EMERGING 2015

The Seventh International Conference on Emerging Networks and Systems Intelligence

ISBN: 978-1-61208-422-0

July 19 - 24, 2015

Nice, France

EMERGING 2015 Editors

Pascal Lorenz, University of Haute Alsace, France

Carla Merkle Westphall, Federal University of Santa Catarina, Brazil

EMERGING 2015

Forward

The Seventh International Conference on Emerging Networks and Systems Intelligence (EMERGING 2015), held between July 19-24, 2015 in Nice, France, continued a series of events focused on presenting and evaluating the advances in emerging solutions for next-generation architectures, devices, and communications protocols. Particular focus was aimed at optimization, quality, discovery, protection, and user profile requirements supported by special approaches such as network coding, configurable protocols, context-aware optimization, ambient systems, anomaly discovery, and adaptive mechanisms.

Next-generation large distributed networks and systems require substantial reconsideration of exiting 'de facto' approaches and mechanisms to sustain an increasing demand on speed, scale, bandwidth, topology and flow changes, user complex behavior, security threats, and service and user ubiquity. As a result, growing research and industrial forces are focusing on new approaches for advanced communications considering new devices and protocols, advanced discovery mechanisms, and programmability techniques to express, measure and control the service quality, security, environmental and user requirements.

The conference had the following tracks:

- Mobility and Ubiquity
- Applications and Services
- Technology and Networking Trends
- Networking and service differentiation
- Security

Similar to previous editions, this event attracted excellent contributions and active participation from all over the world. We were very pleased to receive top quality contributions.

We take here the opportunity to warmly thank all the members of the EMERGING 2015 technical program committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to EMERGING 2015. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the EMERGING 2015 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that EMERGING 2015 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of Emerging Networks and Systems Intelligence. We also hope that Nice, France, provided a

pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

EMERGING 2015 Chairs

EMERGING Advisory Chairs

Tulin Atmaca, IT/Telecom&Management SudParis, France
Carl James Debono, University of Malta, Malta
Robert Bestak, Czech Technical University in Prague, Czech Republic
Zoubir Mammeri, IRIT - Toulouse, France
Constandinos X. Mavromoustakis, University of Nicosia, Cyprus
Raj Jain, Washington University in St. Louis, USA
Phuoc Tran-Gia, University of Wuerzburg, Germany
Norihiko Yoshida, Saitama University, Japan
António Nogueira, DETI-University of Aveiro/Instituto de Telecomunicações, Portugal
Ioannis Moscholios, University of Peloponnese, Greece
Henrik Karstoft, Aarhus University, Denmark
Jean-Michel Dricot, Université Libre de Bruxelles, Belgium
Anne James, Coventry University, UK
Anna Medve, University of Pannonia, Hungary
Nikolaos Tselikas, University of Peloponnese, Greece
Jelena Zdravkovic, Stockholm University, Sweden
Rolf Drechsler, University of Bremen/DFKI, Germany
Christian Blum, IKERBASQUE - Basque Foundation for Science University of the Basque Country, Spain

EMERGING Industry/Research Chairs

Robert Foster, Edgemount Solutions - Plano, USA
David Carrera, Barcelona Supercomputing Center (BSC) / Universitat Politecnica de Catalunya (UPC), Spain
Preetha Thulasiraman, Naval Postgraduate School - Monterey, USA
Anastasiya Yurchyshyna, University of Geneva, Switzerland
Stephan Hengstler, MeshEye Consulting, USA
Haowei Liu, Intel Corp, USA
Jin Guohua, Advanced Micro Devices - Boxborough, USA
Yannick Naudet, Public Research Centre Henri Tudor (CRP Henri Tudor) - Luxembourg-Kirchberg, Luxembourg
Theodor D. Popescu, National Institute for Research & Development in Informatics - Bucharest, Romania
Patrick Senac, ISAE (Institut Supérieur de l'Aéronautique et de l'Espace) - Toulouse, France
Euthimios (Thimios) Panagos, Applied Communication Sciences, USA
Christophe Guéret, Vrije Universiteit Amsterdam, The Netherlands

EMERGING Publicity Chairs

Ines Ben Jemaa, INRIA, France
Ken Katsumoto, Osaka University, Japan
Stefan Frey, Hochschule Furtwangen University, Germany
Zhihui Wang, Dalian University of Technology, China
Eric Veith, Wilhelm Büchner Hochschule, Germany

EMERGING 2015

Committee

EMERGING Advisory Chairs

Tulin Atmaca, IT/Telecom&Management SudParis, France
Carl James Debono, University of Malta, Malta
Robert Bestak, Czech Technical University in Prague, Czech Republic
Zoubir Mammeri, IRIT - Toulouse, France
Constandinos X. Mavromoustakis, University of Nicosia, Cyprus
Raj Jain, Washington University in St. Louis, USA
Phuoc Tran-Gia, University of Wuerzburg, Germany
Norihiko Yoshida, Saitama University, Japan
António Nogueira, DETI-University of Aveiro/Instituto de Telecomunicações, Portugal
Ioannis Moscholios, University of Peloponnese, Greece
Henrik Karstoft, Aarhus University, Denmark
Jean-Michel Dricot, Université Libre de Bruxelles, Belgium
Anne James, Coventry University, UK
Anna Medve, University of Pannonia, Hungary
Nikolaos Tselikas, University of Peloponnese, Greece
Jelena Zdravkovic, Stockholm University, Sweden
Rolf Drechsler, University of Bremen/DFKI, Germany
Christian Blum, IKERBASQUE - Basque Foundation for Science University of the Basque Country, Spain

EMERGING Industry/Research Chairs

Robert Foster, Edgemount Solutions - Plano, USA
David Carrera, Barcelona Supercomputing Center (BSC) / Universitat Politecnica de Catalunya (UPC), Spain
Preetha Thulasiraman, Naval Postgraduate School - Monterey, USA
Anastasiya Yurchyshyna, University of Geneva, Switzerland
Stephan Hengstler, MeshEye Consulting, USA
Haowei Liu, Intel Corp, USA
Jin Guohua, Advanced Micro Devices - Boxborough, USA
Yannick Naudet, Public Research Centre Henri Tudor (CRP Henri Tudor) - Luxembourg-Kirchberg, Luxembourg
Theodor D. Popescu, National Institute for Research & Development in Informatics - Bucharest, Romania
Patrick Senac, ISAE (Institut Supérieur de l'Aéronautique et de l'Espace) - Toulouse, France

Euthimios (Thimios) Panagos, Applied Communication Sciences, USA
Christophe Guéret, Vrije Universiteit Amsterdam, The Netherlands

EMERGING Publicity Chairs

Ines Ben Jemaa, INRIA, France
Ken Katsumoto, Osaka University, Japan
Stefan Frey, Hochschule Furtwangen University, Germany
Zhihui Wang, Dalian University of Technology, China
Eric Veith, Wilhelm Büchner Hochschule, Germany

EMERGING 2015 Technical Program Committee

Mohd Helmy Abd Wahab, Universiti Tun Hussein Onn Malaysia, Malaysia
Mouhamed Abdulla, University of Québec, Canada
Smriti Agrawal, Chaitanya Bharathi Institute of Technology, India
Adel Al-Jumaily, University of Technology, Australia
Eiman Tamah Al-Shammari, Kuwait University, Kuwait
Cristina Alcaraz, University of Malaga, Spain
Firkhan Ali Bin Hamid Ali, Universiti Tun Hussein Onn Malaysia, Malaysia
Mayada Faris Ghanim Alomary, University of Mosul, Iraq
Mercedes Amor-Pinilla, University of Málaga, Spain
Richard Anthony, University of Greenwich, UK
Eleana Asimakopoulou, University of Derby, UK
Tulin Atmaca, IT/Telecom&Management SudParis, France
M. Ali Aydin, Istanbul University, Turkey
Eduard Babulak, Sungkyunkwan University, South Korea
Susmit Bagchi, Gyeongsang National University, South Korea
Zubair Baig, Edith Cowan University, Australia
Valentina E. Balas, Aurel Vlaicu University of Arad, Romania
Kamel Barkaoui, Cedric-Cnam, France
Nik Bessis, University of Derby, UK
Robert Bestak, Czech Technical University in Prague, Czech Republic
Christian Blum, IKERBASQUE - Basque Foundation for Science University of the Basque Country, Spain
Indranil Bose, Indian Institute of Management – Calcutta, India
Kechar Bouabdellah, Oran University, Algeria
Lars Braubach, University *of* Hamburg, Germany
Mieczyslaw Brdys, University of Birmingham, UK
Francesc Burrull i Mestres, Universidad Politecnica de Cartagena (UPCT), Spain
Horia V. Caprita, "Lucian Blaga" University of Sibiu, Romania
Chin-Chen Chang, Feng Chia University - Taichung, Taiwan
Chi-Hua Chen, National Chiao Tung University, Taiwan, R.O.C.
David Chen, University of Bordeaux – Talence, France

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

A New Internet of Things Architecture with Cross-Layer Communication

Alberto Messias da Costa Souza
Cruzeiro do Sul University
São Paulo, Brazil
Email: linuxstring@gmail.com

José Roberto de Almeida Amazonas
Escola Politécnica of the University of São Paulo - USP
São Paulo, Brazil
Email: jra@lcs.poli.usp.br

***Abstract*—This paper describes a new Internet of Things architecture with cross-layer communication. This architecture shows the importance of cross-layer communication between physical, middleware and application layer and explores this functionality to improve the decision make process. Our implementation extends the LinkSmart Middleware, aggregates pattern recognition services to the middleware and introduces a cross-layer communication structure and associate parameters. The cross-layer communication permits to modify the behaviour of the middleware and physical layers by means of control functionalities implemented in the application layer. In this paper, we validate the cross-layer communication in a new IoT architecture with pattern recognition services. We develop an application, that used a real database.**

Keywords-***Internet of Things; Cross-Layer Communication; IoT Middleware; Pattern Recognition.***

I. Introduction

The Internet of Things (IoT) refers to the next generation of the Internet [1], which interconnects trillions of nodes, represented by small ubiquitous devices, equipped with sensors, interconnected Web servers, supercomputers or clusters [2].

This revolution will not affect only the amount of information, but also their quality. Many small processors embedded in objects will be integrated in daily routines.

According to Smith [3], data management in IoT is a crucial aspect. Considering a world of interconnected objects which constantly exchange many kinds of information, the volume of generated data and involved processes makes the data management to become critical. New services to process and analyse the massive data generated by the communication between devices will be needed. These services will need to have open interfaces and will have to be able to provide a simple integration between many applications.

The pattern recognition mechanisms are implemented in the lower layers of the IoT model, namely the physical, middleware and services layers [4]; therefore, and these capabilities need a new architecture in which communication between lower layers and the application layer is enabled. This communication needs a contextualized control implemented in the application layer. In this paper, we introduce this new IoT architecture and its implementation emphasising the communication aspects between layers.

This paper is organised as follows: after this brief Introduction, Section II introduces IoT concepts and IoT middlewares. The proposed architecture and its implementation details are shown in Section III. Conclusions and future works are presented in Section IV.

II. Background

This section introduces the background definitions about IoT and IoT middlewares.

A. Internet of Things

As stated in [5], "the IoT is a global network infrastructure, linking physical and virtual objects through the exploitation of automatic identification, data capture and communication capabilities. This infrastructure includes the existing and evolving Internet and other network developments. It will offer specific object-identification, sensor and connection capability as the basis for the development of independent federated services and applications. These will be characterized by a high degree of autonomous data capture, event transfer, network connectivity and interoperability, actuation and control ".

Figure 1 illustrates the inclusive model as proposed by the CASAGRAS EC funded project [5].

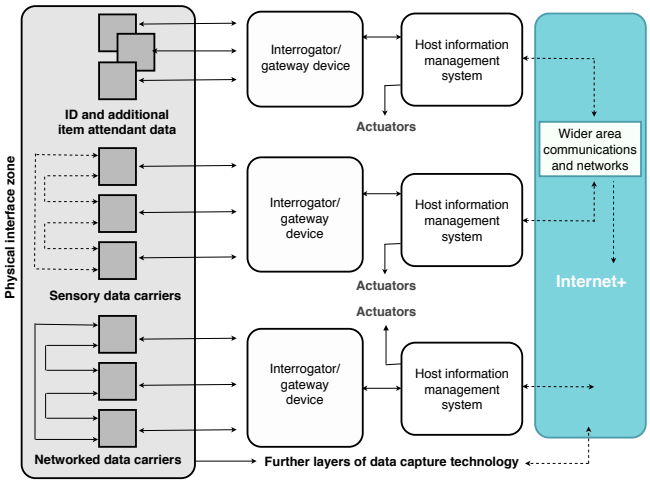


Figure 1. CASAGRAS inclusive model [6].

 ISBN: 978-1-61208-422-0

According to the CASAGRAS inclusive model, a real-world object has its identification ID and associated information stored on some kind of item-attendant data carrier as, for example, on a RFID [7] tag. It is important to realize that the identification technology is not restricted to RFID. Biometry and bar codes are other examples of ID technology that can be employed. The information is retrieved from the object by means of an interrogator that acts as a gateway device and sends it to be stored in a host management system. The Internet is used both to allow access to the retrieved information and to search for further information and associated applications and services. The end result is that an action will take place either displaying new information and/or acting upon the object and/or the environment [6]. The whole process is context-aware and the final action depends on the object itself and its present status in the current environment.

B. Internet of Things middleware

As shown in [8], there are many middlewares, which are defined as software systems that provide an abstraction layer between the operating system and development applications environments in the context of pervasive computing, whose focus is to provide an useful and abstracted suite of procedures that can deal with the heterogeneity of devices and contexts of information.

For this work, it is of relevance the IoT middleware *Network Embedded System for Heterogeneous Physical Devices Middleware in the Distributed Architecture - (HYDRA)* created by the FP6 IST [9], which started in July 2007 and finished in December 2010.

As observed in [10], the first objective of the Hydra Project was the development of a software middleware based on the Service-Oriented Architecture (SOA), in which the communication occurs transparently between the lower layers.

The framework should support centralised and distributed architectures, security and trust, and model driven applications development. One of the framework's development premises was its applicability in current networks and novel network models with interconnected devices that operate with reduced computational power, energy and memory capacity.

The resulting product of this project was called LinkSmart middleware, a name that will be used to refer to the developed middleware from this point onward.

The elements of the LinkSmart middleware are placed between the application and physical layers. The physical layer is related to network communication resources, while the application layer contains modules related to the management of information flow, user interface, application logic and configuration details. Between the two layers is the LinkSmart middleware, consisting of three sub-layers, network, service and semantics, each responsible for specific functions and purposes [9].

The LinkSmart Middleware funcional structure is divided into two parts: (1) Application elements describe components deployed on hardware which is performance-wise capable of running the application that the solution-provider creates. This means these components are meant to be run on powerful machines; and (2) The device elements describe components deployed based on LinkSmart Middleware. These components are running on small devices which have limited resources [9].

III. Cross-layer communication in IoT architecture

This section describes the communication model of the new proposed IoT architecture with cross-layer communication. In our proposal, we insert pattern recognition services in the IoT architecture, specifically in physical and middleware layers. The focus of this paper is the communication model of the new IoT architecture and the cross-layer communication implemented between the physical, middleware and application layers.

To implement this new IoT architecture, we extended the LinkSmart middleware by creating a new pattern recognition services implementation and an abstraction of the following algorithms: outlier detection, values estimation and clustering. This solution is able to apply these algorithms to many kinds of environments and devices. New applications can retrieve contextualized information from the middleware rather than receiving raw data from either the physical or middleware layers.

Figure 2 shows the proposed change in the layer structure of the LinkSmart middleware.

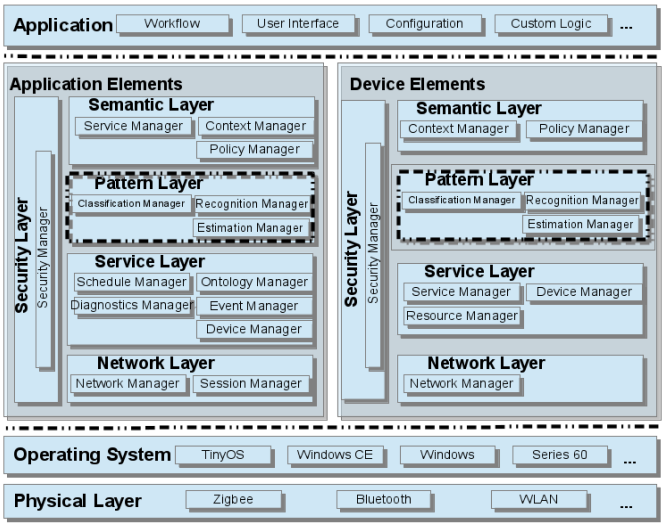


Figure 2. A new layer structure of the LinkSmart middleware, adapted from [9].

In Figure 2, it is shown a new box called *Pattern Layer*, highlighted by a dashed rectangle. This new layer has three managers: classification, recognition and estimation, which implement the pattern recognition functionalities.

The value estimation, behaviors recognition, classification and outlier detection algorithms [11] [12] contribute to network traffic minimisation in the IoT context, as the upper application layer will not receive raw data but pre-processed information by the LinkSmart middleware pattern services.

 ISBN: 978-1-61208-422-0

These algorithms have been implemented with a distributed processing data architecture using the Big Data technology [13] [14].

In our implementation we used the following techniques: linear regression for values estimation [11], k-means algorithm for clustering [11] and contextualize for retrieved values from sensors and others devices. To detect outliers we used clustering distance [15] [16] [17].

The values estimation and outlier detection algorithms have also been implemented in the physical layer.

Figure 3 shows the implemented architecture.

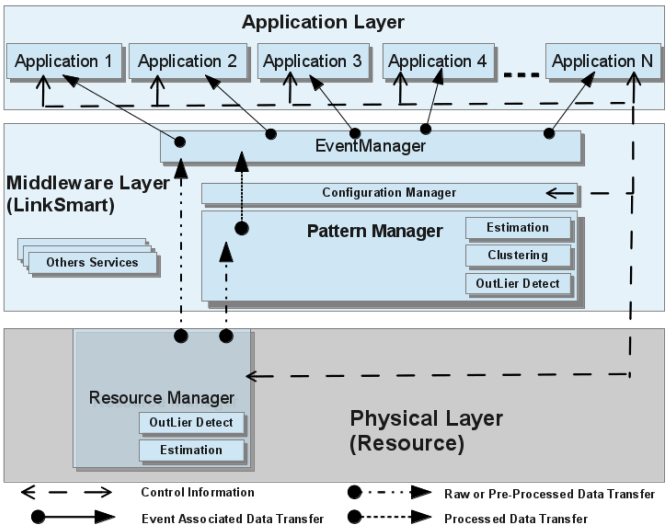


Figure 3. Implemented architecture

The most important aspect of this implementation represented by the Figure 3 is a new pattern recognition module inserted in the LinkSmart middleware. This implementation follows the IoT-A reference model [18], and has the following layers:

- Physical layer: hosts the resource layer represented by sensors and smart objects. The resource manager box represents the driver or software responsible to connect with the LinkSmart and to send raw data or pre-processed data by the value estimation and/or outlier detection algorithms. If data generated by a physical resource is processed in this layer, the resource manager informs the configuration manager in the LinkSmart middleware or directly the application. Both the configuration manager and the application can change parameters in the resource manager to disable the pre-processor so the resource manager proceeds to forward the raw data. Note that from this layer data can flow to the middleware or to the application layer.
- Middleware Layer: it is represented by the LinkSmart Middleware, modified in this implementation by the inclusion of the pattern recognition and configuration managers, the event manager which is quite important in this architecture, and others services that are less relevant in the present case. The pattern recognition manager implements three services: value estimation, clustering and outlier detection. The configuration manager enables the applications or the resource manager to configure parameters in the pattern recognition manager, defining when the algorithms are run or how much data must be stored, enabling or disabling pattern recognition services, cleaning the stored data. After processing data the pattern recognition services can send the contextualized information to the event manager which by its turn will detect and send the new events to its clients. The event manager tackles the scalability issue as it can be in charge of one application client or millions of them. The event manager can be directly accessed by the resource manager which in this case does not activate the pattern recognition services. This feature can be chosen by the application designer and later changed in the configuration manager.
  This layer implements the cross-layer communication as it can receive or send configuration parameters both from the resource manager and applications. However, the most relevant aspect of cross-layer communication is the fact that raw data can be open, processed and interpreted in this layer which otherwise should be a specific function of the application layer according to the ISO/OSI layer model [19].
- Application Layer: the application layer is represented by client applications and configuration applications. The client applications receive the events from the event manager, either raw data or processed information. The configuration application is responsible to configure parameters in the configuration manager or send parameters to the resource manager to control its behaviour or the pattern recognition services activation/de-activation. This bidirectional communication along the associated configuration feature represents the cross-layer communication.

The cross-layer communication is a requirement of this architecture and is implemented as previously described. The proposed architecture enables a fully distributed implementation as the information can be processed at each physical node, or at network access points, or at middleware service delivery nodes, or at application nodes. The cross-layer communication of our proposed architecture enables this distinctive IoT feature.

A. Implementation aspects

We created an object-oriented structured programming middleware implementation. Figure 4 illustrates the specific cross-layer communication structure.

Figure 4 shows the class diagram that represents the developed object-oriented programming structure. There are two classes that represent the cross-layer parameters: (1) the interface *CrossLayerParameter* that creates a main structure of the cross-layer parameters. It can be seen the contract of abstract methods *setLayer* that is responsible to define

 ISBN: 978-1-61208-422-0

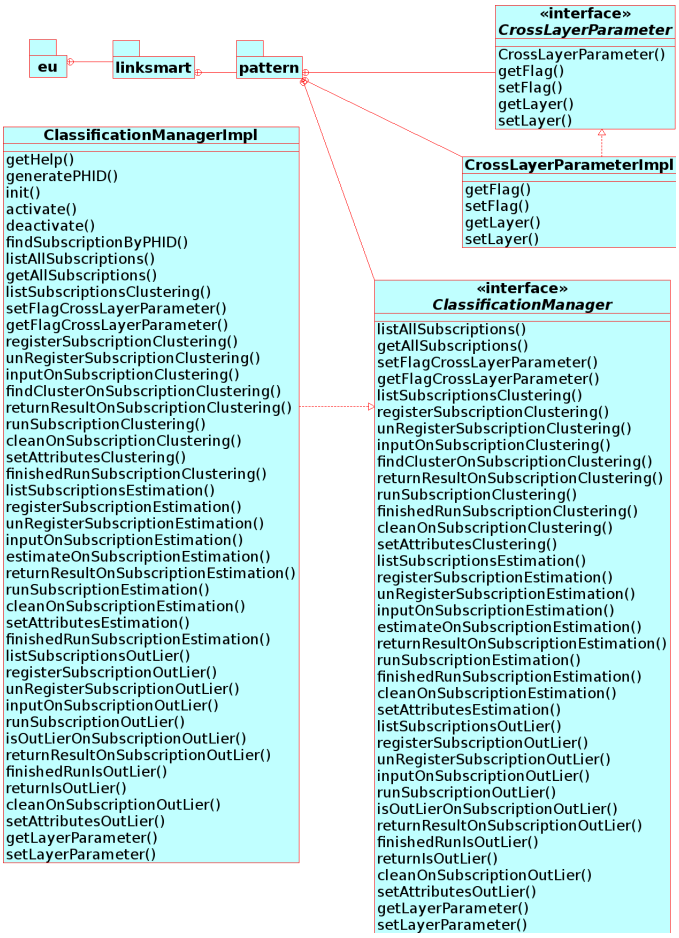


Figure 4. Class diagram of the cross-layer parameters and service implementation in the pattern recognition module.

a layer of either processed or not processed data, *getLayer* that is responsible to return the defined layer, *setFlag* that is responsible to define a boolean flag to inform if the pattern service is active or not in this layer, and *getFlag* that is responsible to return the state of the flag. The class *CrossLayerParameterImpl* is the actual implementation of the methods defined in the interface *CrossLayerParameter*.

The classes *ClassificationManager* and *ClassificationManagerImpl* define the structure and implementation of the pattern recognition services inserted in the LinkSmart Middleware, but they are not the focus of this paper.

New applications to use this structure and services implemented in the LinkSmart Middleware need to define parameters of the middleware or physical layers. Anytime, during the execution of the application, the parameters and behaviors of the physical or middleware layers can be changed. The driver program responsible to connect devices and the LinkSmart middleware must use and interpret those parameters and change the devices' behaviour.

B. A testbed implementation

This section describes the developed testbed implementation, with a resource manager and a test client application to demonstrate the functionality of the proposed cross-layer communication.

Figure 5 shows the web page with the LinkSmart status. It can be seen that the service, highlighted by a red line, called by *ClassificationManagerImpl* has started and was registered in the middleware with Hardware Identification (HID) *0.0.0.8650460202121146535*.

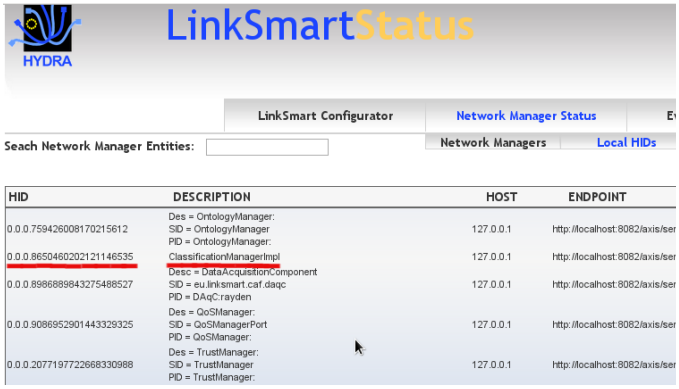


Figure 5. The LinkSmart middleware webpage with ClassificationManager services started.

