ziegeldorf, O. Morchon, & K. Wehrle, (2014). Privacy on the internet of things: Threats and challenges. *Security and Communication Networks* 7.12 (2014): 2728-2742.

[15] M. Covington, & R. Carskadden, (2013). Threat implications of the internet of things. In *Cyber Conflict (CyCon), 2013 5th International Conference on* (pp. 1-12). IEEE

[16] A. Patel, M. Taghavi, Bakhtiyari, & J. Júnior, (2013). An intrusion detection and prevention system in cloud computing: A systematic review. *Journal of Network and Computer Applications 36* (2013) 25-41

[17] P. Corcoran, (2016). The internet of things: Why now, and what's next? IEEE Consumer Electronics Magazine, January 2016, pp 63-69.

[18] V. Cerf, (2015). Prospects for the internet of things. *XRDS*, Winter 2015, 22 (2).

[19] J. H. Hoepman, (2011). *In things we trust? Towards trustability in the internet of things* (pp. 287-295). Springer Berlin Heidelberg.

SESSION

INVITED PAPERS

Chair(s)

TBA

SiTASENSE: Hierarchical 6LoWPAN and Sensing Model for Intelligent Women Safety Services

Dhananjay Singh
Dept. of Electronics Engineering
Hankuk Univ. of Foreign Studies
Yongin-si, South Korea
dan.usn@ieee.org

Gaurav Tripathi
Div. of Central Research Lab.
Bharat Electronics Limited
Ghaziabad, India
gayravtripathy@gmail.com

Hoon-Jae Lee
Div. of Information Networks Engg.
Dongseo University, Busan
South Korea
hjlee@dongseo.ac.kr

Abstract— Internet of Things (IoT) is a new concept of computing technology which is fast emerging as a successful extension to existing Internet in embedded devices. We have visualized interconnections of billions of smart IoT devices to change the way of life. Machine-to-machine communication, machine-to-real world, machine-to-humans are becoming a fast to serve comfortable human life based on 6LoWPAN communication technology. Human security is the one of important aspects of IoT services. The resultant of the IoT objects would be huge amount of data which can be utilized to monitor and control the safety of women and child. In this paper, we have proposed a SiTASENSE model, which is based on hierarchical 6loWPAN networks and intelligent women safety sensing (IWSS) management station. The proposed model concerned about physical interconnections between objects and attributes to access the location. Finally, we have discussed with a use case scenario on the proposed model to deploy in a specific community for the women safety services and sensing of intelligent security management system.

Keywords—Internet of Things, Smart Devices, Women Safety, Intelligent system, Hierarchical 6lowpan

I. Introduction

IoT refers to a loosely coupled, a system of many devices which have the power of sensing, processing, and network capabilities [1]. IoT is gaining fast attention in the academics as-well-as research communities. By enabling the communicating device using Wireless Sensor Networks (WSN) [2] into common objects of day-to-day use and developing newer protocols IoT has started adding a new dimension into human lives. Smart Objects for IPSO [3] has developed a light weight protocol to communicate thousands of devices. The fusion of sensors and 6LoWPAN (IPv6 Low Power Wireless Personal Area Networks) device combination has used to develop a smart objects for Women safety services. However, IoT is an integrated part and an infrastructure of global machines/objects with capabilities of self-configuration based on existing standard and interoperable communication protocols. Where physical devices and virtual "things" have attributes such as identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the current and evolving information network.