The raw data used by the resource manager are from the Guildford's facility that is a member of the European Commission funded Smart Santander Project [20].

The retrieved data were inserted in the Mysql [21] database and a class to simulate the resource manager was created. The resource manager provides temperature and light intensity values from a single sensor node, designated as *node25*.

To illustrate the proposed cross-layer communication, two applications have been created: (1) ResourceManager representing the physical layer and that is responsible to retrieve the data from the Mysql database and to send them to the Pattern Recognition Manager in the LinkSmart middleware; and (2) the client application that is responsible to control the Pattern Recognition Manager or the Middleware and Physical layers behaviors.

Figure 6 illustrates the execution of the developed resource manager application to represent the physical layer. This application has two services implemented: the value estimation algorithm to estimate a value if the real sample is missing and this value is relevant to the application; the outlier detection algorithm if an erroneous data represents a problem to the application as proposed in [4].

Other relevant aspect of this application is its capability to send and receive parameters both from the Middleware and Application layers. Note that the application starts the connection with the Middleware layer, informs the status of the services and begins to send instances with temperature and light intensity values of the environment. In a certain instant of time, the Physical layer receives a message to change the

 ISBN: 978-1-61208-422-0

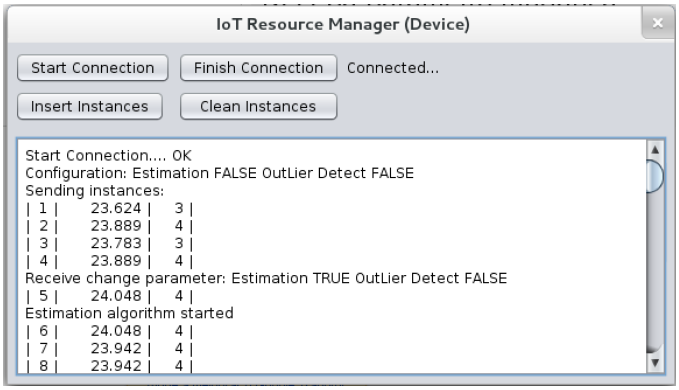


Figure 6. Resource Manager application representing the physical layer.

estimation parameter and to start this service. This execution example illustrates how the cross-layer communication enables the physical layer's behaviour modification in execution time.

Figure 7 illustrates the execution of the client control application developed for the testbed.

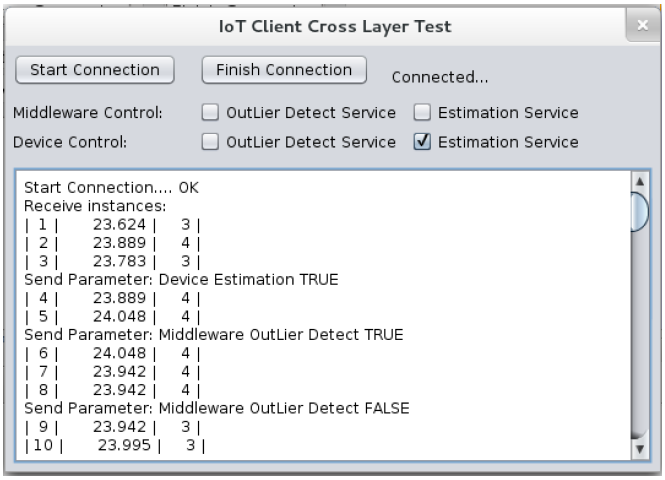


Figure 7. The client control application developed for the testbed.

This application has two functions: (1) it illustrates the client receiving data from the Middleware layer; (2) it implements the control function of the behaviors of the Physical and Middleware layers. In the application's graphical interface, it is shown the control of the Physical and Middleware layers: the application sends a message to start or stop the values estimation or outlier detection services in any of the layers using the developed cross-layer communication structure. The user can enable or disable the checkboxes in the interface.

In the illustrated example of the application execution, one can see that the program established the connection with the LinkSmart and started to receive instances with sample values. Next the user selected the checkbox to start the values estimation service in the physical layer: the application sent this parameter to the Middleware layer which forwarded the new value to the Physical layer. Then, the user marked the checkbox to start the outlier detection service in the Middleware Layer stopping it after a while, changing behaviors in both layers.

Figure 8 illustrates the LOG file generated with the execution of the ClassificationManager services implemented in the LinkSmart.

```
2015-01-09 21:35:50,378 INFO  - ClassificationManager Services Started
2015-01-09 21:35:51,129 INFO  - ClassificationManager Service OutLierDetection Started
2015-01-09 21:35:51,835 INFO  - ClassificationManager Service Estimation Started
2015-01-09 21:35:52,411 INFO  - ClassificationManager Service Clustering  Started
2015-01-09 21:35:56,984 INFO  - ResourceManager (IoTResourceManager.class)
                                 PHID 9835478513654
2015-01-09 21:35:59,129 INFO  - Client (ClientIoTCrossLayerTest.class) PHID 6584236574189
                                 Subscriber PHID 9835478513654
2015-01-09 21:36:02,455 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:02,991 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:03,393 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:04,856 INFO  - Client (ClientIoTCrossLayerTest.class) Receive
                                 ChangeParameter: Device Estimation TRUE
2015-01-09 21:36:05,384 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:07,032 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:08,775 INFO  - Client (ClientIoTCrossLayerTest.class) Receive
                                 ChangeParameter: Middleware OutLier Detect TRUE
2015-01-09 21:36:09,562 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:10,135 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:11,221 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:13,341 INFO  - Client (ClientIoTCrossLayerTest.class) Receive
                                 ChangeParameter: Middleware OutLier Detect FALSE
2015-01-09 21:36:14,378 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:15,824 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:16,258 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:17,874 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:18,561 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:19,992 INFO  - PHID 9835478513654 Receive instance...send to subscribers
2015-01-09 21:36:20,324 INFO  - PHID 9835478513654 Receive instance...send to subscribers
```

Figure 8. ClassificationManager LOG.

In Figure 8, one can see the part of the LOG file that refers to the execution of the client control application of the testbed.

This log file shows some lines referring to the initialisation of the *ClassificationManager* and its pattern recognition services, followed by the IoTResourceManager (Physical layer) and ClientIoTCrossLayerTest (Application layer) actions. The ResourceManager sends instances to the LinkSmart and the client application sends messages to the Physical layer with the estimation start parameter. Then it can be seen the messages to change services in the Middleware layer, i.e., to start and after a while to stop the outlier detection service.

The behaviors change of the Physical and Middleware layers, required by the Application layer, are also shown.

IV. Conclusions and Future Work

In this work, we have proposed, implemented and tested a cross-layer communication structure for an IoT architecture. The structure runs integrated with the novel IoT architecture implemented by extending the LinkSmart middleware with the following pattern recognition services: outlier detection, values estimation and clustering in the Middleware layer, and implementing the outlier detection and values estimation services in the Physical layer as well.

The proposed structure and its corresponding implementation allows the Application layer to change the behaviors of the lower layers in the IoT model, specifically the Physical and Middleware layers. The Application layer can enable or disable these services.

The developed object-oriented programming structure introduces scalability in the parameters handling and pattern recognition services aggregated in the LinkSmart middleware.

 ISBN: 978-1-61208-422-0

This framework addresses scalability, contextualisation and flexibility enabling a huge number of different kinds of devices to acquire environment context awareness. The information provided by a single light sensor, for example, can be read by various applications without any interference on each other. The raw data is processed only once in the Physical or Middleware layers, so different applications may be simpler and receive the filtered information, without the need to process the original raw data. This approach reduces the network traffic and the overall energy consumption.

The testbed implementation validated the proposed cross-layer communication integrated with the IoT architecture using real data from the Smart Santander Project. The execution shows that the IoT architecture implementation, the LinkSmart middleware extended with the pattern recognition services and the new communication model work with real data.

As future work, we will validate the novel IoT architecture with pattern recognition services, cross-layer communication and novel tariff systems in a real telecommunication network with real users and scalable applications.

ACKNOWLEDGEMENTS

We acknowledge the ICT- 2009-257992 (SmartSantander) and the REDUCE project grant EP/I000232/1 under the Digital Economy Programme run by Research Councils UK that supported the development and deployment of the SmartCampus testbed.

REFERENCES

[1] C. Pfister, *Getting Started with the Internet of Things*, 1st ed. O'Reilly Media, Inc., 2011.

[2] A. Jammes, J. Cooper, K. Jeffery, and G. Saake, "Research directions in database architectures for internet of things: A communication of the first international workshop on database architectures for the internet of things (dait 2009)," 2009, pp. 225–233.

[3] I. Smith, *The Internet of Things 2012: New Horizons*. CASAGRAS2, 2012, retrieved: July, 2014. [Online]. Available: http://www.internet-of-things-research.eu/pdf/IERC_Cluster_Book_2012_WEB.pdf

[4] A. M. Souza and J. R. Amazonas, "A novel smart home application using an internet of things middleware," in *Smart Objects, Systems and Technologies (SmartSysTech), Proceedings of 2013 European Conference on*, 2013, pp. 1–7.

[5] E. F. P. CASAGRAS, "Casagras final report: Rfid and the inclusive model for the internet of things," 2009.

[6] J. R. d. A. Amazonas, "Network virtualization and cloud computing: Iot enabling thecnologies," *Casagras2 Academic Seminar*, September 2011, retrieved: November, 2013. [Online]. Available: http://www.casagras2.com.br/downloads/day2/2-Jose_Roberto_de_Almeida_Amazonas-Network'_Virtualization_andar_Cloud_Computing_IoT_enabling_echnologies.pdf

[7] B. C. Hardgrave, J. Aloysius, and S. Goyal, "Does rfid improve inventory accuracy? a preliminary analysis," *International Journal of RF Technologies: Research and Applications*, vol. 1, no. 1, pp. 44–56, 2009. [Online]. Available: http://dx.doi.org/10.1080/17545730802338333

[8] K. E. Kjaer, "A survey of context-aware middleware," 2005, retrieved: January, 2015. [Online]. Available: http://hydramiddleware.eu/hydra_papers/A_Survey_of_Context-aware_Middleware.pdf

[9] M. Sarnovsky, P. Kostelink, P. Butka, J. Hreno, and D. Lackova, "First demonstrator of hydra middleware architecture for building automation," June 2005, retrieved: January, 2015. [Online]. Available: http://www.hydramiddleware.eu/downloads.php?cat_id=2&download_id=29

[10] H. Project, "Hydra project overview," June 2007, retrieved: January, 2015. [Online]. Available: http://www.hydramiddleware.eu/articles.php?article_id=68

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*, 2nd ed. Wiley-Interscience, November 2000.

[12] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, 4th ed. Academic Press, 2008.

[13] H. Sun and P. Heller, "Oracle information architecture: An architect s guide to big data," in *An Oracle White Paper in Enterprise Architecture*, 2012.

[14] D. Tracey and C. Sreenan, "A holistic architecture for the internet of things, sensing services and big data," in *Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on*, 2013, pp. 546–553.

[15] R. Pamula, J. Deka, and S. Nandi, "An outlier detection method based on clustering," in *Emerging Applications of Information Technology (EAIT), 2011 Second International Conference on*, 2011, pp. 253–256.

[16] D. Lei, Q. Zhu, J. Chen, H. Lin, and P. Yang, "Automatic k-means clustering algorithm for outlier detection," in *Information Engineering and Applications*, ser. Lecture Notes in Electrical Engineering, R. Zhu and Y. Ma, Eds. Springer London, 2012, vol. 154, pp. 363–372.

[17] A. M. Souza and J. R. Amazonas, "An outlier detect algorithm using big data processing and internet of things architecture," *Procedia Computer Science*, vol. 52, no. 0, pp. 1010 – 1015, 201, the 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050915008959

[18] W. Joachim and S. Walewski, "Internet of things architecture iot-a," *Deliverable D1.4 - Converged architectural reference model for the IoT v2.0*, 2012.

[19] A. Tanenbaum, *Computer Networks*, 4th ed. Upper Saddle River; NJ: Prentice Hall, 2003.

[20] M. Nati, A. Gluhak, H. Abangar, and W. Headley, "Smartcampus: A user-centric testbed for internet of things experimentation," in *Wireless Personal Multimedia Communications (WPMC), 2013 16th International Symposium on*, 2013, pp. 1–6.

[21] S. Tahaghoghi and H. Williams, *Learning MySQL*. O'Reilly Media, 2006.

 ISBN: 978-1-61208-422-0

DIYbetes: a Mobile Platform for Empowering Type 2 Diabetes Patients

Mário Zenha-Rela
RedLight Software Inc.
IPN Business Center
3030 Coimbra, Portugal
mzrela@redlightsoft.com

Ana Paula Matos
School of Psychology
University of Coimbra
3001 Coimbra, Portugal
apmatos@fpce.uc.pt

Lèlita Santos
School of Medicine
University of Coimbra
3004 Coimbra, Portugal
lelitasantos@fm.uc.pt

Miguel Antunes
RedLight Software Inc.
IPN Business Center
3030 Coimbra, Portugal
mantunes@redlightsoft.com

João Diogo Costa
RedLight Software Inc.
IPN Business Center
3030 Coimbra, Portugal
jdcosta@redlightsoft.com

Marta Oliveira
RedLight Software Inc.
IPN Business Center
3030 Coimbra, Portugal
moliveira@redlightsoft.com

***Abstract*—Type 2 Diabetes is a growing pandemic that already affects 347 million people in the world, which means that one in every 20 people is affected by the disease. Since Diabetes is a disease with significant behavioral roots, managing it involves a shift in behavior. A multidisciplinary team of researchers from Medicine, Psychology and Information&Communication Technologies have gathered to build a mobile platform that addresses behavior change of patients by helping them avoid bad decisions and pushing them to have a healthier lifestyle. In this paper, we describe the technology behind the DIYbetes platform and how it is a paradigm of a networked solution and system intelligence that supports a demanding set of non-functional (quality) requirements such as availability, latency, throughput, modifiability, maintainability, testability and security.**

***Keywords*–mobile systems, cloud, dependability, diabetes, healthcare.**

I. INTRODUCTION

Diabetes Mellitus (DM) is a group of metabolic diseases in which a person has high blood sugar levels. Untreated, diabetes can cause many complications, namely cardiovascular disease, chronic renal failure, and retinal damage. Adequate treatment of diabetes is thus important, as well as blood pressure control and lifestyle factors such as stopping smoking and maintaining a healthy body weight. There are three main types of diabetes mellitus: Type 1 DM (T1DM) results from the body's failure to produce insulin, and currently requires the person to inject insulin or wear an insulin pump. Type 2 DM (T2DM) results from insulin resistance, a condition in which cells fail to use insulin properly, sometimes leading to absolute insulin deficiency as in T1DM. The third main form, gestational diabetes, occurs when pregnant women without a previous diagnosis of diabetes develop a high blood glucose level. It may precede the development of T2DM.

All forms of diabetes have been treatable since insulin became available in 1921. T2DM, especially in early stages, may be controlled with a lifestyle change as a number of lifestyle factors are known to be important to the development of T2DM, including obesity, lack of physical activity, poor diet, stress, and urbanization. The World Health Organization estimates that in 2012 about 347 million people had diabetes, with T2DM making up about 90% of the cases [1]. The same report estimates that in 2004 only, about 3.4 million people died as a direct consequence of DM. This disease incidence keeps increasing rapidly, with the direct number of deaths attributable to DM estimated to increase 66% between 2008 and 2030. In terms of costs, the International Diabetes Federation estimates that 418 billion dollars have been used for the management and treatment of this disease in 2010, with this figure expected to grow to about 561 billion dollars in 2030. DM occurs throughout the world, but is more common in the more developed countries. However, the greatest increase in prevalence is expected to occur in Asia and Africa, where most patients will probably be found by 2030. The increase in incidence in developing countries follows the trend of urbanization and lifestyle changes, namely into a 'westernized' diet. In fact, the World Health Organization (WHO) characterizes T2DM, which represents 90% of DM cases, as a disease almost exclusively derived from unhealthy lifestyle practices [2], [3].

Having described the problem and justified its relevance, in the next sections we describe an effort to address it using technology in a sensible way. DIYbetes (http://DIYbetes.org) is a government-funded initiative to promote Type 2 Diabetes patients self-help using mobile platforms (smartphones and tablets). It was designed with inputs from diabetes medical experts and psychologists that have been dealing with the disease for more than three decades, and a team of Information&Communication professionals. In Section II the objectives of the project are highlighted stressing how important is to build a solution that empowers patients, rather than technical wizzardry. Then, in Section III the core non-functional (quality) attributes of the platform are presented, as an effective solution needs to deal with much more than functionality (e.g., privacy, usability). Armed with this background we arrive at the technical core of the paper (Section IV) where the architecture of the solution is presented, and some trade-offs are discussed. Section V closes the paper with the project current status and further evolutions underway.

II. PLATFORM OBJECTIVES

The DIYbetes project main objective is to create a user-centered technical solution that will allow T2DM patients to gain better awareness of their daily habits and track their disease's evolution. This is achieved by promoting self-awareness and empowerment, while reducing the need for unsustainable in-person appointments due to distance, increase of prevalence and unsustainable increase in costs by healthcare organizations and authorities. DIYbetes is an application running mainly on smartphones and tablets (Figure 1) and also accessible via a web portal. It is designed to be integrated into the service-chain of healthcare professionals to allow recording of glucose

 ISBN: 978-1-61208-422-0

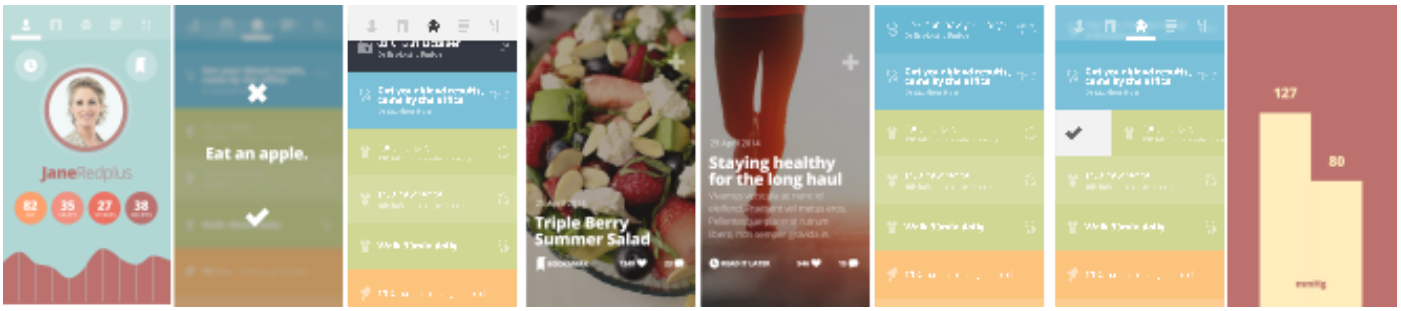


Figure 1. Some screens extracted from the DIYbetes mobile app.

levels and other disease related information, as well as support direct communication with the healthcare professionals [4]. The interactions between patients and their doctors will allow the former to feel more supported. A distinctive feature of this application is the perceived increase in quality of service by the patients due to the follow-up, without costly travel and appointment costs that would be required for such a personalized interaction. To summarize, the DIYbetes application is intended as a bridge between patients and their accompanying health professionals, as it has been designed not only as a logger, but as an information vehicle where diabetics can monitor their information and physicians can have a closer, and continuously up-to-date, look of their patient's records [5], and communicate directly, particularly when distances would prevent this kind of monitoring from happening without such solution (e.g., large area countries without an extensive transportation infrastructure). In synthesis, the technical solution:

- Improves diabetes control by the patient.
- Provides a communication channel between patients and health professionals.
- Aims to reduce the diabetes induced complications and mortality.
- Delivers personalized disease related information.

III. Approach

We shall now present the state-of-art for related approaches, followed by the core requirements that serve as architectural drivers for the solution built.

A. Current approaches

There are several thousand mobile applications to manage DM, both at GooglePlay! and the AppStore [6]-[9]. Most are one person initiatives by technically-savvy patients, others are solid commercial offerings, and a large number is delivered by healthcare institutions to help their patients deal with the disease. However, most of these applications target T1DM patients, due to the criticalness of the condition. This is unfortunate because T2DM represents about 90% of DM patients worldwide. Another reason for the lack of T2DM solutions derives from the fact that this condition is prevalent with age, normally above 45 years, while T1DM occurs at an early age, starting at 9 years or earlier. The lack of technological literacy of elder patients has been preventing the widespread deployment of such solutions. However, several technology observatories have reported a significant growth of tablets among the older population [10], which might be due to the larger screen size and easiness of reading for sight constrained individuals. This growth, while more evident in the more developed coutries (those where the prevalence of T2DM is higher) opens the opportunity for a multiplatform solution as DIYbetes. Currently, most DM solutions focus on logging glucose values or diabetes safe recipes. A remarkable outlier is the Glooko mobile app [11] whose focus is to support people with diabetes interacting with their blood glucose readings and add lifestyle context including carbs, insulin & medication, and activity data. While its main target are T1DM patients, it has been successfully used by T2DM patients too. Despite its very interesting motivational approaches, the fact is that T2DM patients are more receptive to motivation tips targeting their age bracket. That is a major focus of this project and the reason why we have brought into the project psychologists with more than three decades experience with T2DM patients.

As referred above, addressing T2DM requires a behavior change, namely a healthier diet and regular exercise. This requires a service managed by physicians that keep patients motivated. It is manifest that only technology can address the ever growning number of patients [12] by streamlining the treatments, and motivating patients to follow healthier behavior/habits. Hence, more than a logger of glycose readings, this project aims to be a personal lifestyle management tool intended to coach people with diabetes.

B. Requirements as architectural drivers

There are a set of core requirements that have a significant impact in the design decisions associated with the DIYbetes platform. These requirements are the architectural drivers organized as 1) business constrains, 2) technological constrains, 3) functional requirements, and 4) quality attributes.

The *business and technical constrains* are the business and technical decisions made upfront the design. The *functional requirements* specify what the system must execute, but without the details of what does not have an architecture impact (for example colors of the displays and kinds of widgets). The *quality attributes* describe the system's behavior i.e., the properties that will describe how a system achieves a certain functionality. The project's functional requirements include items such as 'support for multiple devices and interfaces' so that users can record and monitor their activities anywhere. This feature allows physicians to follow more closely their patients activities, specify objectives, and act upfront when

 ISBN: 978-1-61208-422-0

the results are not the ones expected. Their role is to specify achievements and objectives in order to keep users committed, engaged, and help them to cure their disease.

The system also works as a coach for the physician suggesting challenges based on the users' objectives and health values. Logging meals, exercise, and health values it is a dull activity, so the system identifies the patterns in users' behaviors in order to streamline those tasks. Finally, the platform tries to influence and motivate patients to feel proud of their achievements.

As highlighted by functional requirements, the platform handles sensitive health and personal data. Due to this fact, security is a key quality attribute and an important technological constrain in the system. Both system's information and infrastructures must respect the legal personal data protection regulations, namely the european General Data Protection Regulation (GDPR) [13].

Additionally to the GDPR compliance, this project aims to gather new sound scientific evidence with regard to the patient's compliance. That is why usability is a top concern, since the daily logging needs to be simple and consistent in the different interfaces so that users keep logging their data.

As the business constraints also highlight, the flexibility to adapt and integrate different business models (commercialize new services, integrate new treatment plans, and integrate multiple devices) allows the platform to adapt to market trends, and explore new markets while keeping support costs low. In fact, the project is aligned with the most recent US healthcare legislation (March 2015), namely the Centers for Medicare and Medicaid Services (CMS) Stage 3 proposed rule [14] that requires that more than 35 percent of all patients seen by the provider or discharged from an hospital should received a secure message using the EHR's electronic messaging function or in response to a secure message sent by the patient. The proposal also calls for more that 15 percent of patients to contribute patient-generated health data or data from a non-clinical setting.

As can be seen from this overview, the complexity of this platform is very significant. In the next Section, we present and describe the architectural solution that supports such demanding requirements.

IV. Technical Solution

In this Section, we present the technical solution that supports the project's goals and constraints. We shall also dicuss some of the design decisions that were taken, as well as some trade-offs that an actual system always imposes.

A. Overview

In Figure 2, we present the context view of the platform. It describes, from a dynamic perspective, the system boundaries and how external actors (people or systems) interact with the system.

Patient to DIYbetes System

This data flow represents log info (health values, meal logs, workout sessions, and drugs intakes), suggestions (examples of healthy restaurants/menus in a given location, workout places, and recipes), and social info (events, new friends, and

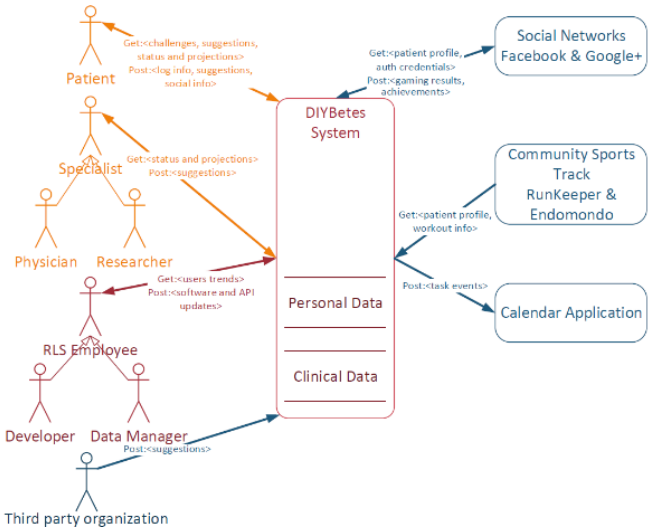


Figure 2. Context view

comments). This information is both stored and managed in order to identify new suggestions/tasks for other users that share the same profile. The log info is stored in the Clinical Data repository while the users' profile and social data is stored in a Personal Data repository.

Patients have tasks like performing a workout, log a meal, take drugs, and/or log their clinical values (instant sugar, pulse, blood pressure, blood sugar level readings, HbA1c, and weight). If the patient accepts the task and provides the requested values, these are processed by the system. Thus, the system can identify patients' trends, calculate projections towards health objectives, and identify the sub-set of tasks that are suitable considering this patients' progress. All this information is then stored in the Clinical Data Repository.

DIYbetes System to Patient

This data flow represents challenges (new tasks that the user has to perform in a daily basis or sporadically, where these tasks are defined in accordance with the patients' objectives), status and projections (based on the patients' trends the system identifies how distant patients are from their objectives), and suggestions (examples include places to workout, restaurants, recipes, menus, exercises, meal plans, and so on). This information is read and updated from Personal Data and Clinical Data repositories.

Sporadically, or in specific time intervals, each patient receives a list of individual and group challenges known as 'tasks'. These tasks coach patients to adopt beneficial behaviors, and target patients' health and lifestyle objectives. The system identifies new tasks based on the patients' progress in previously assigned tasks. Additionally, as users' submit their daily logs, the system calculates current progress and related projections. Thus, patients query this information sporadically to monitor their progress, combine different indicators, and verify the areas where they need to improve. Based on the patients' progress, location, and habits, the system provides new suggestions (meals, plans, workout places, and restaurants) designated to help users to improvie their behaviors.

 ISBN: 978-1-61208-422-0

Specialist to DIYbetes System

This data flow represents the inputs provided by a specialist (physician or researcher). The specialists view patients' progress in the system, submit updates on their patients' plan, accept or reject suggestions provided by the system, and provide some logic to identify new treatment patterns. The system stores these updates, and uses this information to validate and refine their algorithms. Context Information Specialists are studying patients trends and reactions to treatments. This study is conducted based on the logs provided by patients and stored in the Clinical Data repository. Whenever specialists identify new tasks, they submit these improvements in the system. The system then uses this information to update the patients' plan. Additionally, specialists provide objectives and thresholds for their patients, which are used to update patients' plans. Specialists can also request from the technical support team new algorithms, statistical functions, and new ratios. Although these improvements are submitted by the technical team (RedLight Software (RLS) employees) the knowledge comes from the specialists.

RLS employee

The RLS technical staff can develop new Application Programming Interfaces (APIs) to extend the DIYbetes features, or modify the services in order to increase the performance in the system. These APIs are plugged into the system (the required changes are made in the system's interfaces), or when a new infrastructure is deployed, and both updates are ready to be used by the system. This can occur e.g., when context information A specialist requests RLS to include a new statistical function or algorithm, or when the DIYbetes steering commitee identifies a new opportunity that translates into a new feature. In both these situations a developer reviews the system's design, identifies the required software, develops a new API, and deploys the software/hardware in the system, without interrupting the system's execution.

Third Party Organizations

Context Information Third Party organizations like restaurants, health clubs, hospitals, and laboratories, wish to promote/suggest their services. These organizations can use the DIYbetes system interface to promote information related with their products and services. Hence, the DYIbetes system can use third party products as suggestions/recommendations, and forward them to the patients when their profile is somehow related with the suggestion. Therefore they can request RLS to include/extend interfaces that can be used by their services and information sources. When this connection is established, the third party organization is able to send, for example, healthy menus, restaurants, or gyms near the patients' location.

Social Networks

A patient or specialist can choose to register/login using their social network credentials. The DIYbetes system forwards to users (patients, researchers, or physicians) the social network authentication page, which they use to login in their social networks. After authentication succeeds in social networks, they provide user identification to the DIYbetes service, which then forwards a session to the user. The system then requests the user's profile, which is forwarded by the associated social network. The DIYbetes system besides authenticating users, also requests profile information (such as e-mail, age gender, and location). The system uses this profile data, thus refining patient's tasks, identifying related groups (groups based on the patients' profile like location, age, gender, and so on), and managing patients' statistics.

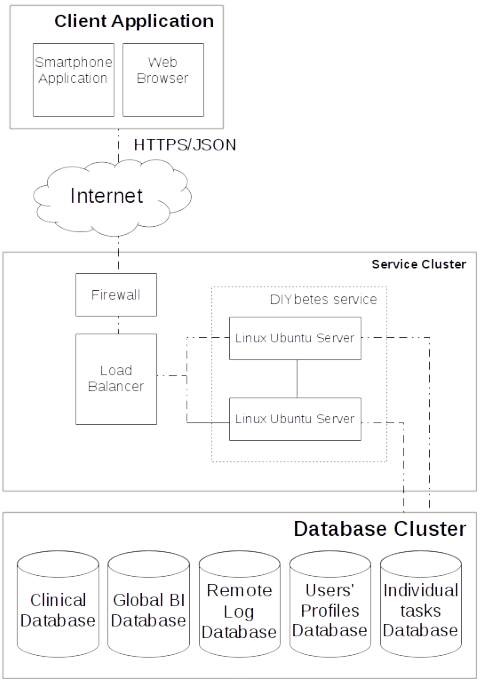


Figure 3. Deployment view

B. Deployment View

We shall now discuss where each part of the architecture will live at deployment. The Deployment View (Figure 3) describes the environment of each system's application and service highlighted in the context view.

Client Application Layer

The *Client Application Layer* is responsible for the support of the DIYbetes client applications, both web and mobile, used by patients and specialists (researchers and physicians).

Smartphone Application

The *Smartphone Application* is responsible for the interface between the patient(s) and the system. It allows patients to log their tasks (workout sessions, daily meals, drugs intakes, and health values), specify and update objectives, and manage their individual and group tasks. The Smartphone Application has location features related with the management of new suggestions (patients receive restaurants, meals, and workout places related with their location), and context aware tasks. It includes connection interfaces designated to support external sensors, which have the objective of complementing daily logs, as well as an internal database, which allows using some application features without connectivity. However, the dynamic information, like social updates, gets deprecated

 ISBN: 978-1-61208-422-0

without connectivity to the DIYbetes Service. This component has several technical restrictions namely a network connection (wireless and mobile networks, 3G and 4G). It uses HTTPS and JSON protocols.

Web browser

The *Web browser* is responsible for the interface between the user (patient, physician, and researcher) and the DIYbetes Service. The Web browser allows patients to monitor their progress (thus, users can view graphs with their progress towards their objectives), verify their task results, manage social and gaming updates, manage objectives and alarms, and configure the connection with their physicians. The Web browser allows physicians to monitor their patients' progress, manage statistics, update patients' plans and objectives, and specify alarms. Each Web browser includes cookies that are responsible for holding the session with the DIYbetes service.

Service Cluster Layer

The *Service Cluster Layer* is responsible for supporting the DIYbetes servers and support infrastructures. This Service Cluster is composed by multiple servers distributed across different geographic locations. Although each global location (Europe, US, and Canada) has their own servers, this layer comprises all the infrastructures for each of these service instances.

Load Balancer

The *Load Balancer* is the service responsible for the distribution of the workload for different DIYbetes servers. The workload distribution is configured based on the geographic location of the services (each area has a DIYbetes service instance), since each region has its own restrictions. The load balancer performs automatic fail over recovery, so when a server is not available, this entity forwards the requests to a different server with lower workload. The deploy of a load balancer should follow three phases, 1) initially it is deployed in one of the service machines. With an increase of the workload 2) a load balancer is deployed in a dedicated machine. Finally, 3) a cluster of load balancers is available if a since load balancer instance is not able to support all the accesses.

DIYbetes service

The *DIYbetes service* is a cluster of service instances running in a group of server, virtual or physical. Each geographic region has their on DIYbetes Service instance. Each DIYbetes Service instance is independent from each other, they only share the Global BI Database.

Linux Ubuntu Server

The *Linux Ubuntu Server* is the server responsible for holding the service infrastructure related with the REST DIYbetes Service. It also manages the HTTPS requests and forwards them to the REST DIYbetes Service and supports the interaction with the Database Cluster through postgres-adapter. This is a critical component of the platform, so the architecture was designed in order to be able to support stringent latency and throughput quality attributes: above 10.000 responses in 1 second; above 10.000 simultaneous requests with a maximum latency of 5 seconds and average latency below 1 second.

Database Cluster

The *Database Cluster* is a cluster of database servers (virtual or physical). Each geographic location has their group of database servers (Individual Tasks/Objectives Database, Group Tasks/Objectives Database, Alarms/Recipes/Drugs Database, and User Profile Database) and clinical database servers (Clinical Database). The Global BI and Remote Log Databases are shared between different DIYbetes services from each region.

Quality Attributes

Due to the criticality of the platform a large set of quality attributes were considered since the projects very beginning, namely latency, throughput, availability, modifiability, maintainability, testability and security. Each quality attribute was thoughly described and relevant test cases performed.

C. Design decisions

A number of design decisions were taken that involved a number of trade-offs. This decision process is detailed below:

1) A relational database was the choice for data repositories that involve queries with multiple joins. Thus, most of the operations require support for these type of relations between data. NoSQL Databases can be used to cache logs when they arrive at the system and before they are processed for statistics purposes. *Trade-off:* Dynamic languages allow to include new features and modules at runtime, and the access to database require less effort to program. However, this simplicity hides poor performance in functions where the complexity is non-linear (e.g., sequence of whiles).
2) The web interface is implemented in a dynamic language (Ruby), and the threads and message queues are using JRuby (Java is used in the thread engine and message queues structures). Thus, threads and message queues have a good access scheduler and performance in Java.
3) Ruby simplifies the integration with non-dynamic languages, so functions where the complexity is non-linear will be later (re)implemented in Java and dynamically plugged into the code. This approach partially handles the performance problem.
4) REST was adopted as the service technology. This approach allows to share logic between web and mobile access interfaces, and it is supported by dynamic languages, which allow to change components at runtime. *Trade-off:* the use of REST requires less effort to implement, but it only should be used when there is a hierarchy of accesses. This hierarchy is appropriate considering the project's information structures, since specific users have access to specific features. If such structure did not exist, web services would be a better option. Moreover, since REST will be used, the service will not have a WSDL that clarifies the structure of the service and the methods that it provides. Thus, we had to devote additional

 ISBN: 978-1-61208-422-0

effort to document the API that is used to integrate services with external parties.

5) The system currently does not support clinical protocols (HL7 or DCOM), since not all healthcare systems implement them, or can be integrated with external systems. Thus, supporting this diversity was postponed due to the size of the technical team.
6) The service will have two keys, which services shall use for read or write information from the Clinical Database. These keys are required to authenticate services in these databases, but no service will have the 2 keys simultaneously, in order to limit database read and write operations. *Trade-off:* The approach described promotes security, but does not protect against attacks that result from the organization staff. Thus, an external database was added that has all the database and OS logs. Therefore, all authentications in the hosting machine, access to the database, all database reads and/or writes are logged.
7) It was decided to deploy two parallel servers, so that we are able to install updates without making the service unavailable. Additionally, no effort will be required to implement complex availability mechanisms. If a server has a fault, the other server is able to support users' requests.
8) The project considers an internal Load Balancer, since it allows the organization specifying custom rules to balance the workload (examples of rules include geography location, identification of the users, hours of the day, and so on). Moreover, a custom approach simplifies the deployment of different A/B testing scenarios. Finally, this approach is essential to promote service uploads without the need to have the service totally unavailable. *Trade-off:* a custom load balancer implies more processing at the server side and additional effort to configure and deploy the rules that will manage Load Balancer operations. Considering the importance of the system's flexibility and performance, this drawback does not justify a different approach. Moreover, a local Load Balancer shall be used, since it allows to reuse the same container technology adopted in the service infrastructure (Nginx). However, this local load balancer has a negative impact in the system's performance. Finally, while Linux OS has more security mechanisms than Windows OS, it has less that OpenBSD. However, we did not consider that OS security mechanisms are a stringent requirement. To manage this issue there is a remote, independent OS BD, which will log all database accesses and server authentications.
9) Each server runs a complete instance of the service, which simplifies the scalability and configurability issues, since if more processing power is needed more servers with more service instances can be added and only minor configurations are required in the Load Balancer.
10) Linux OS was the OS selected for test and product platforms, since most of the organization developers are used to its environment and it adequately fits the purpose.

V. CONCLUSION

In this paper, we presented a multiplatform solution (mobile and web), that addresses behavior change of Type 2 Diabetes patients, and described the set of technologies underlying the DIYbetes platform. It is a complex piece of technology that adresses a set of demanding non-functional (quality) attributes in order to operate in harsh conditions with a low technological literate population. The current solution has just been deployed in Portugal for the patients of the national healthcare system (literally all interested citizens) and will be further expanded to other geographies. A partnership with a USA-based company will allow testing the platform for a different language and geography.

ACKNOWLEGMENT

This project was was supported by a grant from the national COMPETE SI&IDT program (project nr. 34006/2013, DIYbetes). The authors would also like to thank the contribution from all the patients, doctors and nurses involved.

REFERENCES

[1] Diabetes Action Now, World Health Organization and the International Diabetes Federation, 20pp. 2004 ISBN: 924159151X

[2] International Diabetes Federation: Epidemiology and morbidity. [http://www.diabetesatlas.org/content/diabetes-and-impaired -glucose-tolerance]

[3] Berger B, Stenstrom G, Sundkvist G. Incidence, prevalence, and mortality of diabetes in a large population. A report from the Skaraborg Diabetes Registry. Diabetes Care. 1999;22:773-8.

[4] Kerr EA, McGlynn EA, Adams J, Keesey J, and Asch SM, "Profiling the Quality of Care in Twelve Communities: Results from the CQI Study," Health Affairs, Vol. 23, No. 3, May/June 2004, pp. 247-256.

[5] Mosa A., Yoo I.,Sheets L, A Systematic Review of Healthcare Applications for Smartphones, BMC Medical Informatics and Decision Making 2012.

[6] Mobilehealthnews (2010,Oct), http://mobilehealthnews.com/9168/why-the-att-welldoc-deal-matters/ [21] Jasinski,D.J. (2012, Apr) IBM Healthcare, interview.

[7] Wu RC, Morra D, Quan S, Lai S, Zanjani S, Abrams H, Rossos PG: The use of smartphones for clinical communication on internal medicine wards. J Hosp Med 2010, 5:553-559.

[8] Soto RG, Chu LF, Goldman JM, Rampil IJ, Ruskin KJ: Communication in critical care environments: mobile telephones improve patient care. Anesth Analg 2006, 102:535-541.

[9] Hasvold PE, Scholl J: Disrupted rhythms and mobile ICT in a surgical department. International Journal of Medical Informatics 2011, 80:e72-e84.

[10] Adults' Media Use and Attitudes Report 2014, Independent regulator and competition authority for the UK communications industries, Ofcom Research Document, April 2014.

[11] Glooko is a mobile application from a Silicon Valley company that delivers a mobile and cloud based, diabetes & chronic disease management system. https://glooko.com/ [retrived March 2015]

[12] Diabetes: Fatos e Números 2012, Relatório Anual do Observatório Nacional de Diabetes (4a edição), 19 de Fevereiro de 2013.

[13] European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD))

[14] Medicare and Medicaid Programs; Electronic Health Record Incentive Program-Stage 3, online at https://s3.amazonaws.com/public-inspection.federalregister.gov/2015-06685.pdf (March 2015)

 ISBN: 978-1-61208-422-0

An Analysis of Educational Big Data in University Using Mobile E-Portfolio System with Smart Concierge

Noriko Hanakawa
Hannan University
Department of Information Management
Osaka, Japan
Email:hanakawa@hannan.ac.jp

Masaki Obana
Osaka Institute of Technology
Department of Information Science and Technology
Osaka, Japan
Email:obana@is.oit.ac.jp

***Abstract*—We have developed a new educational e-portfolio environment including not only conventional education data but also school life data in university. School life data means non-lectures activities, such as job hunting, club activities, students' communications. A most important feature of the e-portfolio environment is a school concierge named "HapiNan" in smartphone in order to navigate students' school life. The school concierge asks some questions in a smartphone according to student situations. The smart concierge provides mobility and ubiquity of the educational e-portfolio environment. Our goal is to clarify influences of school life activities on university learning. From April to September 2013, a trial version of the system ran. From April 2014, freshmen in our university used the system. We have two phases in order to analyze e-portfolio data. In the first phase, fourth year students used the e-portfolio system. As a result, students' job hunting problems are clear. (1) Job hunting time is 40hours per a week, (2) no relations between Grade Point Average (GPA) and job hunting success, (3) job hunting cost is large. In second phase, freshmen used the e-portfolio system in the first year of education. We found that there was a relationship between GPA and lecture time as well as hobby time. In addition, there was no relation between GPA and home study time or part-time job time. The analysis results are valuable for school advisors and teachers in order to support freshmen's school life.**

Keywords- e-portfolio; smartphone; SNS; GPA; school life; university education.

I. INTRODUCTION

Educational e-portfolio is very important in Japanese universities. Of course, e-portfolio data has to accumulate in educational database. The educational records are study reports, examination results, lecture materials, lecture notes, and goals of lectures and achievement levels of lectures. The portfolio data is useful to improve students' understanding levels and knowledge acquisition. Educational portfolio is becoming more popular and more useful [1] [2]. The efficiency of having an education portfolio is becoming clearer and clearer [7] [8].

However, in Japanese universities, students would like to support not only educational activities but also school life activities, such as club activities, job hunting, part-time job, and communication among students. Especially job hunting activities takes much time for the third and the forth year students. It is a serious problem in Japanese society because university students sacrifice their study time to hunt for job. In addition, school advisors and teachers can not fully understand students' real problems that caused low Grade Point Average (GPA). Of course, we can forecast easily that GPA is low when study time at home is small. However, we cannot understand how the other students' activities, such as part-time job and club activities influence GPA. Our goal is to clarify the influence of school life activities on university learning.

Therefore, we have developed a new e-portfolio environment for supporting school life including various non-educational activities. The most important feature of the e-portfolio environment is a school concierge character called "HapiNan" who urge students to input usual activities information in students' smartphone. For example, if a student does not input job hunting activities, the concierge asks the student "Hello, how is your job hunting?". The concierge asking through the smartphone is useful in order for the student to continue to input data of school life into the e-portfolio every day. In addition, the accumulated data of school life data is useful to analyze GPA. We clarify relationships between GPA values and school life data in e-portfolio system. If we find causes that lead low GPA, for example too much time for part-time job, student advisors and teachers are able to give students advices on school life, especially part-time job.

Section 2 shows related work, Section 3 explains functions of our e-portfolio environment. The trial application of the school life e-portfolio is shown in Section4, Section 5 shows summary and future research.

II. RELATED WORKS

Various e-portfolios for university education have been proposed and applied. Olatz descriptively studied the undergraduate students' perceptions, attitudes and behavior when using an e-portfolio to support their learning and assessment in practice based courses at two traditional Spanish universities [3]. As a result, the students had positive opinions and self-efficiency through the e-portfolio as a tool to manage their learning and assessment during a semester, especially from the second month of use.

Chang studied to use e-portfolios to enhance university students' knowledge management (KM) performance [4]. The research results revealed that the experimental group outperformed the control group in the performances of overall

 ISBN: 978-1-61208-422-0

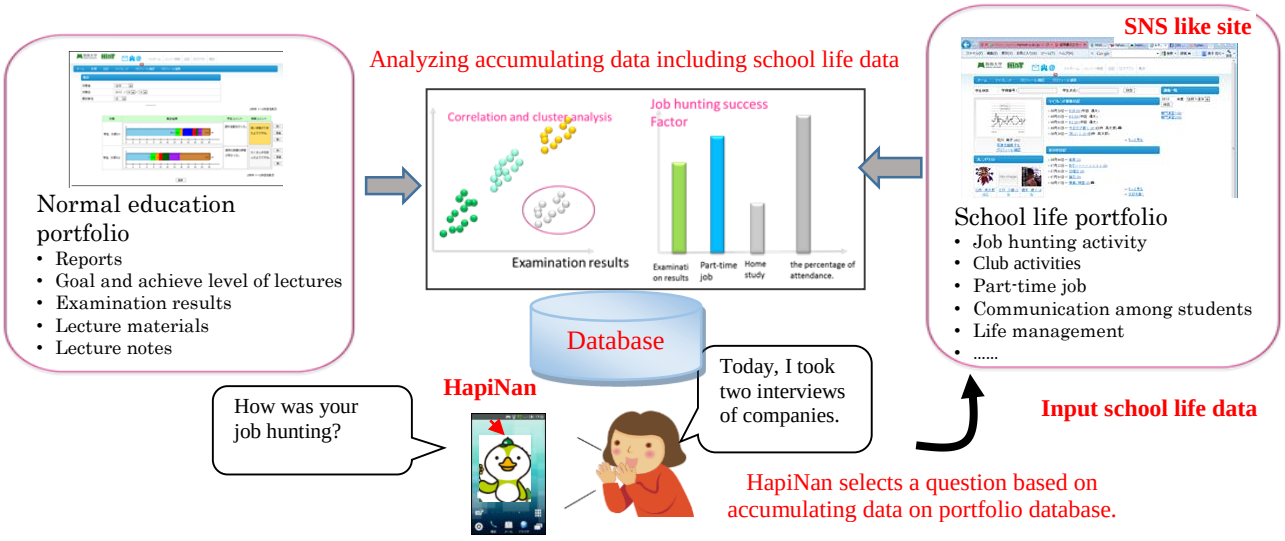


Figure 1 An outline and usage image of the e-portfolio environment

KM and five KM aspects (knowledge sharing, innovation, acquisition, application, and accumulation). This showed that e-portfolios significantly facilitated KM performance.

Carol et al. developed an e-learning system that couples a blog with a learning e-portfolio [5]. They adapted the system to the first year course education. Because the e-portfolio system is based on web-blog system, massive blog data was accumulated in the system. Rodriguez et al. shows usefulness of e-portfolio in university professional education [6]. They claimed that (1) e-portfolio is a complementary tool for student's assessment, (2) e-portfolio is a perfect follow-up device to check student's competences development throughout their degree studies. Other researches also claim there is a benefit of the e-portfolio system and data analysis. Shroff et al. analyzed students' behaviors in examinations using e-portfolio system [7], and Alexiou et al. studied the benefits of the e-portfolio system in university education. [8].

These studies are conventional education e-portfolio systems. The usefulness and effects of educational e-portfolio have already been clear. Therefore, we try adding a new function of the school life e-portfolio to the conventional education e-portfolio in order to support the entire university school life for students.

III. THE EDUCATION E-PORTFOLIO ENVIRONMENT WITH THE CONCIERGE

A. Overview

Figure1 shows an outline of the e-portfolio environment and usage image. The e-portfolio environment includes normal educational portfolio and school life portfolio. In lectures, students and teachers use the normal educational portfolio. On the other hand, when students do not take lectures, students use school life portfolio in smartphone. Students can input school life data anytime and anywhere. Therefore, a tool for inputting school life information is required. Because students are not compelled to input school life data, students may forget to input the data. For this reason, we included in the smartphone application the school concierge "HapiNan". HapiNan is our university mascot character.

B. A normal educational e-portfolio

Figure 2 shows a web page of goals and achievement levels of lectures in the normal educational e-portfolio environment. The educational e-portfolio is a typical education portfolio having the following functions:

(a) Accumulating lecture materials
(b) Accumulating lecture notes
(c) Accumulating reports
(d) Accumulating examination results
(e) Accumulating attendance situation of lectures
(f) Accumulating teachers' comments
(g) Accumulating GPA

Similar to the conventional education e-portfolio system, our e-portfolio environment supports all functions of the conventional education e-portfolio system.

C. A school life e-portfolio

A school life e-portfolio is an original function in our e-portfolio environment. The school life e-portfolio is like SNS (Social Network Service), such as Facebook (See Figure 3). Because the school life e-portfolio supports diary function the same way as Facebook diary function, students input their dairy activities to school life e-portfolio without a feeling of wrongness.

The school life e-portfolio can collect the following data:

(a) Time to take lectures in school
(b) Time to study at home
(c) Time to do extra-curriculum activities
(d) Time to work on part-time job
(e) Time to do hobby
(f) Time to sleep

 ISBN: 978-1-61208-422-0

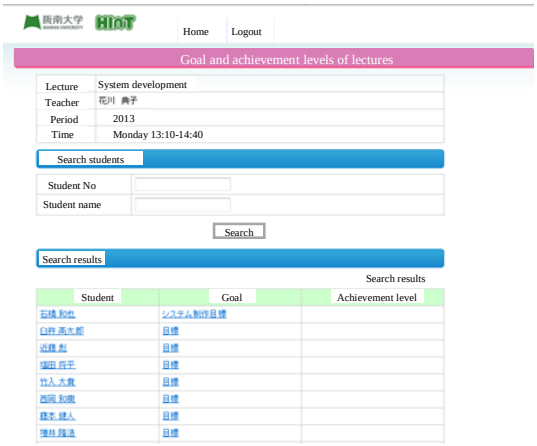


Figure 2 A web page of goals and achievement levels of lectures in normal educational e-portfolio

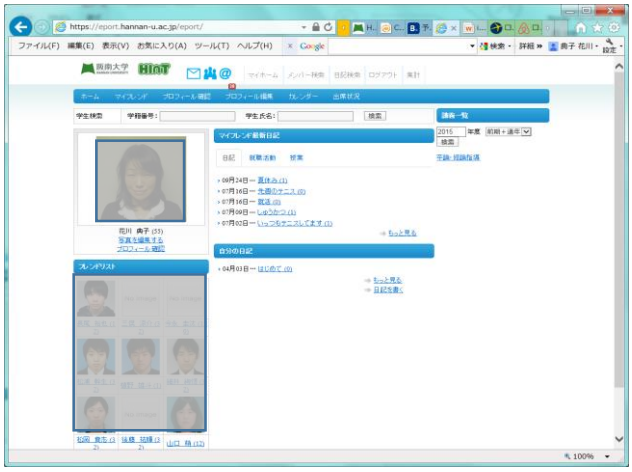


Figure 3 A web page of diary in school life e-portfolio

In addition, the school life e-portfolio includes special functions because students can easily input their dairy activities. The special functions are the following:

(1) Dairy function like social network, such as Facebook
(2) School concierge "HapiNan" in students' smartphones
(3) Time management function in usual life
(4) Analyzing data of educational data and school life data

The following sub-sections explain the above functions.