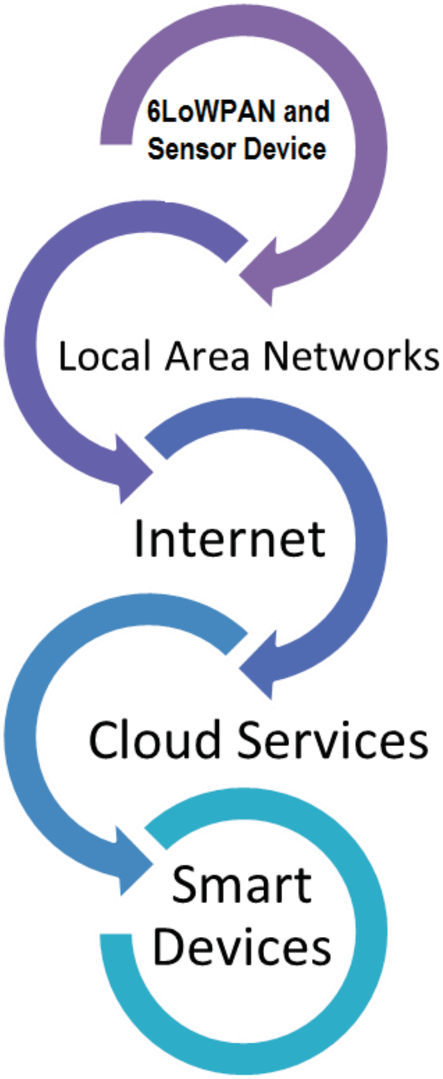


Fig. 1 System overview of IoT and smart objects.

The scope of IoT applications has increased far and wide. It has made impact in the fields ranging from manufacturing, transportation, defence, agriculture, healthcare etc. The world of IoT is getting complex day by day. The sensors here can be called actors which can talk to themselves and can make smart decisions. The major theme of IoT is to encourage information communication which can lead to the data processing and taking up the smart decisions. The gist of the communications can be made stronger and a semantic level can be introduced for IoT to decipher and classify information in semantically rich form [4]. Most of the traditional process of work can now be linked to the prism of IoT and can be transformed for better efficiency of the data. It is the increased usage of the IoT objects that has led to the Ipv6 protocols for linking billions of objects. Homogeneous devices such as smart phones to other smart phones are currently in practice. There are several challenges into the networking of heterogeneous devices such as sensors, and actuators coupled with smart devices. Thus any development framework will need to support the heterogeneous networking. The current technology is evolving to such an extent that IoT has applications of sensors enabled patient identification system and real-time information management of such systems. This change of the technologies and increase in the number of pervasive and ubiquitous sensors has an impact on the current Internet architecture which is slowly and surely migrating towards Future Internet technologies. The future internet will have place for each of the billions of the object that can be connected to the Internet and can be accessed anywhere and anytime. This IoT is making different worlds of humans and machines to integrate. These are commonly known as Machine-to-Machine (M2M) [5], Wireless Sensor Network (WSN) and Web technology and bring them to the same dimensions for enhancing the technological aspects of the IoT domain. The IoT objects designing and integration is all welcome step but we must not forget that it is the security paradigm which needs to develop among the IoT nodes for safe and secure transactions. A standard security platform is missing in the IoT as they are in the development process. We have discussed the issue of cross layer security in our paper [6] where we have specifically mentioned that without cross layer security the data can be corrupted. Ultimately the security paradigm has to converge on robust security mechanisms in coming future.

Fig. 1 shown the example of smart objects internet which is being used for expanding the smart objects by effectively cutting cost and maximizing the output into the networks. The important concern is how do we visualize and actually achieve interoperability between interconnected smart devices, and enhance smartness of these devices by enabling their self-behavior which includes sensing of environment and autonomous behavior, while guaranteeing trust, security, and privacy of the users and their data. Nowadays the current issues of safety of women have been of utmost importance. The rising crimes against women has affected the large sections of the world. The safety of women in any community, area is very important for society[7]. The safety of the women data collected by the government agencies is likely sampled from safe places and then it is forced upon every local population as-well-as village level to society and regional level which is not very useful from the standpoint of trying to understand or change the local dynamics of women safety because it affects all of us. We still lack a viable solution for intelligent women safety sensing (IWSS) system which is efficient in terms of performance, cost and effort needed for maintenance of the system. All machines and equipment's that are part of IWSS known as objects and thus IoT for IWSS system is born. We have developed a framework model to utilize the hierarchical 6LoWPAN and sensing to supports women safety services. The proposed model is capable of solving fundamental issues such as addressing, mobility, routing scalability challenges, security connect and control to support women with internet. We are planning to connect IWSS system with the cloud [8] to support thousands of users simultaneously about the safety conditions for women in any selected area. The hierarchical 6LoWPAN architecture is seen as a set of distributed systems where every information processing is seen as a service. In this paper IWSS system is a SiTASENSE model.