1) Dairy function like social network, such as Facebook

Time management function is an original function. The time management means that students report usual activities, such as job hunting time, home study time, part-time job time, and sleeping time. The time management function's purpose is to understand each student situation in usual life. For example, when a student does not submit a reports by due day, a teacher may understand the reasons for the late submission might be due to, for example, the fact that the student spends a lot of time in a part-time job. Especially, in current Japanese society, job hunting in the third and fourth year students is very serious. Job hunting is a main activity in university on the fourth year instead of educational activities. Teachers and school staff support the serious job hunting activities. In the supporting activities of teachers and staffs, the time management function of the school life e-portfolio is useful. Figure 4 shows an image of time management function when students input the usual time in a web page.

2) School concierge "HapiNan" in students' smartphones

The above functions, such as diary and time management required students input actions their usual activities every day. However, usual activity data is not inputted unless almost all students are forced. Students forget the input, or skip the input. Therefore, we prepared school concierge in students' smartphone. The concierge name is "HapiNan", this is our university mascot character. The HapiNan asks a question to a student based on data of the e-portfolio database (See Figure 5). If a student was absent from a lecture, "Why are you absent from the lecture?", HapiNan asks. Moreover, if a student does not report job hunting activities, "How is your job hunting?", HapiNan asks. This way, students' educational data and school life data are accumulated to the e-portfolio database.

3) Time management function in usual life

Dairy and usual life time management data are added on each student. Graphs of the data of each student are generated in the analyzing data functions (See Figure 6). In the web page of the graphs, teachers and staffs can input comments about the time management data. The function is a kind of communication tools between students and teachers.

4) Analyzing data of accumulated educational data and school life data

The e-portfolio system also has important analyzing functions. Figure 7 shows an example of results of the analyzing functions. The left side graph of the Figure 7 shows a relationship between home study time and examination results. The home study time is a kind of time management data of school life portfolio, examination results (GPA) is a kind of data of the educational portfolio. The right-side graph means success factors of job hunting activities. For example, students who spend a lot of time for part-time job can get

Figure 4 An input function of the time management of school life e-portfolio

 ISBN: 978-1-61208-422-0

Figure 5 School concierge "HapiNsan" in a smartphone

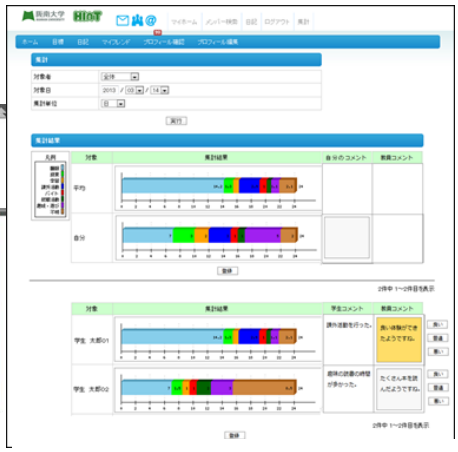


Figure 6 A web page of total sum of time management and teachers comments

earlier success of job hunting than students who get better examination results. Job hunting success data is recorded in school life portfolio, examination results are accumulated in educational portfolio. In this way, the analyzing functions of our e-portfolio environment can be clear the relationship between education data and usual life data.

IV. ANALYSIS OF E-PORTFOLIO DATA

The trial version of the educational e-portfolio system has two phases. The first phase is from April to September 2013. Eleven students used the system in trial. After several functions improved according to the students advices, all freshmen used the e-portfolio system from April 2014 (second phase). The number of the new students is 1000 or more. In this section, at first, we show results of the trial in the first phase. Next, we show the results of all freshmen in the second phase. The data of the second phase was collected for six months, from April to September of 2014.

A. The first phase results (trial phase)

We try adapting the school life e-portfolio to 11 students who are the fourth year students. They were doing job hunting activities. The accumulated educational data and school life data were analyzed. As a result, we found the following points:

(a) Students spend 40 hours a week for job hunting, and students spend 100 hours a month in a job hunting peak period.

(b) Correlation between examination results and success of job hunting is not strong.

Figure 8 Relationship between job hunting success time and job hunting time

(c) The most important challenge in job hunting activities is high transport cost, such as train fee, or bus fee, and airplane fee.

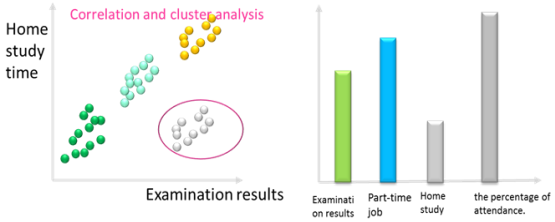


Figure 7 Analysis results of educational portfolio data and school life data

These results are derived from an analysis function of the e-portfolio system. (a) is derived from school life portfolio, (b) is derived from educational portfolio and school life portfolio. (c) is derived from the dairy function like SNS site of school life portfolio. In this way, because school life data is accumulated, students' portfolio becomes more useful in university education.

Figure 8 and Figure 9 show graphs of job hunting time success. These graphs are generated automatically in our e-portfolio environment. Relationships between job hunting success and time spent on job hunting is shown in Figure 8. Figure 9 shows relationships between job hunting success and input rate of school life data to the e-portfolio system. "Student 2" spent a lot of time on job hunting. However, timing of job hunting success was late July of 2013. In contrast, although "Student 3" spent little time onjob hunting, timing of job hunting success was early April of 2013. In this way, students' behaviors that are not only learning activities but also private life activities were clarified. School advisors were able to give precise comments to each student.

B. The second phase results (all freshmen)

In second phase, 1296 freshmen in our university used the e-portfolio system from April 2014. The e-portfolio system was useful supporting the freshmen school life. Especially, school advisors and teachers use the system in the first year education. The education data and school life data of the e-portfolio system were collected for six months. In addition, values of GPA of the freshmen were compared with data of school life in the e-portfolio system.

Relationships between GPAs and school life data of all freshmen are shown in Figure 10. Because all freshmen did not input all school life data, missing data increased (See "No input" of Figure 10). Therefore, we eliminated the missing data from all school life data.

As a result, we found the following:

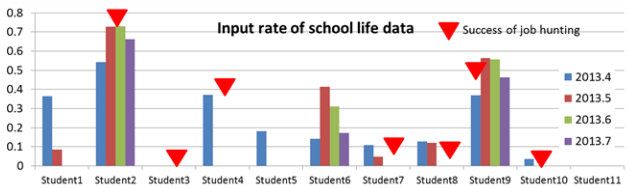


Figure 9 Relationship between job hunting success time and input rate

 ISBN: 978-1-61208-422-0

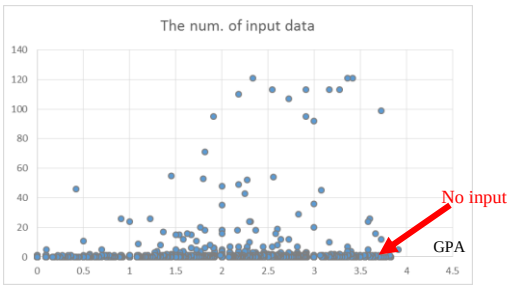


Figure 10 Relationships between GPAs and the number of input data

1) GPA has weak relationship with the number of input data
2) GPA has weak relationship with time of taking lecture
3) GPA has weak relationship with hobby time
4) GPA has no relationship with part-time job
5) GPA has no relationship with home study time

The following subsections show the results of analysis of the data of the e-portfolio system.

1) GPA has weak relationship with the number of input data

"(1) Relationship of the number of input data" of Figure 11 shows relationship between GPAs and the number of input data. Teachers and school advisors encouraged the freshmen to input school life data to the e-portfolio system. Diligent students input the school life data according to teachers' and advisors' suggestions. Of course, the diligent students had also good GPA values. A value of correlation between GPA and the number of input data is 0.336. Therefore, GPA has weak relation with the number of input data.

2) GPA has weak relationship with time of taking lecture

"(2) Relationship of lecture time" of Figure 11 shows relationship between GPAs and time of taking lectures. The earnest students also took many lectures. Naturally, the students that take many lectures get high score of GPA. A value of correlation between GPAs and time of taking lectures is 0.43. Therefore, GPA has weak relationship with time of taking lectures.

3) GPA has weak relationship with hobby time

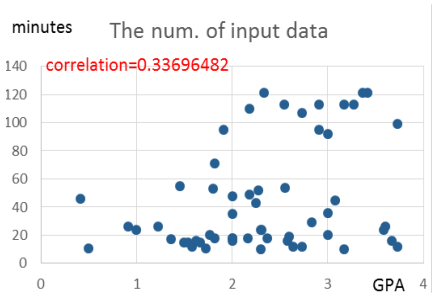
"(3) Relationship of hobby time" of Figure 11 shows relationship between GPAs and hobby time. Hobby includes "playing with friends", "watch TV", "playing computer game", "do sport", and "reading books". Usually, we think that GPA will become low if students spend a lot of time for hobby. However, correlation between GPA and hobby time is 0.36. That is, a student who spends much time for playing his hobby has good GPA. The relationship between GPA and hobby time was unexpected.

4) GPA has no relationship with part-time job

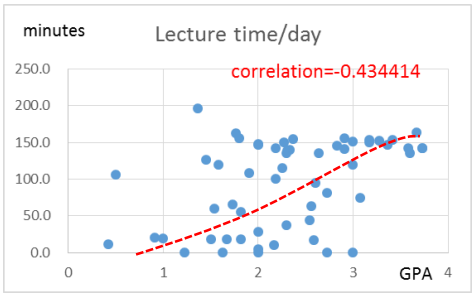
"(4) Relationship of part-time job time" of Figure 11 shows relationship between GPAs and part-time job time. School advisors and teachers often tell "your GPA will become low if you are absorbed in a part-time job. So, you should stop part-time job.". This phrase is repeated in a case of students' consultation with school advisors. However, GPA has no relationships with part-time job time. Even if a student works 176 minutes as part-time job per a day, the student has good GPA (3.6). School advisors and teachers have to change their advices regarding part-time job.

5) GPA has no relationship with home study time

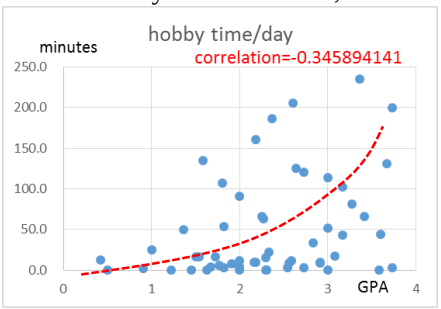
"(5) Relationship of home study time" of Figure 11 shows relationship between GPAs and home study time. Of course, we expected strong correlation between GPA and home study time. That is, when a student hardly studies at home, his GPA



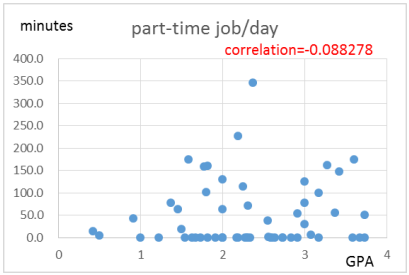
(1) Relationship of the number of input data



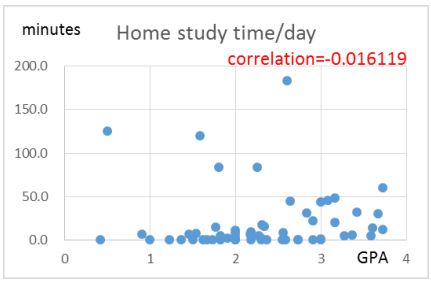
(2) Relationship of lecture time



(3) Relationship of hobby time



(4) Relationship of part-time job time



(5) Relationship of home study time

Figure 11 Relationships between GPAs and the school life data

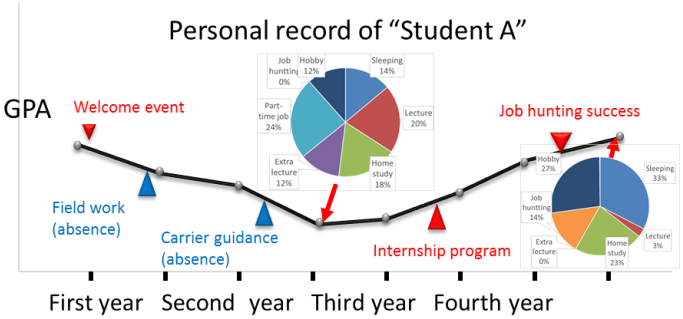


Figure 12 Personal record with GPA and e-portfolio school life

becomes high. However, the analysis result is different from our expectation. GPA has no relationships with home study time. We think that quality of home study is more important than length of home study.

C. *Discussion of results of the analysis*

In second phase, the above 1) to 5) are important results of the analysis of the e-portfolio data. The targets of the second phase were freshmen in our university. The freshmen took the first year's education, not professional education. Because of first year's education data, the results of the analysis of the e-portfolio data are limited to the first year's education. However, from the results, we encourage freshmen to spend sufficient time taking lectures, while at same time reserve sufficient time for hobbies. The amount of time spent on home study is not important. It is good to consider a part-time job.

On the other hand, the first phase analysis of the e-portfolio data focuses on job hunting activities of the fourth years' students. The second phase analysis of the e-portfolio system focuses on the freshmen, and the first year's education. These analysis targets were only same year students. The e-portfolio system has an ability of all year students. Growth of a students during 4 years in university life can be recorded to the e-portfolio system. In addition, there are many events that are not learning activities, for example, job hunting activities, club activities, and hobby activities. Table 1 shows school events that are not learning activities in our university.

TABLE I. SCHOOL EVENTS

Year	Events of each year	All years
The first year	Welcome sport events and party Fieldwork for making teamwork	University festival,
The second year	Carrier-up guidance and program	
The third year	Internship program Carrier-up program,	
The fourth year	Job hunting	

The e-portfolio system accumulates learning data, such as GPA, lecture time, reports. Moreover, students' school life data, such as part-time job, home study, sleeping time, hobby time is also important. The school events, such as university festival, internship program, job hunting are also supported to the e-portfolio system. Therefore, the e-portfolio system can show personal history with a GAP change, a school life data change, and school events (See Figure 12). The personal records are important for university's education. If school advisors found a decline of GPA of "Student A", the advisors can check school life data of "Student A" (See "Second year" point of Figure 12). Advisors may find the intrinsic problems of the student in school life. In contrast, "Student A" can check by himself. During 4 years, he can review his school life. The review results may be useful to judge his important decisions, such as working style, job-change, in future.

V. SUMMARY

We have developed an e-portfolio environment including normal conventional portfolio and school life portfolio. School life portfolio supports non-lecture school life, such as job hinting activities, part-time job, club, and communication among students. In addition, we developed a school concierge "HapiNan" in smartphone in order to support frequent inputs of students. We have two phases in order to analyze e-portfolio data. In the first phase, fourth year students used the e-portfolio system. As a result, we observed the following (1) Job hunting time was 40 hours per a week, (2) there was no relations between GPAs and job hunting success, (3) job hunting cost was large. In the second phase, freshmen used the e-portfolio system in the first year of education. We found that GPA has relationships with lecture time and hobby time. In the future, we will develop a historical personal analysis tools of the e-portfolio data.

ACKNOWLEDGMENT

This research was partially supported by KAKENHI, Grant-in-Aid for Scientific Research(C), 26330093, 2015.

REFERENCES

[1] R. Brian. von Konsky and O. Beverley, "The iPortfolio: Measuring uptake and effective use of an institutional electronic portfolio in higher education", Australasian Journal of Educational Technology, Vol.28, No.1, 2012, pp.67-90.

[2] P. Kim, C. Kee Ng, and G. Lim, "When cloud computing meets with Semantic Web: A new design for e-portfolio systems in the social media era", British Journal of Educational Technology, Vol.41, Issue 6, November, 2010, pp.1018-1028.

[3] O. Lopez-Fernandez and J. Luis Rodriguez-Illera, "Investigating university students' adaptation to a digital learner course portfolio", Computers & Education, Vol.52, Issue 3, April, 2009, pp.608-616.

[4] C. Chang, K. Tseng, C. Liang, and T. Chen, "Using e-portfolios to facilitate university students' knowledge management performance: E-portfolio vs. non-portfolio", Computers & Education, Vol.69, November, 2013, pp.216-224.

[5] N.L. Carroll, R.A. Calvo, and L. Markauskaite, "E-Portfolios and Blogs: Online Tools for Giving Young Engineers a Voice", Information Technology Based Higher Education and Training, 2006. ITHET '06. 7th International Conference on, July, 2006, pp.1-8.

[6] S. Rodriguez-Donaire, B.A. García, and S.O. del Olmo, "e-Portfolio: A tool to assess university students' skills", Information Technology Based Higher Education and Training (ITHET), 2010 9th International Conference on, April, 2010, pp.114-124.

[7] R. H. Shroff, C. C. Deneen, and E. M. W. Ng, "Analysis of the technology acceptance model in examining students' behavioural intention to use an e-portfolio system", Australasian Journal of Educational Technology, Vol.27, No.4, 2011, pp.600-618.

[8] A. Alexiou and F. Paraskeva, "Enhancing self-regulated learning skills through the implementation of an e-portfolio tool", Procedia - Social and Behavioral Sciences, Vol. 2, Issue 2, 2010, pp.3048-3054.

 ISBN: 978-1-61208-422-0

Enhancing Ultrasonic Robot Positioning Accuracy with Parallel Codes Acquisition of Composite Pseudo-noise Sequences

Jen-Fa Huang, Chun-Chieh Liu, Kun-Fong Lin

Advanced Optoelectronic Technology Center,
Institute of Computer and Communications Engineering
Department of Electrical Engineering
National Cheng Kung University, Tainan, Taiwan
Email: huajf@ee.ncku.edu.tw

***Abstract*—Ultrasonic robot positioning with composite codes acquisition is investigated in this paper. The indoor robot positioning system was previously examined with single Pseudo-noise (PN) signal sequence. In views of correlation acquisition, the longer the code acquisition time, the longer the path estimation distance, and the worse the robot positioning accuracy. Under comparable period lengths, acquisition time for composite PN codes can be shorter than that of pure PN codes, thus can largely enhance the robot positioning accuracy. In the devised system configuration, three transmitters continuously send out their ultrasonic coding signals to the robot receiver. The robot evaluates its current position by measuring time difference of arrival (TDOA) among the three paths. Optimization algorithms can then be undertaken over the measured TDOAs to obtain more accurate robot location. Based on correlation characteristics of the proposed composite PN codes, we finally make a general analysis on codes acquisition time to the robot positioning accuracy.**

Keywords -- Indoor positioning system; Composite M-sequences; Parallel codes acquisition; Time difference of arrival (TDOA).

I. INTRODUCTION

With the mature technology, the functionality of robots is more and more pluralism. For example, the navigation robot, the cleaning robot, and other service type of robots, when robots execute their task, they need to move around. Therefore, the accuracy of positioning is very important, and the error of measurements between robot and sensor must be solved. For example, the multipath propagation is caused by the interference, because the ultrasonic wave is transmitted at all direction. As a result, multipath propagation will occur when the ultrasonic wave collide obstacles. Transmitting signals may be cut by obstacles so that a longer distance and a large time delay are produced. Time of Arrival (TOA) [1][2] and Time Difference of Arrival (TDOA) [3][4] positioning are easily influenced by errors so that the positioning accuracy is reduced.

In order to improve indoor ultrasonic positioning accuracy, so the robot object can be more precise positioning, and capture ultrasonic signals in the process. How to confirm the capture of ultrasound echo signals to the correct sources and reduce errors is the most important issue to study.

Several previous works that have used the coding techniques of the ultrasonic signal to determine the robot position, using PN sequences [5][6], Gold sequences [7], Loosely Synchronous (LS) sequences [8], Golay codes [9] and Barker codes [10]. These works represent the development of a Local Positioning System (LPS), based on the transmission of ultrasonic signals.

Pérez et al. [8] explored characteristics of LS sequences which exhibit an Interference-Free Window (IFW) within correlation functions to construct an ultrasonic beacon-based LPS, as well as to reduce the multipath effect. Hernández et al [9] developed system which used Golay codes in the ultrasonic signal processing and obtained features of arbitrary long pseudo-orthogonal sequences with no cross-interference. Hossain et al. [10] found pairs of Barker code with low cross-correlation so that they can be used in multi-user environment.

Huang et al [11] proposed a coding scheme of composite PN code sequences to encode the transmission signals. Such composite codes possess characteristics of mutual codes orthogonality and can asynchronously cancel the mutual interference among transceivers. With sophisticated balanced correlation detections, matched codes with high correlation magnitude can get unique code identification and unmatched codes will be rejected in the receiver end. De Angelis et al. [12] investigated an acquisition system to solve the problem of having more than one BS in the same PN code acquisition system to make it necessary to discriminate between correct detection and false alarm events.

In this paper, we simulate an indoor ultrasonic robot positioning scheme based on Direct Sequence Spread Spectrum (DSSS) system. Through DSSS system architecture, we make our higher power and narrow band of the original signal into a low power and broadband signals. Each transmitter is controlled by central controller. The central controller will select the assigned composite code sequences for the corresponding transmitter's unique code identification. The robot calculates the number of frame peak between local code replica and received summed sequence. The number of frame peak offers estimates of the robot distance to the corner transmitters. With such estimates of transceiver distance, the robot executes TDOA calculation and optimization to obtain its absolute location.

The paper is organized as follows. In Section II, we introduce composite code architecture in detail. Important correlation characteristics are investigated for parallel codes acquisition to estimate robot distance to transceivers. In Section III, composite PN codes are assigned to indoor

 ISBN: 978-1-61208-422-0

corner transceivers to determine the absolute position of robot object by hyperbolic triangulation of the distance obtained from the measurement of the difference in TDOA among a transceiver and the others. In Section IV, with parallel codes acquisition scheme, we analyze the accuracy of the positioning and expect to improve the accuracy of the indoor positioning systems. Finally, in Section V, we present our discussions and conclusions. With parallel PN codes acquisition, robot positioning error is found much reduced to provide more precise movement behaviors.

II. COMPOSITE CODE SEQUENCES

Composite code sequences constructed with M-sequence codes are a particular set of PN sequences. This family of composite sequence codes possesses high magnitude of auto-correlation and low value of cross-correlation characteristics. We assigned different composite codes to transceiver in a DSSS system and controlled by central controller, and signals can be sent simultaneously and be separated at the receiver.

In this paper, we propose a coding method for DSSS indoor robot positioning system. The assigned transceiver composite codes are made up of M-sequence component codes. There are many groups of composite codes that the ultrasonic transceiver can be assigned with. Now, we select two M-sequence codes to illustrate a composite code set of them. Let $\boldsymbol{C}_1$ be an (n_1, k_1) binary M-sequence code and $\boldsymbol{C}_2$ be an (n_2, k_2) binary M-sequence code, where code periods n_1 and n_2 are relatively prime. Let $\boldsymbol{C}_1(X) \in \boldsymbol{C}_1$ and $\boldsymbol{C}_2(X) \in \boldsymbol{C}_2$ denote the basis code words or code vectors in code space $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$. Let $T^i\boldsymbol{C}_1(X)$ denote the i-chips cyclic right-shift of $\boldsymbol{C}_1(X)$, $0 \le i \le n_1\text{-}1$, and $T^j\boldsymbol{C}_2(X)$ the j-chips cyclic right-shift of $\boldsymbol{C}_2(X)$, $0 \le j \le n_2\text{-}1$. With $n=n_1n_2$, let code vector $T^i\boldsymbol{C}_1(X) \in \boldsymbol{C}_1$ repeat itself $n/n_1=n_2$ times and $T^j\boldsymbol{C}_2(X) \in \boldsymbol{C}_2$ repeat itself $n/n_2=n_1$ times, we obtain the repeated binary M-sequences of common period $n=n_1n_2$:

$$\left(T^i\boldsymbol{C}_1(\mathrm{X})\right)^n = (T^i\boldsymbol{C}_1(\mathrm{X}), T^i\boldsymbol{C}_1(\mathrm{X}), \cdots, (n_2 \text{ times})) \quad (1)$$

$$\left(T^j\boldsymbol{C}_2(\mathrm{X})\right)^n = (T^j\boldsymbol{C}_2(\mathrm{X}), T^j\boldsymbol{C}_2(\mathrm{X}), \cdots, (n_1 \text{ times})) \quad (2)$$

By combining (1) and (2) in a chip-by-chip modulo-2 addition, we get a composite code vector defined with the above component M-sequences

$$\left(C^{(i,j)}(\mathrm{X})\right)^n = \left(T^i\boldsymbol{C}_1(\mathrm{X})\right)^n \oplus \left(T^j\boldsymbol{C}_2(\mathrm{X})\right)^n \quad (3)$$

The notation "$\oplus$" represents a modulo-2 summation. The composite code vector of (3) are non-maximal length codes though their constituent component codes are maximal-length ones. In the above equations, we have defined $(T^i\boldsymbol{C}_1(X))^n$ and $(T^j\boldsymbol{C}_2(X))^n$ the repeated component M-sequence codes while $(\boldsymbol{C}^{(i,j)}(X))^n$ the composite codes made up from the above repeated M-sequences codes. Figure 1 depicts a schematic shift register circuit for composite M-sequence codes $\boldsymbol{C}^{(i,j)}(X) = T^i\boldsymbol{C}_1(X) \oplus T^j\boldsymbol{C}_2(X)$, where $T^i\boldsymbol{C}_1(X)$ codes are generated in the upper branch with recursion connection $h_1(X) = 1+X+X^2$ while $T^j\boldsymbol{C}_2(X)$ codes be generated in the lower branch with feedback connection $h_2(X) = 1+X+X^3$.