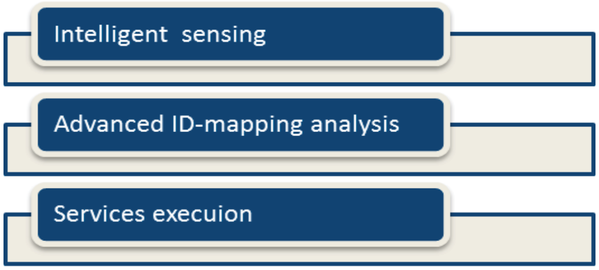


Fig.2. SiTASENSE aspects

IWSS is a part of hierarchical 6LoWPAN model where the application can be generalized in to various spheres of industries. One such disaster management system has been mentioned in [9]. Figure 2 clearly brings out the architecture in the sense that the foremost activity of importance is the sensing of the environment. This sensing becomes intelligent because sensors can periodically forward those data that is of utmost utility. Rest of the data is discarded. Then there is the level of hierarchical mapping which maps the IoT objects in involved in SiTASENSE to have unique identified and follows a hierarchical level of aggregating data.

In this paper we have explore and combined hierarchical 6LoWPAN and IoT techniques to develop IWSS architecture model. The main contribution of this paper is to design and deployment techniques for a real-world IoT based IWSS and which can be integrated in to any new upcoming architecture. Finally we have discussed the challenges and the future plan of our approach.

The remaining of the paper is organized as follows. The section 2 presents a brief introduction of IWSS system and problems in current women security system and their related issues. The section 3 presents hierarchical 6LoWPAN model for women safety in a specific region. Section 4 has discussed women safety related real-time scenario and their solutions. Finally, we have conclude and final declaration of the work shown in the section 5.

II. Intelligent Women Safety Sensing (IWSS)

A. *Problem in IoT Application Space*

The present scenario about the women security in global cities is continuously increasing around the world, especially in large urban areas. The resulting law and order problems have become a major bottleneck to government authorities and decision makers. The existing methods for women safety management, surveillance and control are not meeting the quality standards adequately in terms of cost, maintenance, performance and support. The generation of intelligent women sensing system has yet to evolve in to a proper algorithm. However, SiTASENSE (IWSS) model has to support properly on the IoT platform. The proposed SiTASENSE women sensing system is capable to visualize through a variety of ways specially in real-time sensing data available into the cloud with the concerned Women Safety Cell (WSC) to comprise the actions in case of any emergency or any distress call by the participating women.

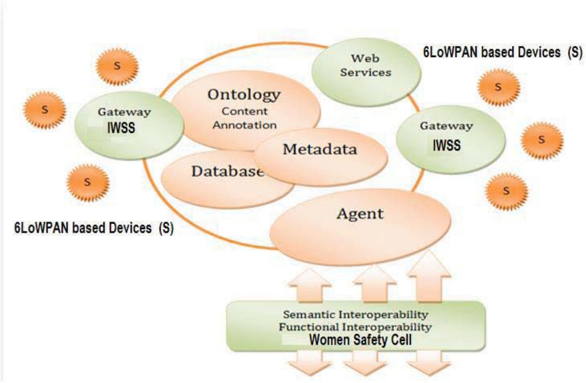


Fig. 3. SiTASENSE processing system (IWSS system)

Fig.3 clearly brings out the idea of the IoT embedded system that we have envisaged for the safety of the women. In the heart of this system there is a microcontroller which can process the information received from the sensors. This information has geo tagged information and power from a battery source. The rules to process information and resultant any threating event can be registered at the Women Safety Cell (WSC) where the alarm can be raised. The movement data is continuously recorded at the cloud level where further data mining can take place.