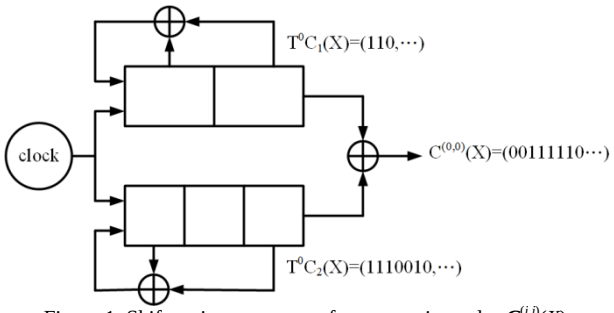


Figure 1. Shift register generator for composite codes $\boldsymbol{C}^{(i,j)}(X) = T^i\boldsymbol{C}_1(X) \oplus T^j\boldsymbol{C}_2(X)$.

On the receiver side, the goal is to capture the matched code signal to estimate position distance while reject interference from other unmatched signal codes. As depicted in Figure 2, we devise a parallel codes acquisition circuit for the robot receiver. Balanced correlators detection/subtraction scheme is adopted. In the upper circuit, the received signals will perform correlation operation with local code signal $\boldsymbol{C}_1 = (1, 1, 0)$ and $\overline{\boldsymbol{C}}_1 = (0, 0, 1)$ to capture acquisition peaks with every 3 bits cycle shift. In the lower circuit, the received signals will perform correlation operation with local signal code $\boldsymbol{C}_2 = (1, 1, 1, 0, 0, 1, 0)$ and $\overline{\boldsymbol{C}}_2 = (0, 0, 0, 1, 1, 0, 1)$ to capture acquisition peaks with every 7 bits cycle shift. The acquisition peaks combined from the upper and the lower correlators will appear at the common periodicity of 21 bits cycle shift. By using this method, the receiver can remove interference of other signals and capture the relative signal $\boldsymbol{C}_1 \oplus \boldsymbol{C}_2$ = (1,1,0, …) $\oplus$ (1,1,1,0,0,1,0, …).

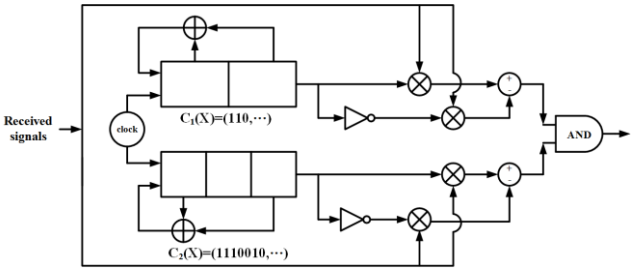


Figure 2. Parallel acquisition circuit for composite codes $T^i\boldsymbol{C}_1(X) \oplus T^j\boldsymbol{C}_2(X)$.

The composite codes $\boldsymbol{C}^{(i,j)}(X) = T^i\boldsymbol{C}_1(X) \oplus T^j\boldsymbol{C}_2(X)$ can be partitioned into proper subsets for assignment to ultrasonic transceiver sets. For example, on referring Table I, with $0 \le i \le n_1\text{-}1$ and $0 \le j \le n_2\text{-}1$, we see the possible code vectors $T^i\boldsymbol{C}_1(X)$ and $T^j\boldsymbol{C}_2(X)$ in Tables I(a) and I(b), and the modulo-2 combined composite M-sequence codes in Table I(c). With respect to Table I(c), transceiver #1 can be allocated with composite codes $(T^i\boldsymbol{C}_1 \oplus T^0\boldsymbol{C}_2)$, transceiver #2 with composite codes $(T^i\boldsymbol{C}_1 \oplus T^2\boldsymbol{C}_2)$, and transceiver #3 with composite codes $(T^i\boldsymbol{C}_1 \oplus T^5\boldsymbol{C}_2)$. Alternative transceiver codes assignment can also be adopted.

 ISBN: 978-1-61208-422-0

TABLE I. (a). $T^iC_1(X)$ CODE SEQUENCES; (b). $T^jC_2(X)$ CODE SEQUENCES; (c). COMPOSITE CODE SEQUENCES $C^{(i,j)}(X) = T^iC_1(X) \oplus T^jC_2(X)$.

(a)

C_1	110	110	110	110	110	110	110
TC_1	011	011	011	011	011	011	011
T^2C_1	101	101	101	101	101	101	101

(b)

C_2	1110010	1110010	1110010
TC_2	0111001	0111001	0111001
T^2C_2	1011100	1011100	1011100
T^3C_2	0101110	0101110	0101110
T^4C_2	0010111	0010111	0010111
T^5C_2	1001011	1001011	1001011
T^6C_2	1100101	1100101	1100101

(c)

$C_1 \oplus C_2$	001111101010011000100
$TC_1 \oplus C_2$	100010000111110101001
$T^2C_1 \oplus C_2$	010100110001000011111
$C_1 \oplus TC_2$	101010011000100001111
$TC_1 \oplus TC_2$	000111110101001100010
$T^2C_1 \oplus TC_2$	110001000011111010100
$C_1 \oplus T^2C_2$	011000100001111101010
$TC_1 \oplus T^2C_2$	110101001100010000111
$T^2C_1 \oplus T^2C_2$	000011111010100110001
$C_1 \oplus T^3C_2$	100001111010100110000
$TC_1 \oplus T^3C_2$	001100010000111110101
$T^2C_1 \oplus T^3C_2$	111010100110001000011
$C_1 \oplus T^4C_2$	111101010011000100001
$TC_1 \oplus T^4C_2$	010000111110101001100
$T^2C_1 \oplus T^4C_2$	100110001000011111010
$C_1 \oplus T^5C_2$	010011000100001111101
$TC_1 \oplus T^5C_2$	111110101001100010000
$T^2C_1 \oplus T^5C_2$	001000011111010100110
$C_1 \oplus T^6C_2$	000100001111101010011
$TC_1 \oplus T^6C_2$	101001100010000111110
$T^2C_1 \oplus T^6C_2$	011111010100110001000

III. ROBOT POSITIONING SYSTEM ARCHITECTURE

In views of correlation acquisition, the longer the code acquisition time, the longer the path estimation distance, and the worse the robot positioning accuracy. According to the proposed composite coding scheme, we devise a parallel composite codes acquisition scheme to implement the indoor robot positioning system; the position of a target can be captured from the distances between the ultrasonic transceivers and a receiver of a target.

Figure 3 depicts a conceptual schematic of the proposed indoor robot positioning system. In the transmitter, the ultrasonic transceivers are installed at the corners and connected to central controller. Three composite PN codes structured from relatively prime-length M-sequence codes are assigned to different transceivers. These composite codes are modulated with ultrasonic carrier wave to generate transmission signals.

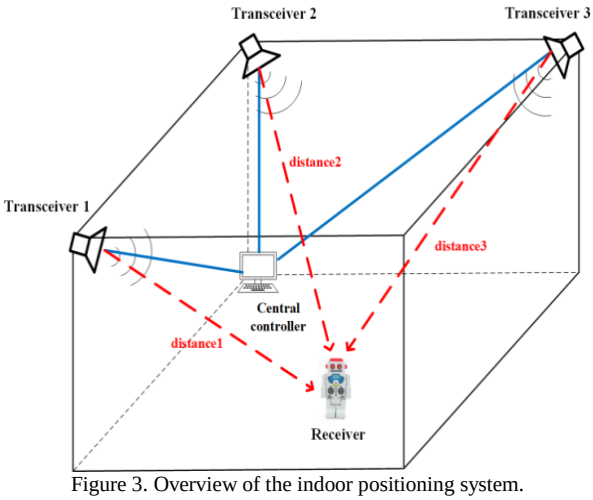


Figure 3. Overview of the indoor positioning system.

The reason we use ultrasound instead of higher frequency modulation signals is for easy visualizing robot codes acquisition under our limited PN code lengths in the transceivers. Take as comparative numerical figures for the high and low modulation rates. With 21-chip lengths per code frame and suppose 5-frames time is needed to confirm code acquisition. On using RF chips rate of 2000-kHz ($2x10^6$ chips/sec), the estimated object distance will be $21x5/2x10^6 = 5x10^{-5}$m. This figure is hardly distinguishable on the robot distance to the transceiver. But on using ultrasonic chips rate of 20-Hz (20 chips/sec), the same code length and acquisition frame will yield an estimated object distance of 21x5/20 = 5m. This figure is something acceptable. In practice, acquisition chips period length in mobile positioning can reach up to 2^{13}-1 = 8191 chips per frame to yield a distinguishable object distance.

In the robot receiver, in order to calculate the distance from each transceiver, the robot needs to separate the incoming signals from different transceivers. The robot bears the same ultrasonic carrier wave and composite PN codes as those of the transceiver signals, which are called the replica signals. On correlating received code signals with local replica signals, the robot can separate correlation peaks for the matched transceiver code from correlation nulls for the unmatched ones. This procedure for correlation detection of code signals is called code acquisition.

The robot positioning block chart for acquiring signal codes and estimating their flight time is as shown in Figure 4. In coding/modulating part of Figure 4(a), every transceiver performs ultrasonic signal modulation with assigned signature code, and emits this ultrasonic signal continuously. Once the signal is received by the robot, the

 ISBN: 978-1-61208-422-0

receiver turns the signal from analog to digital, and demodulates it into a corresponding code sequence. Since the receiver needs to identify the intended sequence code among all received signals, the demodulated code sequence is connected to three parallel correlators to calculate with each assigned code. Figure 4(b) depicts conceptual block chart on correlation decoding processes in the robot side. The output correlation passes through a peak detector to estimate the time of flight from transceiver to the robot. The robot then evaluates its current position by measuring time difference of arrival (TDOA) among the three transceiver paths.

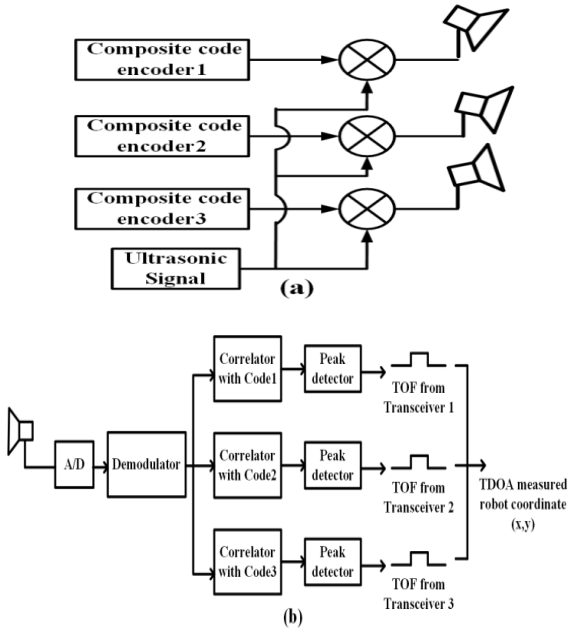


Figure 4. Block chart for robot positioning system; (a). Signals coding in transceivers; (b). Correlation acquisitions in the robot.

With regard to the block diagram of Figure 4 for robot positioning system, we will give detailed descriptions on codes correlation acquisition/detection, acquisition time difference and time error, and relative distance/locations determination of robot object in the following subsections.

A. *Code acquisition with correlation detection*

After the transmission signals transmit to the receiver, the received signals have a transmission time delay so that the received signals are not synchronous with the replica signal. Therefore, how to capture the relative signal and ignore the interference is the main course. We provide solutions to overcome the interference and improve the accuracy in the following sections.

For code acquisition, we note that the correlation characterizations of the assigned composite codes are related with their code weights. If code vectors $T^iC_1(X)$ and $T^jC_2(X)$ have the respective code weights w_1 and w_2, then composite code $C^{(i,j)}(X) = T^iC_1(X) \oplus T^jC_2(X)$ possesses the following code weights

$$\mathrm{W}\left(C^{(i,j)}\right) = w_1(n_2 - w_2) + w_2(n_1 - w_1) \quad (4)$$

$$= \begin{cases} \frac{n_1(n_2+1)}{2}, & if\ w_1 = 0,\ \ w_2 = (n_2+1)/2. \\ \frac{n_2(n_1+1)}{2}, & if\ w_1 = (n_1+1)/2,\ \ w_2 = 0. \\ \frac{(n_1n_2-1)}{2}, & if\ w_1 = (n_1+1)/2,\ \ w_2 = (n_2+1)/2. \end{cases} \quad (5)$$

Here, we have taken advantage that a binary (n_l=2^{ml}-1, k_l = m_l) M-sequence code has all of its n_l nonzero code vectors the same code weight of (n_l+1)/2 = 2^{ml-1}. Corresponding to the weight distribution of (5), the periodic correlation between composite codes $C_u^{(i_u,j_u)}$ and $C_v^{(i_v,j_v)}$ can be derived to be

$$\theta_{u,v} = \begin{cases} \left(\frac{n_1n_2-1}{2}\right), if\ u = v \\ \left(\frac{n_1n_2-n_2-2}{4}\right), \left(\frac{n_1n_2-n_1-2}{4}\right), \left(\frac{n_1n_2-1}{4}\right), if\ u \neq v \end{cases} \quad (6)$$

From the above correlations distribution of (6), we see that correlations between reference transceiver and interfering transceivers can be separated by correlation operation to track the desired transceiver sequences.

When the robot receives the incoming ultrasonic signals the receiver demodulates the received signals and performs correlation operations between the demodulated PN sequences and the replica signals stored in the correlators. The correlation computation will offer codes acquisition information on the periodic correlation peaks, and the receiver calculates the delay time and the codes acquisition error accordingly. Figure 5 illustrates the possible correlation spectra for composite signal sequence been acquired with M-sequence component codes $\boldsymbol{C}_1(X)$=$\boldsymbol{U}(X)$ and $\boldsymbol{C}_2(X)$=$\boldsymbol{V}(X)$. Here we take as example the composite signal sequence of period length $n_1.n_2$=21 and component M-sequences $\boldsymbol{C}_1(X)$ and $\boldsymbol{C}_2(X)$ of period lengths n_1=3 and n_2=7. These code sequences will respectively be assigned to the corner transceivers and the central robot.

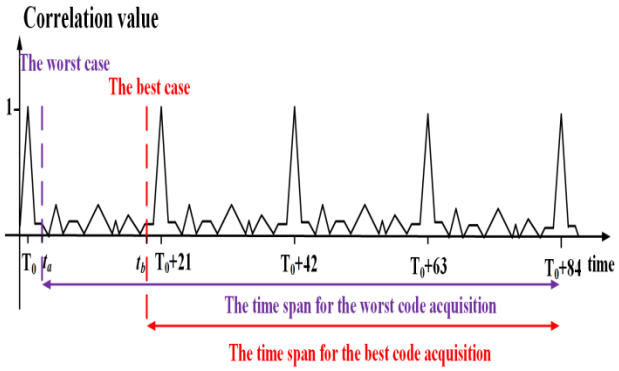


Figure 5. Correlation spectra to illustrate time span for the worst and

the best code acquisitions.

The robot receiver carries out correlation operations between received PN sequence and local M-sequence $\boldsymbol{C}_1(X)$; and in parallel, the robot receiver carries out correlations between received PN sequence and local M-sequence $\boldsymbol{C}_2(X)$. If no correlation peak occurred in either operation, the corresponding local code will advance one chip to another code pattern to continue its correlation computation. On the other hand, if an individual low-level correlation peak occurred, the local code will advance one period cycle of the current code pattern to continue its correlation computation. This process will continue until a high-level common correlation peak is obtained. The local codes in the robot will keep continuing their code sequences advancement for upto three to five common period lengths (3~5 frames) to confirm the final code acquisition status.

B. Code acquisition time difference and time error

In the indoor positioning environment, transmission signals interfere with each other. Every transceiver has different code sequence. In order to capture the relative code sequence to confirm the received signal which the transceiver transmitted. We determine the time between two adjacent peaks interval whether the corresponding transmission signal. If the time interval satisfies the correlation characteristic of the relative signal, we use this signal to calculate the time of flight and the time error of code acquisition.

On advancing code chips for correlation peaks, instant time t_a in Figure 5 depicts the worst case of code acquisition in which twenty-chip advancement is needed to reach an initial common correlation peak. Apparently, the common correlation peak is not captured very soon by the peak detector so that the time span of code acquisition takes much longer. On the other hand, instant time t_b in Figure 5 depicts the best case of code acquisition in which only one-chip advancement will reach the initial common correlation peak. In this case, the common correlation peak can be quickly captured by the peak detector so that the span time of code acquisition is significantly shortened.

A flow chart for the above correlation acquisition processes is as shown in Figure 6. Received summed sequence of period 21 is parallelly correlated with local PN sequences of period lengths 3 and 7. If not getting a correlation peak in either correlator, implies unmatched local and received sequence codes, one-chip relative shift is advanced and correlation magnitude is again calculated. But if a correlation peak is obtained in either correlator, matched local and received sequence codes is assumed, 3- or 7-chips relative shift is advanced and correlation magnitude is again calculated. The processes continue until a highest correlation peak occurs at the common code frame length of 21 chips. From then on, further confirmation of high peak correlation over 3~5 code frames will assure a complete code acquisition.

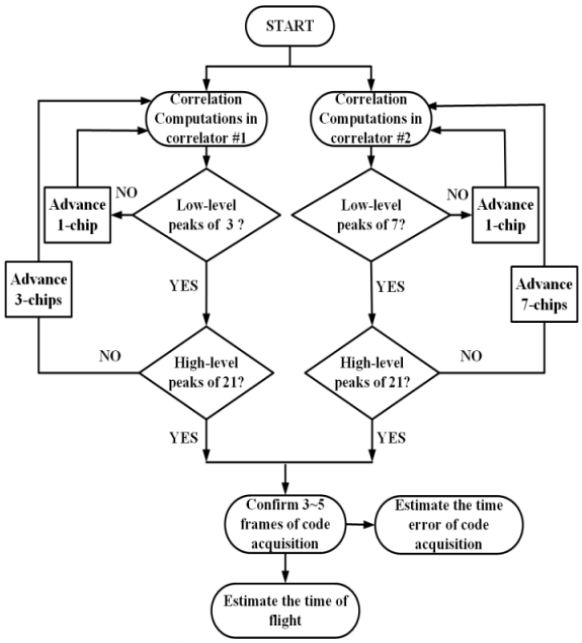


Figure 6. Flow chart for composite PN codes' correlation acquisitions.

In actual, codes correlation acquisition are not necessarily in the best or the worst acquired time, but may fall in the possible instant between the two extremes. Therefore, we will use probability distribution to analyze the possible cases to estimate the average of the time spent. We will further take the average of the time span to improve the indoor positioning accuracy. In the proposed positioning system, the main impact factor is the correlation characteristics of the composite PN sequence codes because the codes periods are not in symmetrical lengths. The time of flight between indoor transmitters and robot receiver are measured from the time instant the ultrasonic signals been emitted from corner transceivers to the time instant the acquisition peak detection been confirmed at the robot.

The time error of code acquisition is caused by the system that spends time searching for the relative sequence. Because the sequence is not sure to fall in the best or the worst acquired time so we need to calculate this acquired time to estimate the time error of code acquisition. We assume the time error of code acquisition about 2~3 code cycle lengths. Therefore, every signal has a different time error of code acquisition because of different code length. For example, the M-sequence code length n_1=31 spends about 62~93 bits shift time to capture the signal and the composite code n_2=3 × 7 spends about 42~63 bits shift time to capture the signal. In our positioning method, we use these values of the time error of code acquisition to enhance our indoor positioning accuracy.

 ISBN: 978-1-61208-422-0

C. *Determine the position of the robot receiver*

In order to obtain the position of the robot, the range measurement is acquired by TDOA of the ultrasonic signals of the transceivers. The TDOA will be biased by the time error of code acquisition that can degrade the positioning estimate. Therefore, the time error of code acquisition needs included in the calculation. Figure 7 is taken to illustrate three transceivers functions in expression (7) below to locate the position of an object receiver.

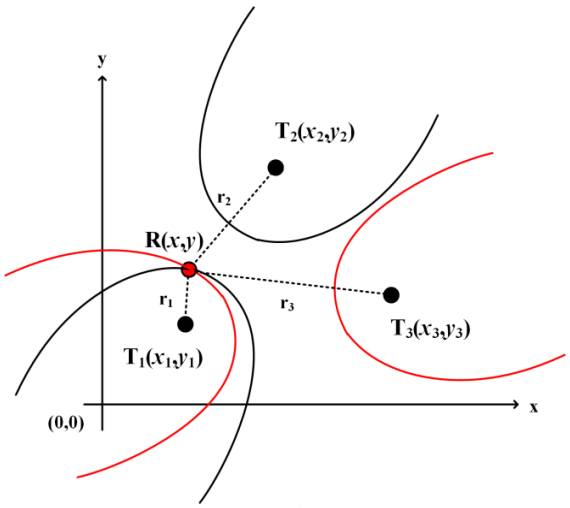


Figure 7. Overviews of TDOA evaluation.

Assume that, r_1, r_2 and r_3 are the estimated time of flight obtained from the number of frame peak between local and received code sequences. Once we get these estimates, we subtract them to each other to obtain ΔT_{12}, ΔT_{13}, and ΔT_{23}. We then substitute these flight time differences into (7) to solve the TDOA:

$$d_{ij} = c * (\Delta T_{ij} + e_{ij}), \quad \text{where } i \neq j$$
$$= \sqrt{(x_i - x)^2 + (y_i - y)^2} - \sqrt{(x_j - x)^2 + (y_j - y)^2} \quad (7)$$

where (x, y), (x_i, y_i) and (x_j, y_j) are respectively the real and the estimated position of robot receiver to the i-th and j-th transceiver, $i,j = 1, 2, 3$; d_{ij} are the value of TDOA; c is the ultrasonic wave speed; ΔT_{ij} is time difference measured by code acquisitions; and e_{ij} is the value of the time error of code acquisition to subtract with each other. The equations above represent hyperbolas, and their intersection gives the estimated positioning of the receiver.

The solution of equation derived a wide variety of algorithms because finding the solution is not easy as the equations are nonlinear. There are many methods to solve equations in this research problem. One direct solving method is Taylor-series method (TSA). It is the simplifying method, but the solutions are not divergent or converge toward a local suboptimal result if the unsuitable initial point was given. Fang Algorithm (Fang), Chan Algorithm (Chan), and Total Least Squares Algorithm (LTS) provides better performance than TSA. In order to optimize location result, evolution computing techniques will be applied to this working. Evolution computing techniques are based on principles of biological evolution, such as natural selection and genetic inheritance, such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Artificial Bee Colony algorithm (ABC).

IV. NUMERICAL SIMULATION RESULTS

In order to analyze the positioning accuracy between traditional M-sequence code and composite M-code which we proposed, we use M-sequence code length n_1=31 and composite code length n_2=3×7 to simulate. The robot was placed in the coordinate (x=3m, y=3m and z=0m). The three transceivers are located at three corners (x=0m, y=0m and z=5m) (x=0m, y=10m and z=5m) and (x=8m, y=0m and z=5m), in the numerical simulation, we assume the robot on the ground so we don't consider the z-axis. Transceiver #1 is assigned with composite codes ($T^0C_1 \oplus C_2$) = (1, 1, 0, ⋯) ⊕ (1, 1, 1, 0, 0, 1, 0, ⋯), transceiver #2 assigned with signature codes ($T^1C_1 \oplus C_2$) = (0, 1, 1, ⋯) ⊕ (1, 1, 1, 0, 0, 1, 0, ⋯), and transceiver #3 assigned with signature codes ($T^2C_1 \oplus C_2$) = (1, 0, 1, ⋯) ⊕ (1, 1, 1, 0, 0, 1, 0, ⋯).

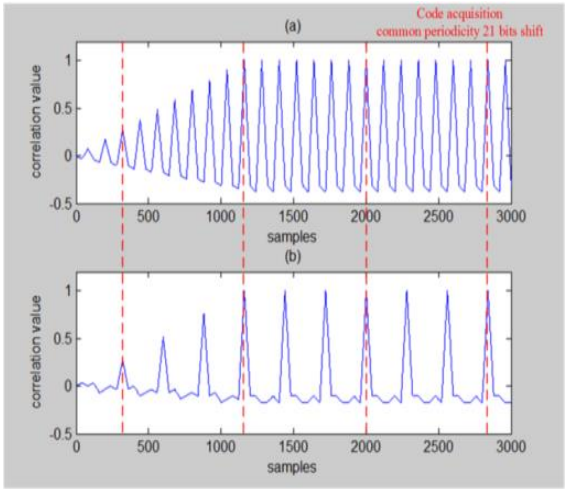


Figure 8. Correlations of composite signal with component signature codes; (a). with M-sequence code C_1 = (1,1,0); (b). with M-sequence code C_2 = (1,1,1,0,0,1,0).

Figure 8 illustrates the receiver performs the correlation operation with transceivers #1. The periodicity of Figure 8(a) is 3 bits shift and the periodicity of Figure 8(b) is 7 bits shift, therefore their common periodicity is 21 bits shift. From Figure 8, the first red line is the first common peak of code acquisition, which will change with the first incoming frame because the order of frame may not be C_1 = (1, 1, 0). Therefore, the receiver will search for

 ISBN: 978-1-61208-422-0

next common peak by common periodicity 21 bits shift to capture their relative signals. These time spent are the error time of code acquisition. Once the receiver captures the peaks completely, the receiver estimates its error time of code acquisition of signals and time of flight.