- *Body sensors:* A small electronic sensing system plug into the body/lockets/dress shall be taking regular safety readings [10]. It has an RF transmitter, which sends the data wirelessly to a local base station.
- *Safe base station*: A base station, which gives this project its name, receives the wirelessly transmitted data from the sensors from the female body. It then relays that data to the Internet via a wired Ethernet connection. The safe base station also acts as a friendly User Interface, which can have manual entry too. These are configurable by applications which will be developed in the future by the SENSE Lab.
- *Data sent to Internet*: The female safety data will be sent in real-time to SiTASENSE, a service, which both stores and provides free access to the data. The service includes embeddable graphs and the ability to generate triggers for tweets and SMS alerts, as well as a robust API which shall have interface with multiple government agencies.

B. *SiTASENSE Model: Women Safety Measures approach*

The SiTASENSE model is a sensor based system designed to allow women safety cell to collect women concentrations outside as well as inside their community area. These places are the most indicative elements related to women safety that can be covered by inexpensive, sensors as-well-as women themselves. The women in the community have to be trained about these self-tagging devices to help in their movement and security patterns.

In SiTASENSE system eeach layer has a huge possibility of data management system and these management systems include virtual sensors known as women tagged device. Human life is precious, it has to be saved. Even while assessing and monitoring crowds, it is the duty of operator at the women safety cell, to monitor around the crowds in and out of the system. More experienced participants will be utilized for areas which shall have more anomalies in the system. The areas where the inferences are less constructive, those areas shall have better participating and trained women for our reports formation.

C. *Women Safety Devices Process Mechanism*

Body sensors can be utilized for self-tagging devices. A small electronic sensing system plug into the body/lockets/dress shall be taking regular safety readings. It has an RF transmitter, which sends the data wirelessly to a local base station. Table 1 shown the different communication protocol to a safe base station, which gives this project its name, receives the wirelessly transmitted data from the sensors from the women body. It then relays that data to the Internet via a wired Ethernet connection. The safe base station also acts as a user friendly interface or manual entry too. These are configurable by applications which are under the development process. In these techniques data sent to the internet in real-time to SiTASENSE, a service, which both stores and provides free access to the data. The service includes embeddable graphs and the ability to generate triggers for tweets and SMS alerts, as well as a robust API which shall have interface with multiple government agencies as well as patrolling agencies. The respective area of interest has to cover by SiTASENSE monitoring cell. SiTASENSE has to involve the process of installing one or more types of sensors for women awareness process. Thus, we are talking about all age groups women in a chosen area of interest for our women surveillance. In this scenario data mining would be a suitable choice to analysis of the huge amount of data in each region at the cloud level.

Table 1. Safety communication devices

Parameter	ZigBee	BT	802.11b	802.11g	UWB
Throughput (Mbps)	0.03	1-3	11	54	200
Max. Range (ft)	75	30	200	200	30
Bandwidth (MHz)	0.6	1	22	20	500

The above table shows the different protocols [1][4][13] of sending information to the respective gateways. Their respective throughput, maximum range of operational efficiency, and bandwidth usage is shown. Traditionally the communication protocol has been derived and discussed in [11].The complexity of the problem can easily understand by the fact that there is a presence of heterogeneous sensors as well as there is a presence of heterogeneous age group of crowds with different upbringings. IoT based system has entertain all the distress calls from the women logged in to the system. We are assuming that the system has seamless at all interfaces, be it smart phones or web based interfaces. In the IWSS, the smartphone request the queries for the safety of the related kith and kin, and both the smartphone and the RFID card reader query the preferences store web service for women sensing preference of the respective user. This type of behavior is often found in pervasive systems where any device can query data from anywhere in the system.