Figure 9 illustrates correlation operations of robot receiver with transceivers #1-#3 on the relative composite codes. The cycle of peak is 21 bits shift as shown in Figure 9. The different numbers of cycle is caused by the different distance between the receiver and transceivers. We use these data to estimate the numbers of cycle, and calculate the time of flight and the error of code acquisition.

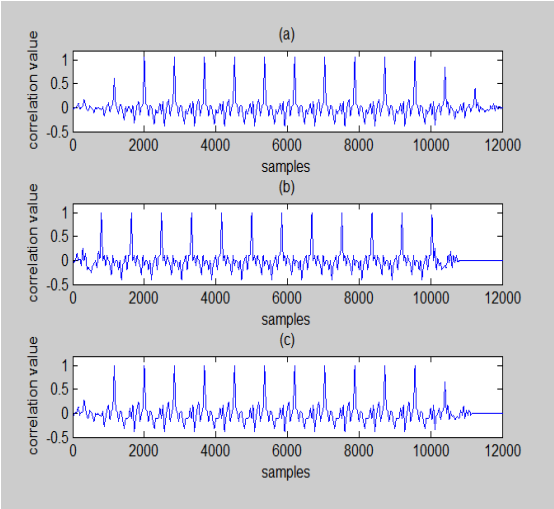


Figure 9. Robot correlation operations with transceivers; (a). with transceiver #1 on composite codes ($T^0C_1 \oplus C_2$); (b). with transceiver #2 on codes ($T^1C_1 \oplus C_2$); (c). with transceiver #3 on codes ($T^2C_1 \oplus C_2$).

Table II shows estimates of time of flight and robot distance to the three transceivers. Through calculating the number of frame peak between local code and received sequence, we estimate the time of flight and then the distance between transceivers and the robot. The estimated position errors are not over 10-cm, thus achieves our goal on enhancing indoor robot positioning accuracy.

TABLE II. ESTIMATE OF THE DISTANCE BETWEEN TRANSCEIVERS AND THE ROBOT.

	Estimate time difference	Estimate range	Real range
With transceiver#1	0.0179 sec	6.1934 m	6.1644 m
With transceiver#2	0.0232 sec	8.0272 m	7.9372 m
With transceiver#3	0.0221 sec	7.6466 m	7.6811 m

For comparison, we assign the central robot and the corner transceivers with comparative M-sequence codes of period length n=31. Figure 10 gives possible correlation spectra on correlating received code sequence from transceiver #1 and local signature codes in the robot. Note that, with such conventional PN code sequences, two-levels of correlation magnitude is possible and single branch correlator circuit can be taken for codes acquisition operation.

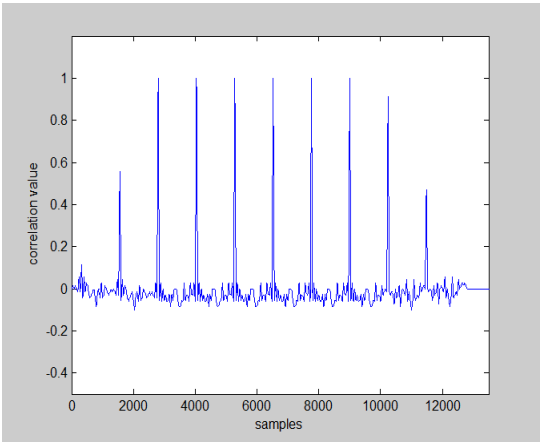


Figure 10. Correlation operation with transceiver #1 on 31 bits M-sequence code.

On comparing Figures 9 and 10, we find that the number of frame cycles of 31 bits M-sequence code is less than those using composite codes to do correlation operations at the same distance. Because of a large code length cycle, the receiver spends much shift time to capture the signal so that the error time of code acquisition is more than using composite code to do correlation operation. The mean of the distance errors is about 20 cm, so using composite code is more precise than using M-sequence code.

V. CONCLUSIONS

We have proposed a composite code acquisition to implement indoor ultrasonic robot positioning based on DSSS system. Each transceiver is modulated the ultrasonic signal with a 3×7 bits composite code, which has a particular auto-correlation and cross-correlation in a cycle. By using code acquisition the robot receiver detects the arrival time of codes and the error time of code acquisition, and the robot will use these information to determine its absolute location.

By comparing our solution with traditional M-sequence code, we find that composite codes behave more advantages. First, the code length is more flexible, it is not limited by 2^m-1. Second, other robot users are difficult to acquire the location of the designated robot because the code combination is more complex. Third, under the same location distance, the positioning accuracy and the code acquisition time-error are more precise with composite coding than the conventional M-sequence coding. This is

 ISBN: 978-1-61208-422-0

because correlation acquisition takes more cycles than that using pure M-sequence codes.

REFERENCES

[1] P. C. Chen, "A non-line-of-sight error mitigation algorithm in location estimation," *Proc. IEEE Wireless Communications and Networking Conference*, Sept. 1999, vol. 1, pp. 316-320.

[2] S. Al. Jazzar, J. Caffery, and H. R. You, "A scattering model based approach to NLOS mitigation in TOA location systems," *Proc. IEEE Vehicular Technology Conference*, May, 2002, vol. 2, pp. 861-865.

[3] Aatique, Muhammad, "Evaluation of TDOA techniques for position location in CDMA systems," Diss., Virginia Polytechnic Institute and State University, 1997.

[4] K. Sanhae, L. Jungwoo, Y. Myungsik, and S. Yoan, "An improved TDoA-based tracking algorithm in mobile-WiMAX systems," *20th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept. 13-16, 2009, pp. 561-565.

[5] J. K. Werner and B. Markus, "Using Pseudo-random Codes for Mobile Robot Sonar Sensing," *the* 3*rd IFAC Symposium on Intelligent Autonomous Vehicles* (*IAV*'98), Madrid, Spain, March 25-27, 1998, pp. 231-236.

[6] A. Heale and L. Kleeman, "A Sonar Sensing with Random Double Pulse Coding," *Australian Conference on Robotics and Automation*, Melbourne, Australia, Aug. 30 – Sept. 1, 2000, pp. 81-86.

[7] J. M. illadangos, J. Urena, M. Mazo, A. Hernandez, F. Alvarez, J. J. Garcia, C. De Marziani, and D. Alonso, "Improvement of ultrasonic beacon-based local position system using multi-access techniques." *IEEE International Symposium on Intelligent Signal Processing* (*WISP* 2005), Algarve, Portugal, Sept. 1-3, 2005, pp.352-357.

[8] M. C. Pérez et al., "Ultrasonic beacon-based Local Positioning System using Loosely Synchronous codes," *IEEE International Symposium on Intelligent Signal Processing*, (*WISP* 2007), October, 2007, vol. 1, no. 6, pp. 3-5.

[9] M. C. Pérez, J. Urena, A. Hernández, C. De Marziani, A. Jimenez, J. M. Villadangos, F. Alvarez, "Ultrasonic signal processing using configurable computing," *15th Triennial World Congress of the International Federation of Automatic Control* (*IFAC* 2002), Barcelona, 2002.

[10] Md. Alamgir Hossain, Md. Shariful Islam, and Md. Sadek Ali, "Performance Analysis of Barker Code based on their Correlation Property in Multiuser Environment," *International Journal of Information Sciences and Techniques* (*IJIST*), January, 2012, vol.2, no.1, pp. 27-39.

[11] J. F. Huang, K. S. Chen, Y. C. Lin, and C. Y. Li, "Reconfiguring Waveguide-Gratings-based M-Signature Codecs to Enhance OCDMA Network Confidentiality," *Optics Communications*, February, 2014 vol. 313C, pp. 223-230.

[12] G. De Angelis, G. Baruffa, and S. Cacopardi, "Parallel PN code acquisition for wireless positioning in CDMA handsets," *the* 5*th advanced satellite multimedia systems conference* (*asma*) *and the* 11*th signal processing for space communications workshop* (*spsc*), September 13-15, 2010, pp.343-348.

 ISBN: 978-1-61208-422-0

A Method for Custom Movement Generation in Wireless Mobile Network Simulation

Hawra Alseef, John DeDourek, Przemyslaw Pochec
Faculty of Computer Science
University of New Brunswick
Fredericton, Canada
e-mail: {hawra.alseef, dedourek, pochec}@unb.ca

***Abstract*—Random motion is often used in evaluating performance of Mobile Ad hoc Networks (MANETs). Mobility pattern of nodes significantly affects performance of MANETs. In the simulation of large mobile networks, automated movement generators are used, such as the *setdest* utility in ns2. In this paper, we investigate the modifications to the standard *setdest* generator that specifies the motions along straight-line paths. We propose and implement a new method for movement generation for the ns2 simulator that specifies the node movements along curved paths generated using simple fractals. The new generator was successfully tested with the ns2 simulator. The results show that, in the random way point motion, only the node speed significantly affects MANET performance, and not the shape of the individual path segments taken by a node.**

Keywords-movement generator; network simulation; ns2; fractal path; setdest utility; MANET

I. INTRODUCTION

A Mobile Ad hoc Network (MANET) is a set of mobile devices that cooperate with each other by exchanging messages and forwarding data [1][2]. Mobile devices are linked together through wireless connections without infrastructure and can change locations and reconfigure network connections. During the lifetime of the network, nodes are free to move around within the network and node mobility plays a very important role in mobile ad hoc network performance. Mobility of mobile nodes significantly affects the performance of a MANET [2].

Simulation is a commonly used evaluation tool for mobile networks. It allows for modelling existing networks as we as future networks. Using simulation, different network configurations working under different traffic load conditions and using different routing protocols can be quickly and easily modelled and evaluated. For mobile networks, if no other design constraints are present, random motion of the nodes is usually used. Using a common and specific random motion model allows to create the base condition for comparison between different network evaluations.

ns2 is an open source simulator well suited for modelling wired and wireless networks [3]. It includes a motion scenario generator *setdest* designed to automatically generate random motion paths for a large number of nodes. This tool generates a random motion path for each node by selecting a random destination for the node and then moving the node towards this destination along a straight line. Once this destination is reached by the node, a new destination is randomly selected and, after an optional pause time, the node starts moving again to the new destination.

In this paper, we propose to use the random motion generated by the ns2 *setdest* utility [3] to create a new trajectory for the mobile nodes. The waypoints are kept the same but the path followed by the node between two waypoints is no longer defined by one straight line segment, and is replaced by a fractal curve composed of a number of shorter line segments.

In Section 2, we review different random movement models commonly used in simulation. Section 3 introduces properties of the Koch fractal. Section 4 describes the new movement generator based on generation of the node movement along a fractal path. Section 5 presents a study of two mobile networks: one with the conventional random movement and the other with the fractal movement. Conclusion is presented in Section 6.

II. STATE OF THE ART

Any model of a MANET requires a mobility model specifying the movement pattern of the nodes [4]. The most realistic models are trace driven but cannot be always applied because of their *a posteriori* nature. On the other hand, the synthetic models [5] are not trace driven but instead rely on assumptions about the node movement mode. Among these are the random (random-based) models where the nodes move randomly and without restrictions and where the destination and the speed are chosen randomly.

There are many different types of random mobility models that are used in MANETs. The main ones are the Random Walk, the Random Waypoint, and the Random Direction. The Random Walk model [5] mimics the Brownian motion of particles found in nature. Each node travels in a straight direction for a specified time interval before randomly changing the speed and the direction, and then continuing for another time interval. In the Random Waypoint model [6], each node selects a destination within the simulation area and then follows a straight path to it; once the destination is reached the node may pause and then select a new destination (waypoint). In the Random Direction model [7], instead of selecting a random destination, the node selects a random direction and then moves along this

 ISBN: 978-1-61208-422-0

direction until it reaches the simulation area boundary where, possibly after a pause, it selects a new direction for the next move.

The ns2 *setdest* utility generates the node movements following the Random Waypoint algorithm [3]. In the Random Waypoint Movement (RWP), each node moves from its randomly selected initial starting position towards the randomly selected at a randomly selected speed. Once at the target destination the node may pause for a randomly selected time, and then start the next random move. This process will be repeated until the end of the simulation by the ns2. One notable aspect of the RWP movement is that the nodes following this pattern tend to concentrate in the center region of the deployment area [8][9].

III. THE KOCH FRACTAL

We propose to use fractals for the movement generation for mobile network simulation based on the RWP model. Instead of moving the nodes along a straight line between the waypoints, the nodes are moved along a fractal path. We selected the Koch fractal because, like a line segment, it has a defined starting and ending points, and because of the simplicity of its generating algorithm [10][11].

A. Construction of the Koch curve

The construction of the Koch starts with a straight line that is then converted to the Koch fractal curve, Figure 1.

Figure 1. Step 1.

Figure 2. Step 2.

Figure 3. Step 3.

This process is then repeated for each of the 4 segments generated at the first iteration, leading to the curve shown in Figure 3. These steps can be applied repeatedly and eventually result in a complex shape. When the Koch curve generating algorithm is applied to an equilateral triangle it results in a closed curve called the Koch snowflake [11].

B. Properties of the Koch snowflake

Number of Sides (n): for each iteration, every segment of the curve from the previous iteration will be converted to four segments in the following iteration. Since we begin with three sides, the formula for the number of sides in the Koch curve is:

$$n = 3 * 4^a \quad (1)$$

where a indicates the number of iterations. For iterations 0, 1, 2, and 3, the numbers of sides are 3, 12, 48, and 192 respectively.

Length of Sides (L): In every iteration, the length of a side is 1/3 the length of a side from the previous iteration. If we begin with an equilateral triangle with side length x, then the length of a side in iteration a is:

$$L = x * 3^{-a} \quad (2)$$

For iterations 0 to 3, length = x, x/3, x/9, and x/27.

Perimeter (p): The key features of the Koch curve lies in having the same length of all sides in each iteration, this leads to a perimeter, which is simply the number of sides multiplied by the length of a side:

$$p = n * L \quad (3)$$

For the snowflake, from the previous formulas, we get:

$$p = (3 * 4^a) * (x * 3^{-a}) \quad (4)$$

In the same manner, for the first 4 iterations (0 to 3) the perimeter is 3x, 4x, 16x/3, and 64x/9. We notice that, the perimeter increases by 4/3 times for each iteration, so we can rewrite the formula as

$$p = (4/3)^a * 3x \quad (5)$$

IV. CUSTOM MOVEMENT GENERATION WITH FRACTALS

The main objective of this research is to implement a new method for movement generation in MANET simulation in ns2. Indeed, the standard way for movement generation is to use the *setdest* utility that generates a set of setdest commands that are then "executed" in the ns2 simulator. setdest commands generate a movement along a straight line between the current location and the designated destination point. This research aims at providing a new tool for modifying the simulation environment by modeling motion in wireless network simulations, specifically for generating movement files for ns2 simulation that specify the motion along curved (fractal) paths. Typically, defining the node movements needs to be done ahead of the ns2 simulation. In general a curved path can be approximated by a series of short line segments, which determine the final shape of the curve. Therefore, a Java program was implemented that reads the

 ISBN: 978-1-61208-422-0

movement file with random movements generated, for example, by the *setdest* utility. Then, as each movement in the movement file is specified by a separate setdest command, we will replace each one of these setdest commands, each specifying a movement along a straight line, with a series of setdest commands specifying the movement along a curved path (fractal). Once the new movement file is generated the ns2 simulation can proceed in a standard way.

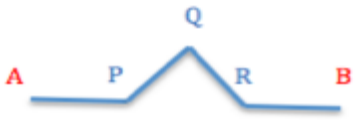


Figure 4. The result of fractal transformation of a line segment AB.

Let's consider the original setdest command for the direct movement from A to B (Figure 4):

$ns_ at T "$node_(#) setdest X_B Y_B S"

where T indicates the starting time at which the node starts moving towards the destination X_B,Y_B at the specified speed S. While splitting the initial path (line segment AB) into four segments (AP, PQ, QR and RB) and defining the destination of each of the four moves is a simple geometry, the other setdest command parameters require careful consideration. More precisely, the need of updating the time and speed in the setdest commands arises when applying the fractal transformation. In order to make the fractal movements arrive at the final destination (point B) at the same time that the original straight movement would have arrived, we need to do the following modifications:

$ns_ at T_P "$node_(#) setdest X_P Y_P S_{new} "
$ns_ at T_Q "$node_(#) setdest X_Q Y_Q S_{new} "
$ns_ at T_R "$node_(#) setdest X_R Y_R S_{new} "
$ns_ at T_B "$node_(#) setdest X_B Y_B S_{new} "

The four (fractal) movements should proceed sequentially, each having a starting time after the previous movement ends. To calculate the precise time of each move and the new speed we need to determine the new speed and the new starting time for each of the four new setdest commands. First, we need to calculate the time the node would take to travel from A to B at speed S along the original straight line path AB:

$$t_{AB} = \text{sqrt}((X_B - X_A)^2+(Y_B-Y_A)^2)/S \quad (6)$$

then the start times for each move are calculated as:

$$\begin{aligned} T_P &= T \\ T_Q &= T_P + t_{AB}/4 \\ T_R &= T_Q + t_{AB}/4 \\ T_B &= T_R + t_{AB}/4 \end{aligned} \quad (7)$$

and the new speed, due to the distance travelled increased by 1/3, is:

$$S_{new} = 4 * S / 3 \quad (8)$$

(Obviously, when the intermediate point Q would fall outside the predefined simulation region then the corresponding segment of the fractal path cannot be generated, as shown on Figures 5 and 6.)

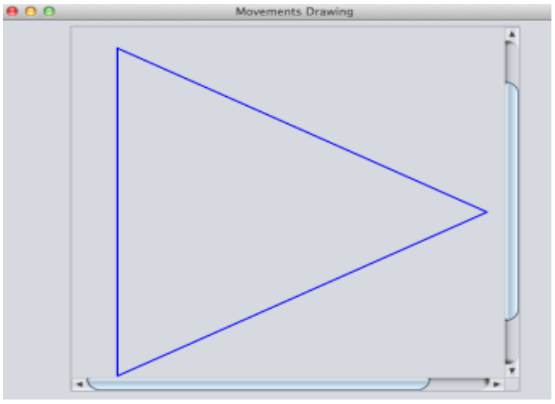


Figure 5. Trace of sample simple node movement.

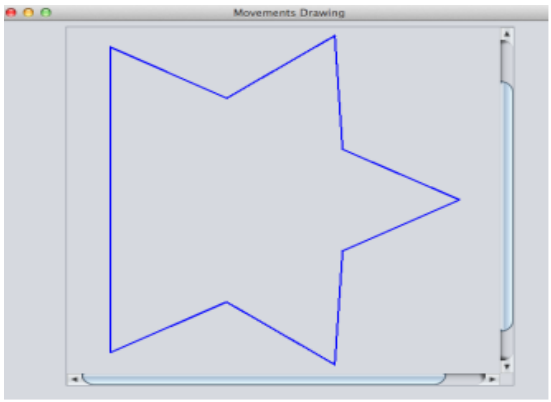


Figure 6. Fractal movement generated from Figure 5.

For example, consider the following movement statement taken from a movement file generated by the setdest utility:

$ns_ at 0.400000 "$node_(0) setdest 100.00000000 400.0000000 1000.02523710421"

This line specifies that at time 0.40000s, node0 starts to move from the starting point (100,100) towards the destination (100,400) at a speed of 1000m/s (this can be one single random movement in a straight line). This single command in the movement file is then replaced by four new commands generating the movement along the path corresponding to the shape of the Koch fractal (one iteration of the Koch fractal generation algorithm). The four movements are listed below.

 ISBN: 978-1-61208-422-0

$ns_ at 0.40000000 "$node_(0) setdest 100.00000 200.00000 1333.366982805613"

$ns_ at 0.474998107265 "$node_(0) setdest 13.3974596215 250.0000 1333.36698280"

$ns_ at 0.549996214530 "$node_(0) setdest 100.00000 300.00000 1333.366982805613"

$ns_ at 0.624994321795 "$node_(0) setdest 100.00000 400.00000 1333.366982805613"

V. EVALUATION OF MANET PERFORMANCE UNDER FRACTAL MOVEMENT

We evaluated the performance of a sample MANET under different motion generation conditions. A MANET with the number of nodes ranging from 5 to 80 was simulated over the area of 800 by 800 meters with two fixed communicating stations at (100,500) and (700, 500). Constant Bit Rate (CBR) traffic was generated over the User Datagram Protocol (UDP) and routed with Ad hoc On-Demand Distance Vector (AODV) (Table I).

TABLE I. SIMULATION PARAMETERS

Parameters	
Simulator	NS-2.33
Channel Type	Channel / Wireless Channel
Network Interface Type	Phy/WirelessPhy
Mac Type	Mac/802.11
Radio-Propagation Type	Propagation/Two-ray ground
Interface Queue Type	Queue/Drop Tail
Link Layer Type	LL
Antenna	Antenna/Omni Antenna
Maximum Packet in ifq	50
Area (n * n)	800 x 800
Source Type	(UDP) CBR
Simulation Time	100s
Routing Protocol	AODV

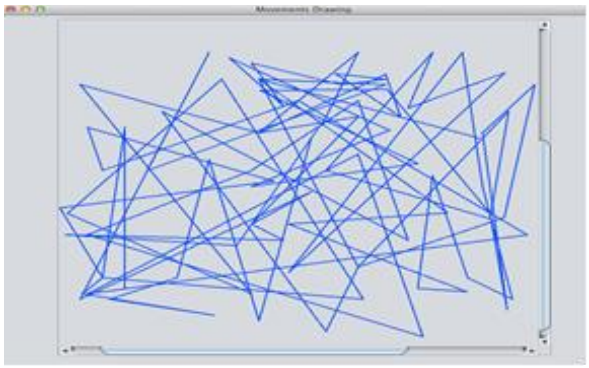


Figure 7. Trace of complex random movement.

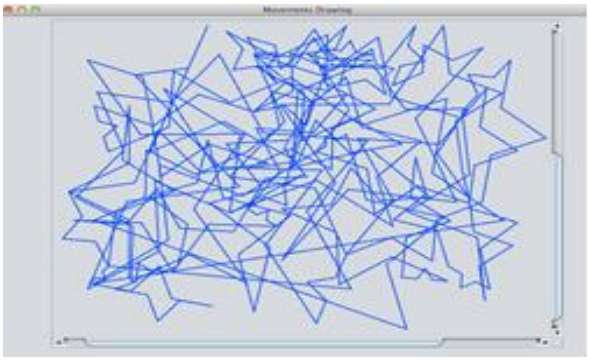


Figure 8. Fractal movement generated from Figure 7.

Standard RWP movement was generated with setdest and then the standard movement was converted to the fractal movement using one step of Koch generating algorithm, as shown in Figures 7 and 8. Two scenarios were investigated: (i) low speed (10m/s), and (ii) high speed (30m/s).

Figure 9 illustrates the difference in the number of packets received at the destination when using the original movement and the new fractal movement at low speed. It shows that most of the time the packet delivery for the fractal movement is higher than the original linear movement. Although the speed of the fractal path is higher than the original (because of the increased path length along the fractal curve between the original waypoints), we observed a higher number of packets delivered at the destination for the fractal movement at speed of 13m/s. However, applying the t-test for the comparison of two paired means representing the packets received in the linear motion and the fractal motion experiments with 25 nodes gives 8%, which indicates that the observed difference is not statistically significant. Also, comparing the average packet delivery across all node densities does not show a significant difference (t-test value 49%).

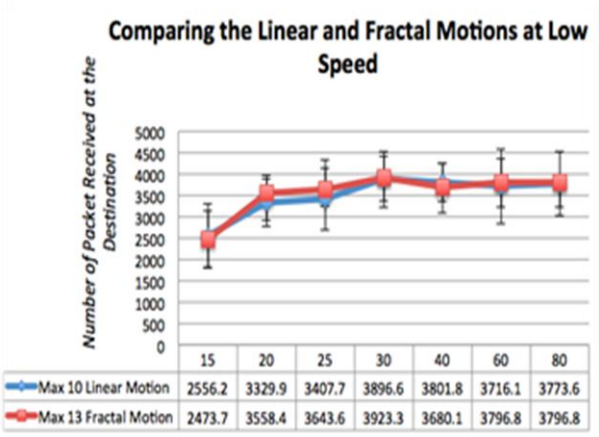


Figure 9. Throughput comparison at low speed

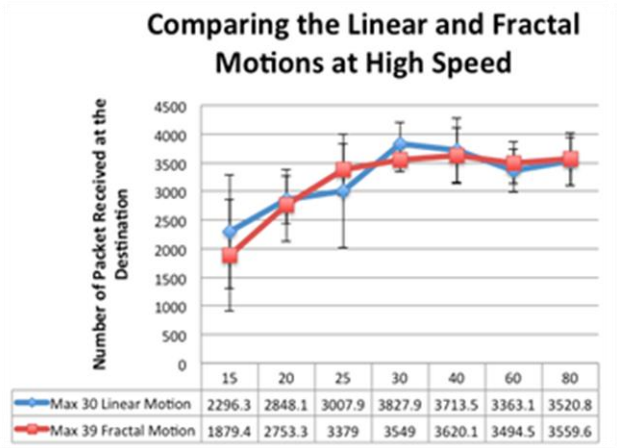


Figure 10. Throughput comparison at high speed.

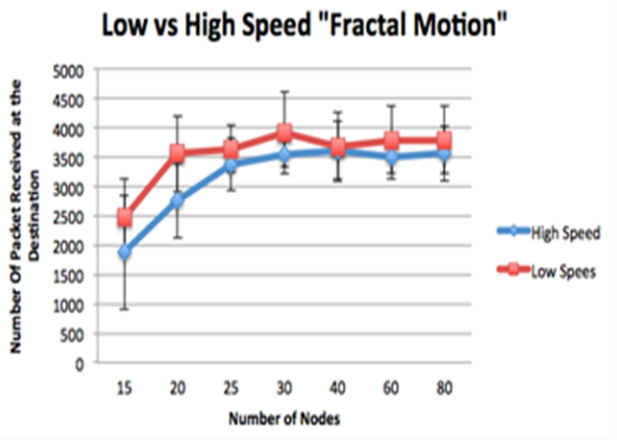


Figure 12. Throughput comparison for fractal motion.