III. Hierarchical 6LoWPAN Model

The goal of the hierarchical 6LoWPAN is to facilitate understanding and communication among acquisition managers, theoreticians, designers, evaluators, and users of data fusion techniques to permit cost-effect system design, development, and operation. Fig. 2 shows the hierarchical 6LoWPAN model where each tier has a specific role for identification of objects to overall process refinement of the system. When we are talking about virtual sensors like all sensors -tagged women, they will actually be the sensor reports to the layer system. The SiTASENSE system has utilized hierarchical model by using sensors as-well-as smart embedded device for sensing the events, person, or thing for a better situational awareness. At the top tier these systems must combine sources data with varying temporal, spatial, spectral and radiometric characteristics.

To design a network topology in a discrete layering forms because each layer allows focusing on specific functions. For example, Top tier (tier 0) can carry traffic across the enterprise (wide area) backbone, medium-tier (tier 1~n-1) can connect buildings at each campus, and End tier (last) can connect user devices and servers within buildings. A typical hierarchical topology is based on following issues.

- A core layer of high-end routers and switches that are optimized for availability and performance.
- A distribution layer of routers and switches that implement policies.
- An access layer that connects users via lower-end switches and wireless access points.

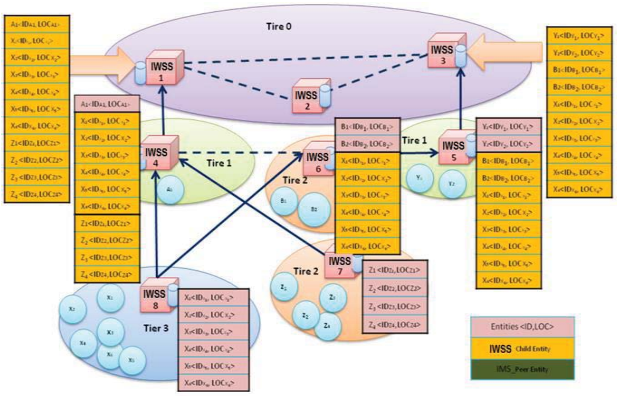


Fig.4. Hierarchical IWSS Layer Structure.

Algorithm -1: //Adding the entities from lower IWSS tier to upper IWSS tier

```
  T[i]<-- $i^{th}$ Tier
  W[i]<-- $i^{th}$ IWSS server
T[i].W[j]<-- $j^{th}$ IWSS server belonging to $i^{th}$ Tier
max_parent <-- maximum no. of parent of any IWSS
max_child <-- maximum no. of child of any IWSS
init(i,j) <-- do the initialization of both i and j
add_child (i,j) <-- make child of i also of j
  while i < max_parent, j < max_child
      do initialization (tier, IWSS)
        input a, b
        if a < b
        add_child (T[b].W[i], T[a].W[j])
       end if
   end while
```

Fig.4 shows the overview of hierarchical layer structure. In this structure each Tier has its set of intelligent women safety sensing (IWSS) management station. There are two kind of relationship between IWSS, Parent-child and Peer relations. Tier0 is the Top-Level-Tier where all IWSS are fully peered and have parent relation.Tier1 and Tier2 are Intermediate

Tiers. They have both parent child relation and peer relation. Tier3 and Tier2 are leaf Tiers and connected to entities. Therefore, leaf tier cannot a parent of another IWSS. Each IWSS has to stored there entities and child node entities ID and location information as-well-as they can share their database with peer's node too. In general , if we denote Tier (IWSS) to be Tier of given IWSS then it should satisfy the relation Tier(IWSS) = MAX (Tier (parents)) + 1 where parents mean parents IWSS of given IWSS. This implies that when new IWSS is added its parent should be in one level upper Tier.

Community IWSS scenario, to detect an isolated lady at a specific location and classifying it as a risk level or safe levels and even identifying it specifically in a specific age bracket is all covered under an object assessment lower tier-3 would be verify/assessment a situation that can use priory information to indicate home of the women in question such as the unique tag number found and location would further indicate its safety parameters levels , if not the exact level , and possibly the women disposition such as movement to contact. The impact assessment modules at the surveillance control modules at the women only police controlled room could use this information then and then indicate that the route and risk levels detected by IoT based sensors and raise an alert or alarm. Thus nearest patrolling unit would be directed to reach the spot immediately.As soon as the panic message is received at the Women Safety Cell (WSC) , the message shall be deciphered, the name of the women , her location as well as the environment factors will be taken under consideration and due protective task force shall be dispatched immediately.