Figure 10 shows the packet delivery for linear and fractal motions at high speed. This time we observe a lower packet delivery for fractal motion recorded in most of the experiments. One possible explanation of lower performance with fractal motion is that the increase in movement speed of 10m/s, from 30 to 40, results in more frequent link disconnections and consequently lower packet delivery. Applying the t-test for the comparison of two paired means representing the packets received in the linear motion and the fractal motion experiments with 25 nodes gives 32%, which indicates that the observed difference is not statistically significant. Also, comparing the average packet delivery across all node densities does not show a significant difference (t-test value 60%).

Figure 11 illustrates the advantage of using lower speed in a network with linear motion. It shows that the packet delivery is consistently higher at low speed for almost all node densities. Applying the t-test for the comparison of two means representing the packets received in the linear motion and the fractal motion experiments with 20 nodes gives 4%, which indicates that the observed difference is statistically significant. The average packet delivery for all node densities is 3176 at high speed and 3497 at low speed,

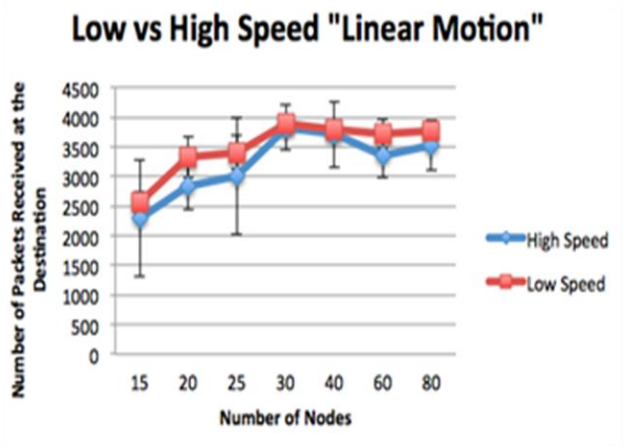


Figure 11. Throughput comparison for linear motion

and this difference in performance is statistically significant (t-test value 0.076%).

Figure 12 illustrates the advantage of using lower speed in a network with the fractal motion. The packet delivery is consistently higher at low speed for all node densities. Applying the t-test for the comparison of two means representing the packets received in the linear motion and the fractal motion experiments with 20 nodes gives 1%, which indicates that the observed difference is statistically significant. The average packet delivery for all node densities is 3186 at high speed and 3553 at low speed, and this difference in performance is statistically significant (t-test value 1.7%).

V. CONCLUSION AND FUTURE WORK

In this paper, we presented a tool for transforming linear movements into fractal movements based on the Koch curve. The new tool reads a standard ns2 movement file, decodes each movement, and replaces it with a series of new movements forming a fractal curve, and then outputs a new movement file. The newly generated movement file satisfies the ns2 specifications and can be used in the ns2 simulator. Both standard movement files generated with *setdest* and new movement files generated with the new fractal tool were used in simulating a MANET with varying number of nodes (i.e. with different node densities). We compared the MANET performance in terms of packet delivery under two different motion scenarios and at different speeds. We observed marginally higher performance of MANET with fractal motion at low movement speeds. However, the statistical tests show that the difference observed in our limited experiments is not significant. We observed that the packet delivery is lower at higher speeds for both motion types, and after the application of the t-test for the difference of the means, we concluded that the observed lower packet delivery at higher speed is statistically significant.

From our results, we conclude that only the node speed significantly affects the MANET performance, and not the shape of the path taken by a node.

 ISBN: 978-1-61208-422-0

The work presented in this paper demonstrated a new experimental approach for investigating performance of mobile networks: applying transformations to the node movement paths. The future work on transforming the node movement paths will include using more than on iteration of the generating function of the Koch fractal, calibrating the node speed when it starts moving on the new curved path and testing if the new path generators reduce the tendency observed in the RWP model of clustering the nodes towards the center of the experimental area.

ACKNOWLEDGMENT

This work is sponsored and funded by the Ministry of Higher Education of Saudi Arabia through the Saudi Arabian Cultural Bureau in Canada.

REFERENCES

[1] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic (Eds.), Mobile Ad Hoc Networking, New York, Wiley-IEEE Press, 2001.

[2] F. Bei and A. Helmy, A survey of mobility models in wireless Ad hoc Networks, University of California, USA, 2004.

[3] H. Ekram and T. Issariyakul, Introduction to Network Simulator NS2, Springer, 2009.

[4] N. Aschenbruck, E. G. Padilla, and P. Martini, "A survey on mobility models for performance analysis in tactical mobile networks", Journal of Telecommunications and Information Technology, vol. 2, 2008, pp. 54-61.

[5] T. Camp, J. Boleng, and V. Davies, "A Survey of Mobility Models for Ad Hoc Network Research", Wireless Communication and Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications , vol. 2, no. 5, 2002, pp. 483-502.

[6] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols", Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking (Mobicom98), ACM, October 1998, pp. 85-97.

[7] E. M. Royer, P. M. Melliar-Smith, and L. E. Moser, "An Analysis of the Optimum Node Density for Ad hoc Mobile Networks", Proceedings of the IEEE International Conference on Communications(ICC), Helsinki, Finland, June 2001, pp. 857-861.

[8] C. Bettstetter, "Mobility Modeling in Wireless Networks: Categorization, Smooth Movement, and Border Effects", ACM Mobile Computing and Communications Review, vol. 5, no. 3, July 2001, pp. 55-67.

[9] R. Alghamdi, Movement Generator for Mobile Network Simulation. Master's Report, Faculty of Computer Science, University of New Brunswick, Fredericton, Canada, 2012.

[10] G. Edgar, Measure, Topology, and Fractal Geometry, 2nd Ed., Springer 2008

[11] Koch's Snowflake, online, http://en.wikipedia.org/wiki/Koch_snowflake, retrieved: May 2015.

 ISBN: 978-1-61208-422-0

A Longitudinal Study on Flipping the Classroom in a College Level English Course: Performance of undergraduate students at the Lebanese International University (LIU)

Fawziya Tarhini, Dina Shouman, Anwar Kawtharani
School of Education
Lebanese International University
Beirut / Nabatieh, Lebanon
{Email: fawziah.tarhini, dina.shouman,
anwar.kawtharani}@liu.edu.lb

Hanadi Saleh
College of Education
Florida Atlantic University
Boca Raton, FL, USA
Email: wsaleh@fau.edu

Hassan Khachfe
Center for Quality Assurance, Institutional Assessment, and Scientific Research
Lebanese International University (LIU)
Beirut, Lebanon
Email: hassan.khachfe@liu.edu.lb

Abstract— **Technological trends have given rise to the development of flipped learning classrooms. An inverted (or flipped) classroom is a specific type of blended learning that uses technology to separate lectures from learning activities. While lectures will take place outside the classroom, learning activities and concepts practice will take place inside the classroom. This paper compares the learning achievement of two flipped English (ENGL 201) classes with traditionally taught classes at the same university. A longitudinal panel research study is used to investigate the learning environment of these two classrooms. At the onset of the pilot, students were less satisfied with the new orientation, but, they – later - became more open to the new learning method. These findings are discussed in terms of how they contribute to the stability and connectedness of classroom learning communities.**

Keywords-Blended Learning; Flipped Classrooms; Educational Technology; Inverted Classrooms

I. INTRODUCTION

The Ministry of Education and Higher Education in Lebanon (MOEHE) regulates the education institutes through a regional education system. The education system in Lebanon is centralized, and this regulation is not direct. The education system is managed through regional education bureaus leaving the integration of technology into the education system up to the public or private institution. Given that the incorporation of technology is an integral part of the curriculum, instruction, and assessment, technology addition is inevitable.

Furthermore, the idea of integrating diverse approaches to teaching and learning has been one of the major goals of education. With the advancement of technological means come new trends to advance the goal of education, the goal of personalizing instructions, and leading up to the introduction of flipped classrooms. The university system at the chosen educational institution (LIU, Lebanese International University) piloted the flipped learning – a part of blended learning - for the first time in fall of 2014/2015.

In subsequent sections, we will address areas of research related to student reaction of integrating flipped classroom techniques with English 201 students at LIU (Lebanese International University). In the first section, research related to the impact of general goals of education will be addressed. In the second section, the goal of personalizing instruction with students will be discussed. Finally, the last section focuses on research-based strategies that have been implemented in flipped classrooms.

A. The Goal of Education

According to Paul O'Keefe [6], individuals have certain approaches to goal pursuit. Two factors - cognitive and affective – influence how people pursue goals. Individuals need to be interested to maximize motivation and self-regulation, and the structure of achievement context influences motivation for attainment of goals as per O' Keefe and Garcia [7]. In short O'keefe's research examines motivational processes involved in the pursuit of goals.

B. The Goal of Personalizing Instruction

According to Keefe and Jenkins [4], personalization of instruction and learning is the effort on the part of an educational institution to take into account individual student characteristics and needs and flexible instructional practices in organizing the learning environment. Teachers committed to personalizing instruction help their students develop personal learning plans, assist in diagnosing their cognitive strengths and weaknesses and other style characteristics, help adapt the learning environment and instruction to learner needs and interests, and mentor authentic and reflective learning experiences for their students. Personalization is broader in scope, more

 ISBN: 978-1-61208-422-0

systematic in organization, and more authentic in its goals and strategies.

C. Asynchronous Learning

In an asynchronous learning environment, students are able to actively participate in their own learning, giving them the opportunity to interact with their peers, provide peer feedback, and reflect on the status of their personal learning goals and outcomes [2]. In many learning environments, there are learning activities and expectations that require students to create, synthesize, explain, and apply the content or skills being taught [3]. Asynchronous technologies support learning and allow more time for student reflection, collaboration, and student-to-student interactions [1].

D. Educational Technology

Technology has the potential to not only offer access to resources for learning in a superficial sense, but also to provide increased affordances for autonomous learning. Opportunities for interaction, situated learning, and support for learning outside formal contexts, have greatly improved because of technology. These affordances are not yet always capitalized on. However, they offer the opportunity to support the learning process [8].

E. Flipped Classroom Approach

The flipped classroom approach has been used for years in some disciplines, particularly within the humanities. Barbara Walvoord and Virginia Johnson Anderson encouraged the implementation of this method in their book *Effective Grading* [9]. They suggest a model in which students gain *first-exposure learning* prior to class and focus on the *processing* part of learning (synthesizing, analyzing, problem-solving, etc.) in class. To ensure that students do the preparation necessary for productive class time, Walvoord and Anderson propose an assignment-based model in which students produce work (writing, problems, etc.) prior to class. The students receive productive feedback through the processing activities that occur during class, reducing the need for the instructor to provide extensive written feedback on the students' work.

According to Aronson and Intern [5], the Flipped Learning model of instruction is gaining consideration among instructors and professors at the college and university levels. In this model, some or most of the direct instruction is conveyed outside the group learning space using multiple modes of delivery. Class time is optimized for students to engage in hands-on learning, collaborate with their peers and evaluate their progress rather than traditional direct instruction delivery. Instructors can offer one-on-one support, guidance and motivation. This enables a shift from an instructor-centered classroom to a student centered learning environment. Flipped Learning is principally well-suited to higher education settings for a variety of reasons. The in-class discussion and enrichment activities allowed by moving content delivery outside of class time provide opportunities for students to develop vital skills needed in the 21st century, including critical thinking, creativity, communications, and collaboration. The model can also be especially useful in large lecture courses where student engagement and interaction is usually minimal. When students receive the lecture outside of class, they can use time in class with their peers more effectively by breaking up into smaller discussion groups or engage in other in-class activities. Instructors also make more effective use of their time by reviewing content that students actually need help with and guiding student discussions. The Flipped Learning model also permits for differentiated learning in classes of all sizes, since students can preview the lecture content at their own pace and ask questions on their own time [2].

II. METHODOLOGY

A. Research Purpose and Questions

The purpose of this study was to examine the effectiveness of flipping a college course designed to teach composition and research skills when compared to the traditional classroom orientation. The research question was: Does flipping the classroom impact learning effectiveness? The following point was considered: how much did students perceive what they learnt? The motivation was based on the expectation that flipping the classroom will have a positive impact on the students' achievements, quantitatively and qualitatively.

B. Design of the Study

This research used a short - term longitudinal panel study in which students of two flipped classrooms at Nabatieh and Beirut campuses were tracked over a period of four months and, eventually, compared to their traditional counterparts. Inferential statistics were used to determine the significance of any differences found between and among groups. Observational data supplemented assessment data to help better interpret and understand the results. The setting for the study was an English (ENGL 201) course, a general University requirement that teaches the principles of "Composition and Research Skills", taught by two instructors at the Lebanese International University Nabatieh and Beirut Campuses. English 201 is a course where students study different types of discourse: narrative, description, argumentative, cause – effect, process, etc… and evaluate them based on the four bases of good writing: unity, coherence, sentence skills, and support. Students also learn how to conduct research and write essays.

C. Description of the Instructional Approaches

This section describes the approaches tested in this study: 1) traditional instruction in the form of classroom lectures and large group based instruction. 2) a flipped classroom enabling technology using videos, pdfs, web links, narrated power points, word docs with classroom

 ISBN: 978-1-61208-422-0

support. Students of both sections used the same material: all used a textbook, syllabus, assignments and exams.

D. Form of Instruction

In the two chosen sections, instruction following the traditional approach took place inside the classroom. On the other hand, in the flipped approach, the instruction was provided using a specialized platform called "Coursesites" [10] where assorted materials were uploaded and announcements to students were made regularly keeping them updated of upcoming online sessions. Instructors had to send invitations to their students so that they can log onto the platform.

E. Assessment

The primary reasons for evaluating students are those reasons which are an essential part of a teacher's main responsibility, helping students improve in knowledge and skills, feelings and attitudes, and hence, helping students learn. In both treatments, students were assessed formatively based on a criterion - referenced benchmark. Students sat for two types of exams: a midterm and a final one. The validity of the exams was taken into consideration.

F. Class Process

In both orientations, students were asked to read the textbook materials before attempting to complete the homework. In the traditional classroom orientation, the teacher provided instruction in the classroom. Students were expected to be active learners: that is - come to class prepared, participate in class discussion, and ask questions. In the flipped orientation, students did not attend class, but tackled the starting material and completed some assessments online. In the flipped classroom orientation, in addition to reading the textbook material, students were able to watch narrated power points, videos, demonstrating how to accomplish the task. Pdf's, weblinks, and other attachements were also provided to aid the stds in understating the starting material. There was also a discussion forum that they can use in order to raise any question on mind. Answers to the forum can either be from fellow students or from the teacher her/himself. In addition, students were – through the forum able to read other perspectives and carefully consider a final response.

G. Participants

Subjects in this research were undergraduate students taking the composition and research skills English course (ENGL 201) during the Fall semester of 2015 at the Lebanese International University (LIU), Nabatieh and Beirut campuses. This course was divided into a fourteen weeks term. Participants of both orientations were taught simultaneously. The participants took the course as a required part of their program. 36 participants were involved per each traditional and flipped class in Nabatieh and 35 participants per each section in Beirut.

H. Data Collection and Analysis

To answer the primary question regarding the students' achievement, Pearson correlation was measured to identify any statistically significant correlation specially that the class sample was parametric. Final average grades were calculated based on the unified grading system of the four sections: Participation %10, Research Presentation %20, Essay Quizzes %20, Midterm Exam %25, and Final Exam %25. The tests were identical and designed to assess students' achievement of the learning outcomes of the course. The tests were formative and summative, online and in-class, assessments the instructors gave students throughout the learning process.

I. Limitations

Course redesign took large time investment and effort since extra resources and material other than the traditional ones had to be found and used on part of the instructors. Student Commitment – to a certain extent - rendered the reliability of the instructions for students with high absence rate. Email was the main form to send announcements, where % 5 of the students still faced difficulty in checking their emails regularly, but, eventually, doing so before attending the sessions in question. There was difficulty in reserving the computer lab – in one of the campuses - to accommodate students in varied and convenient timings. Students felt threatened – at the beginning - by the use of technology, but – later - got accommodated with it. Some students had difficulty accessing internet in and outside the university, but had access to IT support though phone or mail.

III. RESULTS

Quantitative results were used to identify patterns and explore the research findings.

A. Students' Achievements

Given the fact that all students submitted their assignments, Pearson correlation was used to examine the effect of the instruction time (i.e., the regular classroom; the flipped classroom) on test scores. The main effects of instruction type in both Campuses: Nabatieh & Beirut were significant as shown in the tables below:

TABLE I. FINAL AVERAGE GRADES: NABATIEH

CORRELATIONS

		NAB201E	NAB201A
NAB201E	Pearson Correlation	1	.097
	Sig. (2-tailed)		.573
	N	36	36
NAB201A	Pearson Correlation	.097	1
	Sig. (2-tailed)	.573	
	N	36	36

 ISBN: 978-1-61208-422-0

Table I shows that there is positive correlation between the traditional section (NAB201E) and the flipped one (NAB201A); the correlation as shown is (Sig .573)

TABLE II. FINAL AVERAGE GRADES: BEIRUT
CORRELATIONS

		Beirut201PF	Beirut201GT
Beirut201PF	Pearson Correlation	1	.310
	Sig. (2-tailed)		.070
	N	35	35
Beirut201GT	Pearson Correlation	.310	1
	Sig. (2-tailed)	.070	
	N	35	35

Table II shows that there is also positive correlation between the traditional section (Beirut201PF) and the flipped one (Beirut201GT); the correlation is (Sig .070).

IV. DISCUSSION AND CONCLUSIONS

This paper explored how technology can be used to teach composition and research skills and what impact flipped leaning might have for students taking a college course in comparison to the traditional instruction. As shown in the tables above, both traditional sections in Nabatieh and Beirut showed positive correlations with their flipped counterparts. Hence, it demonstrated how technology integrated into class instruction has a similar impact as the traditional method, which was manifested by Pearson Correlation - .5 at Nabatieh section and .07 at Beirut's. Compared to the traditional treatment, the flipped classroom approach provided an effective method for delivering the class; it allowed students to learn course content at their own pace in which they had access to the online session either on campus computer labs (in which the broadband was made sure to be feasible to upload or download material) or at the convenience of their own home, allowing them to make a better use of their time and become more teacher independent. At first, students used to rely on the teacher to deliver the required material in class in which they were quasi passive recipients, whereas and after the very first two sessions, students were participating, commenting, giving their opinion on the class material used on Coursesites. It is of utmost importance to mention that to meet the students' needs, they were oriented on creating student accounts, surfing the site, downloading and uploading material, using discussion forums, & receiving online announcements, in the university lab before the official onset of the online sessions.

We expected the flipped approach to be better than the traditional one, but, this result was surprising. We found no statistical difference between the novice and the traditional ones when it comes to the students' final averages achievements, and that is – of course – quantitatively. Qualitatively, we are in the process of collecting students' feedback using a questionnaire "Students Satisfaction Survey Form – SSSF) which will allow us to have a more thorough view of the qualitative difference between the flipped and the traditional approaches from the students' perspective.

The evidence suggests that the flipped approach is at least as effective as the traditional one for delivering this class and somewhat more scalable which is impressive given the limitations stated above. Despite the fact that few students faced difficulty in logging onto their emails, most of them were able to accomplish the intended tasks in the modules. Moreover, students received vis-a-vis support and had access to their instructors during their office hours to fill in any gaps. It is a newly founded approach for students and they still managed to fair out the same as a regularly delivered class. Students had ownership of their own learning resulting in experiencing independent educational experiences. As the semester progressed, it was evident via the teachers' observation in class that students were able to tackle various directions and topics independently. Another point to elucidate, that prior to administering the final exam, a mock-final exam was uploaded in one of the modules which required higher order thinking skills, and the student fared very well.

We also expected that the blended approach might be inferior to the regular one in achieving the required outcomes and objectives especially with the limitations we had. But, the results of this study seem to suggest that this is not the case and students became well acquainted with technology and common educational means used globally.

In summary, our findings suggest that the flipped approach was as instructional as the traditional method given all the stated limitations.

V. FUTURE RESEARCH

While our study provides evidence that the flipped classroom is as efficient as the non - flipped, we should be cautious in generalizing the findings beyond the scope of the context. Future studies will show if one method has efficacy over the other, when it comes to students' performance and motivation. Future research in this area is required. Due to the positive impact of the flipped course in Fall 2014/2015, the flipped program will continue as of Fall 2015/2016. It is recommended to proceed with the study tackling both quantitative and qualitative sides in accordance with LIU (Lebanese International University) students' profiles. Future recommendations include widening the scope of flipped learning into telecommunication reaching students in other countries.

 ISBN: 978-1-61208-422-0

REFERENCES

[1] C. Bonk and K. Zhang, "Introducing the R2D2 model: Online learning for the diverse learners of this world," Distance Education, 2006, 27(2), pp. 249-264. doi:10.1080/01587910600789670 e-Learners.com 2012, April 2. Synchronous vs. asynchronous classes [blog]. [retrieved: January, 2015] http://www.elearners.com/online-education-resources/online-learning/synchronous-vs-asynchronous-classes/

[2] Er. E. Özden, and A. Arifoglu, "A blended e-learning environment: A model proposition for integration of asynchronous and synchronous e-learning," International Journal of Learning, 2009, 16(2), pp. 449-460.

[3] J. Harris, P. Mishra, and M. Koehler, "Teachers' technological pedagogical content knowledge and learning activity types: Curriculum-based technology integration reframed. Journal of Research on Technology in Education" 41(4), 2009, pp. 393-416, [retrieved: January, 2015] http://learnonline.canberra.edu.au/file.php/5963/TPACK_UC/pdf/harris_mishra_koehler_jrte

[4] J. W. Keefe and J. M Jenkins, "Personalized instruction: Changing classroom practice. Larchmont,"2000, NY: Eye on Education.

[5] N. Aronson, P. Intern, and M. K. Arfstrom, "Ph.D. Flipped Learning Network & Kenneth Tam, " Summer 2013, Pearson.

[6] P. A O'Keefe, C. S. Dweck, and G. Walton, "Implicit theories of interest and motivation. Accepted presentation at the Society of Personality and Social Psychology, Austin, TX," 2014.

[7] P. A O'Keefe, & Linnenbrink- L. Garcia (in press), "The role of interest in optimizing performance and self-regulation," Journal of Experimental Social Psychology".

[8] H. Reinders, "Big brother is helping you. Supporting self-access language learning with a student monitoring system," 2007, 35(1), 93–111.

[9] B.E. Walvoord, and VJ Anderson, "Effective grading: A tool for learning and assessment," 1998, San Francisco: Jossey-Bass.

[10] https://www.coursesites.com/webapps/Bb-sites-course-creation-BBLEARN/pages/index.html [retrieved January 2015]

 ISBN: 978-1-61208-422-0

Media Form Effect on Children's Attention

Chun-Chun Wei Min-Yuan Ma
Department of Industrial Design
National Cheng Kung University
Tainan, Taiwan
email:p38991123@mail.ncku.edu.tw email:mamy@mail.ncku.edu.tw

***Abstract*—Picture books are frequently used as teaching materials and have thus been published in diverse media forms. By exploiting visual, auditory, and tactile modalities, children can construct a reading context in which their attention is enhanced and they can learn effectively. In this study, picture books of various media forms are used as stimuli for examining the effect of various picture book types on children's attention. On the basis of the results of this study, the audiobook group demonstrates the highest attention performance. In the gender difference, a significant difference in attention response exists between the boys and girls. In particular, the attention performance of the boys is significantly higher than that of the girls in the e-book group.**

Keywords-children's attention; picture books; media forms; electroencephalography (EEG).

I. INTRODUCTION

Picture books are one of the most commonly used teaching materials from preschool to elementary school [1]. Because picture books are designed for children, they cater to children's characteristics and have features that are relevant to language development [2], cognitive engagement [3], artistic thinking, and entertainment [4]. The senses are used to communicate with the external world during early reading development in children. Various types of sensory stimulation help children construct a reading context, which enables children to learn independently [5]. Picture books are published in various media forms; thus, information contained in books can be expressed by text, images, symbols, sounds, and texture. Accordingly, children who easily receive sensory signals can directly communicate with the external world in visual, auditory, and tactile manners [6]. For example, pop-up books featuring three-dimensional (3D) images and operable interfaces, audiobooks with voice dubbing and sound, toy books comprising various tactile materials, and e-books containing multimedia animation can appeal to the senses of children and enhance their comprehension of the content.

This study determines that effectively using picture books that provide sensory stimulation to instruct children can enhance their attention and learning performance. Numerous studies on picture book instruction have revealed that picture books can increase children's interest in learning.

According to previous electroencephalography (EEG) studies [7][8], attention is a type of brain function, and based on the activation of the cerebral frontal lobe, α and β waves can be measured to investigate the degree of attention in learning. The EEG device, is used in the EEG experiment in this study for measuring children's attention. We anticipate that the results of this study can be applied to children's education and picture books design, help parents, teachers and designers select and design picture books for children, and improve the attention of children.

The research report comprises the following five sections: (1) introduction, (2) media forms of picture books, (3) experimental design, (4) quantitative analysis, and (5) conclusion.

II. MEDIA FORMS OF PICTURE BOOKS

In the current mature publication market, because of reader sensory requirements and improved multimedia technology, the content of picture books is presented through text, illustration, symbols, and multimedia (e.g., various book materials, sounds, music, and animation), and diverse sensory stimuli and reading methods are used to attract children's attention and interest. Picture books are not limited to a two-dimensional (2D) design and are integrated media that provide various sensory stimuli. According to previous studies, picture books can effectively help children develop their language, cognitive, and aesthetic appreciation abilities [2][3].