In the context of IWSS, we are layered based architecture to resolve our issue of congestion and preferential path advice to the user. IWSS system controlled by layer based architecture and accessed by the smart phone or RFID smart card[12]. At each cross roads, a motion sensor/counter sensor samples the number of women assembling in their respective area of interest crossing periodically. The sampled readings are aggregated and processed at a women sensing server. The women sensing server processes the aggregate sensor reading and produces an actuation action, which is sent to the alarm/alert/traffic patrolling parties or nearest police stations.

These are actually are actuator of the system. However, there are other feasible solutions to identify in to the system.

- *Data aggregation using hierarchical identification mapping server:* An IoT system carries the unique features having sensing and actuating nodes. The sensors detect the physical/measurable quantity. This is the raw information which is then processed in to meaningful information. The most common approach is by using a hierarchy of data aggregation nodes.

- *Automatic sensing management:* When the number of women enrolled in IWSS system then the IWSS system can query the wellbeing of the respective women in their area of Interest.

- *Safety and Query via Smart Phone:* The family members can query the status of the women of their family from the IWSS system. More of smart phones functions have been discussed in [14].

- *Actuation results by the Processed Sensor Information:* We can visualize that in any IoT based application system we can have one or more actuators which will work as a results of sensor information. Any change in the sensor information can trigger the corresponding changes in the actuator systems. In the IWSS, a change in the density of women in a related area from any of the Sensors can change the aggregate definition of resultant womensecurity density or the inferences dependent upon the women security density. Thus actuator can release any alarms, alerts or traffic condition light management system.

IV. Women Safety Scenario Discussion

We believe that in order to achieve the vision of a future SiTASENSE fully suited to future needs, several aspects needs to be focused. For example the scenario of distress management in which there is a woman which has met with an accident in an area where our sensors are minimal and real time information of the causalities and rescue operations is being hampered by the bad weather. However, in this case the SiTASENSE comes in to play and inspired from [15] decides to send the Distress Report Teams (DRT). The logged in women participants in the SiTASENSE participants shall send us their safe feeling report to the SiTASENSE server at the Women Cell. This whole mechanism of transfer of information has the basis of hierarchical mapping id server which facilitates the information exchange.

This communication medium shall make use of sensor and any other communication mediums. Therefore there is a need to convert the information from data collected over a period of time of women to mobility models. These models shall be known as Women Mobility Models (WMM). As we know WMM are an important tool of profiling human movements and in this case of any issues, they may place a message to the Women Safety Cell (WSC). Several research works has been done in human mobility analyses which have been captured by real life mobility traces by academic experts and Internet communities. This resolves our participating and training issues but the consent of women being traced and important.

SiTASENSE can be a working solution in which the devices can be modeled into the women body and her safety concerns can be addressed specifically. To actually implement these through layers, the following challenges exist.

- *Limits of current Network:* The current Internet architecture has challenges in terms of mobility, scalability which is being challenged in the present IWSS based architecture.
- *Security Issues:* Layer being a scalable architecture, the security needs to be specifically posted for the Identification based keys which must not be manipulated by some external resources or midway in their respective transmission. However proactive identification of the threats needs to be envisaged.

- *SiTASENSE needs to have participating women:* The mobility model data for each women can be unique in terms of the choices of their movements. This must not be sold to marketing companies so as to influence their mindset. There should not be any breach of trust for sensitive and personal; information about the women cannot be compromised.
- *SiTASENSE can be applied to a local area of interest*: Many such security sites of multiple SiTASENSE can be grouped over to form a nodal agency for simultaneous and bigger master SiTASENSE.

Therefore, in the real-time collection of data is vital for preparation of patrolling party and subsequently forwarding to dedicated route or ad-hoc networks has explored and examined in details. In the WMM, we can well understand how we can create genuine women for SiTASENSE. It is imperative that this process when repeated from time to time can help to train the participants in the system.

V. Final Remarks and Conclusions

This paper has discussed a SiTASENSE model for women sensing and safety. The proposed model has capability to response and managing the women security scenario in our desires area of interest especially in the developing countries like India, China, and Russia etc. where we can save lots of resources using local women participants. There is several challenge of using superior WSC system which must provide support for data processing where women beings or their aggregates can be considered as virtually one sensor which has to be saved at any cost.

The problem of false inputs, ambiguous inputs, redundant information and then to take a decision based on the received input is an important task in future. The goal of the system is to enhance the decision support system using various communication technologies like WiFi, Cellular, satellite and GPS etc. for an efficient WSC. In general each one of the women has the potential to understand the evolving situation.

The SiTASENSE model has public private partnership to enhance the safety of women in the concerned area of interest and widen economic opportunities over the networks. Thus, it would be increasing effectiveness of business processes and enabling cross-sector of smart infrastructures in industry. The WMM based schemes are improving human-centered and sensor networks regulatory. The use of sensor oriented hierarchical 6LoWPAN model is very useful for distress monitoring applications. Thus SiTASENSE can be made as the comprehensive women sensing and safety mechanism.

VI. Acknowledgements

This research was supported by Hankuk University of Foreign Studies research fund and the National Research Foundation of Korea NRF2011 project (grant number: NRF-2011-0023076) and NRF2016 project.

VII. References

1. Atzori, L., Lera, A., & Morabito, G. The internet of things: A survey. Computer Networks, 54(15), 2787–2805, 2010.
2. Akyildiz, Ian F., et al. "Wireless sensor networks: a survey." Computer networks 38.4 (2002): 393-422.
3. Vasseur, Jean-Philippe, and Adam Dunkels. Interconnecting smart objects with ip: The next internet. Morgan Kaufmann, 2010.
4. Singh, Dhananjay, Gaurav Tripathi, and Antonio J. Jara. "A survey of internet-of-things: future vision, architecture, challenges and services." Internet of Things (WF-IoT), 2014 IEEE World Forum on. IEEE, 2014.
5. Ali, Anum, Ghalib A. Shah, and JunaidArshad. "Energy efficient techniques for M2M communication: A survey." Journal of Network and Computer Applications 68 (2016): 42-55.
6. Singh, Dhananjay, Gaurav Tripathi, and Antonio Jara. "Secure layers based architecture for Internet of Things." Internet of Things (WF-IoT), IEEE 2nd World Forum on 2015.
7. Symonds, Judith. "Why IT doesn’t appeal to young women." Women, work and computerization. Springer US, 2000, pp. 70-77.
8. J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of Things (IoT): A vision, architectural elements, and future directions,” Future Gener. Comput. Syst., vol. 29, no. 7, pp. 1645–1660, Sep. 2013.
9. Tripathi, G., Singh D. “Heterogenious Crowd-sourcing and Data Fusion Model for Disaster Management Services”, Journal of Theoretical and Applied Information Technology 83.2 (2016).
10. Appelboom, Geoff, et al. "Smart wearable body sensors for patient self-assessment and monitoring." Archives of Public Health 72.1 (2014): 28.
11. Molisch, Andreas F., et al. "IEEE 802.15. 4a channel model-final report." IEEE P802 15.04 (2004): 0662.
12. INFSO D.4 Networked enterprise and RFID INFSO G.2 Micro and Nanosystems. In: Co-operation with the working group RFID of the ETP EPOSS, Internet of Things in 2020, roadmap for the future, Version 1.1, May 27. 2008
13. Hui, J., Culler, D., & Chakrabarti, S., “6LoWPAN: Incorporating IEEE 802.15.4 into IP architecture– internet protocol for smart objects (IPSO) Alliance”, White Paper # 3. http://www.ispo-alliance.org 2009.
14. Jeffrey Nichols and Brad A. Myers, "Controlling home and office appliances with smart phones," Pervasive Computing, IEEE , vol.5, no.3, July 2006, pp.60-67.