This study explores the selection of picture books and media types of picture books that suit children and attract their attention. We collect the various forms of picture book designs that are currently commercially available. These picture books are then classified into four types according to the sensory perception methods used when reading books, as determined by a focus group of three designers with more than 5 years of design experience. These four types are (a) conventional book (visual perception), which is the most common book format and involves page turning when reading; (b) pop-up book (visual and tactile perceptions), which transcends the limitations of conventional books by including interactive components to present 3D concepts, thereby providing an enjoyable and interactive reading experience; (c) audiobook (audial and visual perception), which offers a multimedia presentation of the traditional storybook format with the addition of speech feedback (e.g., CD and MP3 formats) so that children can elect to hear the story read to them; and (d) e-book (multisensory perception), which includes multimedia effects such as oral reading, written text, oral discourse, music, sound effects, and animations, thereby enriching the content of picture books.

 ISBN: 978-1-61208-422-0

III. EXPERIMENT DESIGN

An EEG experiment is conducted to investigate whether sensory stimulation provided by conventional books, pop-up books, audiobooks, and e-books affect children's attention performance.

A. Experimental Participants and Procedures

48 third-grade elementary school children aged 9 to 10 years are recruited as participants in this study. The participants are evenly divided into four groups (12 in each group). Each group comprises 6 boys and 6 girls. The participants are divided into four groups: the conventional book, pop-up book, audiobook, and e-book groups. The experimental procedure is as follows:

Step 1: The researcher explains the purpose and procedure of the experiment to a participant and advises the participant to relax.

Step 2: The researcher places an EEG cap on the head of the participant.

Step 3: The participant begins reading a picture book from the first page. During the EEG measurement, the participant continues to read until he or she reaches the end of the book. The reading duration is not fixed.

B. Experimental Tool

In this study, the EEG device, NeuroSky MindBand [7], is used for the EEG experiment. The sensors of the device are placed at the frontopolar (FP1) area of the forehead, which is the frontal lobe area of the attention network, to measure attention performance [8][9]. MindBand is lightweight, stable, easy to wear, and highly appropriate for children. Unlike a traditional EEG device, MindBand does not require wearing an electrode cap, applying gel, or washing hair after testing. For convenience, NeuroSky MindBand and related products are widely applied in EEG studies on education, psychology, and sports [10][11][12].

C. Electroencephalographic Data Collection

In the EEG experiment, we collect the EEG data of the participants while they read the books. Combined with the MindBand, a Universal Serial Bus (USB) brain-wave assessment and measurement system developed by Alchemy Technology [7] is used to collect data. The data is ranged from 0 to 100, enabling the collection of real-time information and analysis of EEG signals. Through the USB transmission hardware interface, the EEG charts and data on visual attention state are output [13].

D. Picture Book: Guess How Much I Love You

To prevent the story content and children's comprehension of the story content from affecting the attention performance of the children, a focus group method (based on the discussion of three language teachers at an elementary school) is employed for determining the use of the conventional book, pop-up book, audiobook, and e-book versions of Guess How Much I Love You [14].

IV. QUANTITATIVE ANALYSIS

A total of 48 children participate in this experiment and are assigned to four groups of 12 people each. The average score of all the participants for attention is 44.02. According to the average score of each group, the audiobook group exhibits the highest attention performance (52.92), followed by the e-book group (42.17), pop-up book group (41.42), and conventional book group (39.58).

A. Picture Book Media Types Independent Samples t-Test between Gender

An Independent Sample *t*-Test [15] is performed to examine whether a significant difference in attention response induced by picture books exists between the boys and girls. As shown in Table I, the number of boys is 24 and the average score is 52.08; the number of girls is 24 and the average score is 35.96. The *t* value is 2.60 (df = 46, $p < 0.01$), which is significant. Thus, a significant difference in average attention score exists between the boys and girls. The *t* value is a positive value (2.60), indicating that the attention performance of the boys are superior to that of the girls.

TABLE I. INDEPENDENT SAMPLES T-TEST OF GENDER

	t	df	Sig.
Gender	2.60	46	*0.01

Independent Sample *t*-Tests are performed to examine whether differences in attention response exist between the boys and girls in the various groups. As shown in Table II, a significant difference in attention response exists between the boys and girls in the e-book group ($p < 0.01$). The *t* value (3.36) is a positive value, indicating that for the e-book group, the boys demonstrate a significantly higher performance than the girls (the average score for boys is 57.83, and the average score for girls is 26.50). No significant difference in attention response exists between the boys and girls in the other groups.

TABLE II. T-TEST OF GENDER FOR VARIOUS MEDIA PICTURE BOOKS

Media Forms	t	df	Sig.
Conventional	0.92	10	0.38
Pop-up	0.30	10	0.77
Audiobook	1.32	10	0.22
e-book	3.36	10	*0.01

B. The interaction effect between media forms and gender： Two-Way ANOVA (MANOVA)

To examine whether an interaction effect between media forms of picture books and gender on attention response exists, a two-way ANOVA is performed. Table III summarizes the results of the two-way ANOVA. As shown in Table III, the F

 ISBN: 978-1-61208-422-0

value of the group effect is 0.93 ($p < .44$), indicating that no significant difference in attention response exists between the various groups.

TABLE III. THE INTERACTION EFFECT OF MEDIA FORMS AND GENDER

Source	SS	df	MS	F	Sig.
Media forms	1308.56	3	436.19	0.93	0.44
Gender	3120.19	1	3120.19	6.63	*0.01
Media forms * Gender	1149.40	3	383.13	0.82	0.49
Error	18812.83	40	470.32		
Corrected Total	24390.98	47			

The F value of the gender effect is 6.63 ($p < .01$), indicating that a significant difference in attention response exists between the boys and girls. In addition, the attention performance of the boys is superior to that of the girls. The results are consistent with those presented in Section IV-A. The F value of the interaction effect of gender and group is 0.815 ($p < .49$), indicating that no interaction effect between gender and group on attention response exists.

V. CONCLUSION

In this study, third-grade elementary school children are recruited as participants for examining the relationship between children's attention and picture book reading. MindBand is used in the experiment to collect EEG data from the participants reading various picture book types.

Considering the attention performance of the boys, the audiobook group exhibits the highest attention performance, followed by the e-book group, conventional book group, and pop-up book group. Considering the attention performance of the girls, the audiobook group achieves the highest attention performance, followed by the pop-up book group, conventional book group, and e-book group. Both the boys and girls in the audiobook group display excellent attention performance. The result shows that audiobooks should be selected to help children establish reading contexts using the acoustic stimuli provided by reciting story contents, thus improving their reading comprehension and concentration [16][17].

For the e-book group, the attention performance of the boys is significantly higher than that of the girls. The results suggest that visual and auditory stimulation provided by the e-book elicit excellent attention performance in the boys. The auditory stimulation provided by the audiobook elicits excellent attention performance in the girls, whereas dynamic visual stimulation provided by the e-book elicits the poorest attention performance in the girls. The result show multimedia effects of audiobook and e-book with immediate feedback might help male learners to increase attention and comprehension [18].

According to the results of this study, we suggest that the audiobooks can be adopted to instruct third-grade elementary school children in reading to enhance the attention performance of children. Audiobooks and e-books that provide auditory and visual stimulation can be used to instruct boys and audiobooks can be used to instruct girls. The results of this study can serve as a reference for parents, teachers and designer to select and design picture books and to enhance the attention of children.

ACKNOWLEDGMENTS

Funding of this research work is supported by the Ministry of Science and Technology, Taiwan, under grant numbers NSC 101-2410-H-006-072-MY2.

REFERENCES

[1] R. Barr, “Memory Constraints on Infant Learning from Picture Books, Television, and Touchscreens,” Child Development Perspectives, vol. 7, 2013, pp. 205-210.

[2] A. G. Bus, M. H. van IJzendoorn, and A. D. Pellegrini, “Joint book reading makes for success in learning to read: a meta-analysis on intergenerational transmission of literacy”, Review of Educational Research, vol. 65, 1995, pp. 1–21.

[3] I. Elia, M. van Den Heuvel-Panhuizen, and A. Georgiou, “The role of pictures in picture books on children’s cognitive engagement with mathematics,” European Early Childhood Education Research Journal, vol. 18, 2010, pp. 125–147.

[4] M. Y. Lin, Appreciation and application of picture books, Taipei: Psychological Publishing, 2000.

[5] J. Holt, “The underachieving school,” Sentient Publications, 2005.

[6] C. Briggs and D. Elkind, “Cognitive development in early readers,” Developmental Psychology, vol. 9, 1973, p. 279.

[7] Alchemy Technology: http://www.alchemytech.com.tw/.

[8] M. B. Sterman, D. A. Kaiser, C. A. Mann, B. Y. Suvenobu, D. C. Beyma, and J. R. Francis, “Application of quantitative EEG analysis to workload assessment in an advanced aircraft simulator.” In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 37, 1993, pp.118-121.

[9] A. Gevins, M. E. Smith, L. McEvoy, and D. Yu, “High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, ” type of processing, and practice, Cerebral Cortex, vol. 7, 1997, pp. 374-385.

[10] J. I. Ekandem, T. A. Davis, I. Alvarez, M. T. James, and J. E. Gilbert, “Evaluating the ergonomics of BCI devices for research and experimentation. Ergonomics”, vol. 55, 2012, pp. 592-598.

[11] D. V. Poltavski, D. Biberdorf, and T. V. Petros, “Accommodative response and cortical activity during sustained attention,” Vision research, vol. 63, 2012, pp. 1-8.

[12] C. S. Lin, Y. C. Lai, J. C. Lin, P. Y. Wu, and H. C. Chang, “A novel method for concentration evaluation of reading behaviors with electrical activity recorded on the scalp.” Computer methods and programs in biomedicine, vol. 114, 2014, pp. 164-171.

[13] K. Crowley, A. Sliney, I. Pitt, and D. Murphy, “Evaluating a brain-computer interface to categories human emotional response”, In Advanced Learning Technologies, 2010 IEEE 10th International Conference, July 2010, pp. 276-278.

[14] S.McBratney, Guess how Much I Love You, Candlewick Press, 2000.

[15] H. Timothy, and R. D'Agostino, “Robustness of the two independent samples t‐test when applied to ordinal scaled data.” Statistics in medicine, vol. 6, 1987, pp. 79-90.

 ISBN: 978-1-61208-422-0

[16] M. Furini, "Digital audiobook: from passive to active pursuit," Multimedia Tools and Applications, vol. 40, 2008, pp. 23-39.

[17] C. C. Lu, Y. Y. Chen, and C. W. Chen, "A Correlative study of CD-ROM picture books in classrooms and school children's formation of descriptive concepts," International Journal of Science and Mathematics Education, vol. 9, 2011, pp. 47-67.

[18] S. P. Lee, H. K. Su, and S. D. Lee, "Effects of computer-based immediate feedback on foreign language listening comprehension and test-associated anxiety," Perceptual and motor skills, vol. 114, 2012, pp. 995-10

 ISBN: 978-1-61208-422-0

Rehabilitation System in 3D Natural Scenes

Amin Safaei, Q. M. Jonathan Wu

Department of Electrical and Computer Engineering
University of Windsor
Windsor, ON, Canada
Email: {safaeia, jwu}@uwindsor.ca

***Abstract*—In this paper, we present a rehabilitation system for patients who suffer wrist injury. The idea to use computer vision for rehabilitation is not new; however, the method proposed in this paper differs significantly from previously proposed methods. We propose a 3D hand model evaluation method that can recognize soft and elaborate representations of hand motions. In practice, hand motion recognition in an unconstrained environment is a difficult task because of intra-class variation. It becomes more challenging when we lose depth data because of projection. However, the emergence of commercial depth sensors, such as Microsoft Kinect and SoftKinect, has overcome this issue. In previous work, we used the data of tip and joints, which was sufficient for simple motion; however, in complex motion, such as grabbing and rotation, it is not possible to track and estimate the depth of tips and joints. In this work, we modify the algorithm that is proposed by Rodriguez *et al* and Hadfield. Instead of using 2D data, we extend the method for 3D data, and for elevation information, Hidden Markov Model (HMM) is used.**

Keywords–Machine Vision; SoftKinetic; Motion Recognition; Video Processing; Rehabilitation.

I. INTRODUCTION

Rehabilitation has been emphasized recently in the field of computer vision. Rehabilitation is defined as a dynamic process that helps patients recover normal functional capability. To reach this milestone, it is necessary to monitor patient activity continuously and correct motions.

In this paper, we proposed a system to evaluate hand motions via depth data (3D) for rehabilitation. Given natural hand features and an uncontrolled environment, the proposed system classifies and differentiates unnatural slowness of motions.

To obtain the data, two main methods are available: sensor based and vision based. Sensor based methods use electromechanical or magnetic sensors to capture activity and then convert the motions to digital signals. The main drawbacks are that this method is expensive and requires calibration and a setup procedure. In contrast, vision based methods require only a camera. The advantages of this method are more natural, unencumbered, non-contact interactions, whereas the disadvantage of this method is requiring environments that are insensitive to lighting. Vision based methods can generally be categorized into appearance hand models (2D Mapping) and 3D hand models. Appearance hand models attempt to learn mapping from feature vectors, such as the positions of tips and joints of the hand. In contrast, 3D hand models rely on a 3D kinematic description of a hand. This method estimates depth data using either a stereo camera or 3D depth cameras (IR cameras). Using a stereo system for hand motion evaluation was tested in [1]; however, because of some issues, such as calibration and complexity, in this work, we used 3D depth cameras [2] [3] for ease of use and higher accuracy than that of the stereo system. To describe hand motions, we use the HMM, which is a state-based model that analyzes data and recognizes patterns. As a special case of human-computer interaction and rehabilitation, several constraints are imposed, which include complexities, background, variable lighting conditions, transforming gesture structures, real time implementation and dependency on user and device characteristics. The remainder of this paper is organized as follows. First, the proposed method for segmentation is discussed and then, the theory behind depth camera is described. Next, feature descriptors and classification method are explained and finally the proposed method for evaluation is discussed.

II. METHOD OF ANALYSIS

The proposed methodology is developed for single-hand motion recognition and evaluation. The procedure consists of segmentation, depth data, UV mapping, classification and recognition. In this section, the theory and implementation of the proposed method are described.

A. Segmentation

Segmentation is defined as the division of an image into different regions, each having different futures. In this work, we need to isolate the hand from the background. Our proposed methods for segmentation are K-means clustering and L*a*b* color space. The image is first transferred from RGB to L*a*b* color space, and then K-means clustering is used to isolate the hand from the background [4][5].

B. Depth Data

A 3D depth camera can provide 3D data using a low-cost CMOS pixel array with an active modulated light source. Its compact construction, ease of use and high accuracy and frame rate make it an attractive solution for a wide range of applications. A 3D depth camera operates by illuminating the scene with a modulated light source and observing the reflected light. In the proposed system, hand images and depth images are captured simultaneously and are then merged via UV mapping[6].

The letters U and V in a UV map denote the axes of the 2D texture because X, Y and Z are used to denote the axes of the 3D object in the model space. For any point P on the sphere, the unit vector from P to the sphere's origin can be calculated. Assuming that the sphere's poles are aligned with

 ISBN: 978-1-61208-422-0

the Y-axis, UV coordinates in the range of [0, 1] can then be calculated using (1).

$$u = 0.5 + \frac{\arctan2(d_z, d_x)}{2\pi};\ v = 0.5 - \frac{\arcsin(d_y)}{\pi} \quad (1)$$

C. 4D Feature Descriptors

In this work [1], we proposed a system that uses the position of joints and tips and then uses classification to recognize hand motions. The primary problem of this method is that it was not reliable for complex hand motions, such as grabbing and rotation. It was based on the assumption that all fingertips are visible and can be detected in the image, so it could not track the position of the all joints and tips when fingers occluded each other. To cope with this problem, two extended future descriptors are used. The first one is the extended method of Laptev et al., which provides a descriptor ρ of the visual appearance and local motion.

$$\rho(\upsilon, \nu, \omega) = (G(I(\upsilon, \nu, \omega)), F(I(\upsilon, \nu, \omega)), D(I(\upsilon, \nu, \omega))) \quad (2)$$

where G is a Histogram Oriented Gradient (HOG), F is Histogram Oriented Flow and D is Histogram Oriented Depth. A bag of words is employed on each ρ. Each ρ represents one type of hand pose. To cluster these poses, K-Means Clustering is performed on all with a Euclidean distance function.

The second descriptor is defined based on the algorithm that was developed by Oshin *et al*[7] and Hadfield[8], which is called the Relative Motion Descriptor (RMD). We extended the algorithm for pose estimation in 3D data. For each frame, we consider volume i_{xyz} with 3D dimensions. The sum of interest point s is defined based on the interest point detection and their strengths.

$$s(x, y, z, t, \tfrac{X}{\sigma}, \tfrac{Y}{\sigma}, \tfrac{Z}{\sigma}) = \sum_{x'=x}^{x+\frac{X}{\sigma}} \sum_{y'=y}^{y+\frac{Y}{\sigma}} \sum_{z'=z}^{z+\frac{Z}{\sigma}} \sum_{t'=t}^{t+\frac{T}{\sigma}} \iota(x', y', z', t') \quad (3)$$

where ι is the representation of the frame of the specific pose.

D. Classification

To classify each of the motions, the extended method of MACH filter proposed by [9] is used and is given by:

$$F(u, \upsilon, \omega, q) = \sum_{t=0}^{T-1} \sum_{z=0}^{N-1} \sum_{y=0}^{M-1} \sum_{x=0}^{L-1} f(x, y, z, t) \quad (4)$$

where f(x,y,z,t) are the 3D data corresponding to the temporal derivative and depth of the input sequence, and F(u,v,w,q) is the result in the frequency domain. To detect similar action in a testing video sequence, Inverse Fourier is applied to the filter and then to the video sequence.

$$F(u, \upsilon, \omega, q) = \sum_{t=0}^{T-1} \sum_{z=0}^{N-1} \sum_{y=0}^{M-1} \sum_{x=0}^{L-1} s(x, y, z, t) H(x, y, z, t) \quad (5)$$

where H is the filter in the time domain and s is the test video of hand motion.

E. Evaluation

To evaluate and differentiate any unnatural slowness of motions, after we classify the motion, we employ a HMM, which is widely used with time series data, such as speech and gesture recognition. We consider that hand motion consists of discrete hand poses and that each hand pose can be represented as a state. We define a bounded left-right model with the well-known transition model[10][11].

III. Conclusion

We have proposed a framework for 3D automatic hand motion evaluation with a SoftKinetic camera that solves some of the drawbacks of the existing methods. It can evaluate complex motions, whereas previous models are suitable for only simple motions. The proposed system captures images and depth data and then implements segmentation to extract the object of interest. UV mapping was used to merge RGB data with depth data. The classifier learned the characteristics of the points of interest based on the extracted features and then classified the hand postures. Finally, an HMM was used to evaluate and recognize movements based on the rate of the evolving motions.

Acknowledgment

The work is supported in part by the Canada Research Chair program, AUTO21 Networks of Centers of Excellence, the Natural Sciences and Engineering Research Council of Canada.

References

[1] A. Safaei and M. Jahed, "3D Hand Motion Evaluation Using HMM," Journal of Electrical and Computer Engineering Innovations (JECEI), vol. 1, 2013, pp. 11–18, ISSN: 2322-3952.

[2] "Kinect," 2015, URL: https://www.microsoft.com/en-us/kinectforwindows/ [accessed: April, 2015].

[3] "SoftKinetic," 2015, URL: http://www.softkinetic.com/ [accessed: April, 2015].

[4] "CIELab," 2015, URL: http://www.optelvision.com/documents /optelvision-s-explanation-on-cielab-color-space.pdf [accessed: April, 2015].

[5] T.-W. Chen, Y.-L. Chen, and S.-Y. Chien, "Fast image segmentation based on k-means clustering with histograms in hsv color space," in Multimedia Signal Processing, 2008 IEEE 10th Workshop on, Oct 2008, pp. 322–325.

[6] "UVMapping," 2015, URL: http://wiki.blender.org/index.php/Doc:2.4 /Manual/Textures/Mapping/UV [accessed: April, 2015].

[7] O. Oshin, A. Gilbert, and R. Bowden, "Capturing the relative distribution of features for action recognition," in Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, March 2011, pp. 111–116.

[8] S. Hadfield and R. Bowden, "Hollywood 3d: Recognizing actions in 3d natural scenes," in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, June 2013, pp. 3398–3405.

[9] M. Rodriguez, J. Ahmed, and M. Shah, "Action mach a spatio-temporal maximum average correlation height filter for action recognition," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, June 2008, pp. 1–8.

[10] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A hidden markov model-based continuous gesture recognition system for hand motion trajectory," in Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, Dec 2008, pp. 1–4.

[11] G. A. Fink, Markov models for pattern recognition: from theory to applications. Springer Science & Business Media, 2014.

 ISBN: 978-1-61208-422-0

Performance Study of Channel-QoS Aware Scheduler in LTE Downlink Using NS3

Adi S.M.Y., Kuokkwee Wee, Ee Mae A., Mohd. F.A.A.
Faculty of Information Science & Technology
Multimedia University
Melaka, Malaysia
adi.syukri.yusof@gmail.com, wee.kuok.kwee@mmu.edu.my, jessicaaem@live.com,
mfikriazli.abdullah@mmu.edu.my

***Abstract*— Recent mobile technologies allow users to interact with each other through voice and video by using smart-phones or tablets. The applications available on the Internet are mostly using a similar protocol which is best-effort data network by service provider. Best effort service does not guarantee data to be delivered within the allocated time. Real-time applications compete equally for network resources in best effort, hence real-time application perform poorly using best effort services. In this paper, two classical packet schedulers and two channel-aware schedulers for Orthogonal Frequency-Division Multiple Access (OFDMA) are studied. The schedulers are simulated using real-time packets with a various channel condition of user equipment (UE) in NS3 simulator to test the performance of each scheduler. The Channel-Quality of Service (QoS) Aware scheduler outperforms all other schedulers for real-time traffic.**

Keywords-VoIP, Real-Time, Video, NS3, simulation, LTE.

I. INTRODUCTION

Long Term Evolution also known as LTE was introduced by 3GPP earlier in 2004 but it was only finalized and approved in 2008 [1][2]. LTE network consists of LTE devices and System Architecture Evolution (SAE). SAE is an evolution from third generation mobile internet (3G) packet core network. LTE has become the new standard for mobile network. The downlink and uplink peak rate of LTE are able to achieve 300Mbit/s and 75Mbit/s respectively. Latency of LTE is controlled by the Quality of Service (QoS) that permits less than 5ms latency in the Radio Access Network (RAN) [3].

LTE is able to cater fast moving mobiles as well as to support both Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD) [4]. Although TDD seems to be overall a better choice, FDD is more widely implemented due to earlier technologies. LTE consists of a user plane called evolved Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access Network (eUTRAN) and a control plane called evolved packet core (EPC). In the eUTRAN, there is an evolved NodeB (eNodeB) which handles radio resource management (RRM) [5].

Real-time traffic has a strict requirement for delay. QoS ensures high-quality performance for critical applications. Traditionally network traffic uses best effort services which do not guarantee any reliability, delay, jitter, or other performance characteristics. On the Internet Protocol (IP) based network, the Integrated service (Intserv) and differentiated service (Diffserv) are able to provide preferential treatment to specified traffic.

The outcome of this paper is to analyze the capability of channel-QoS aware (CQA) in a mixed traffic environment. As a result of the simulation, CQA scheduler is able to allocate resources efficiently when considering the delay, channel condition, and guaranteed bit rate. We simulate Voice over IP (VoIP), video and File Transfer Protocol (FTP) traffic using NS3 to test the capability of schedulers as shown in Figure 1. The CQA scheduler is good for real-time traffic as it is able to provide better QoS as compared to other schedulers [6][7].

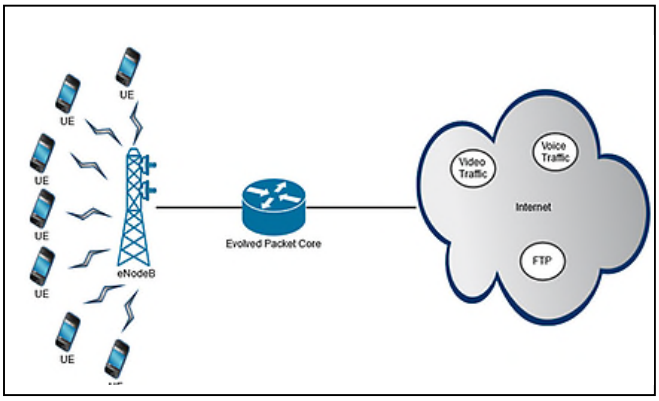


Figure 1. NS3 LTE Simulation Scenario.

This paper focuses on the downlink scheduler of the LTE network. The main focus is the QoS performance as it plays an important role in determining the improvement of QoS in LTE. This includes the networks throughput rate, delay, fairness and packet loss. The rest of the paper is organized as follows: Section II briefly describes several related works. Section III discusses the schedulers, while Section IV explains and elaborates on the methodology. Section V concludes the paper and indicates future works.

II. RELATED WORKS

The increase in the popularity of LTE technology system worldwide has gained more interest in the design of LTE packet scheduling algorithms. There are various downlink schedulers that were proposed by researchers. The advanced

 ISBN: 978-1-61208-422-0

technology of high speed wireless network has caused the trend of VoIP calling to grow. Apart from having VoIP, video conference is also a need in this era. VoIP and video conferencing is categorized under real-time traffics and has a strict delay requirement in order for it to perform well. However, classical schedulers are not designed to be aware of QoS requirement for real-time traffic.

Various studies have been performed in OFDMA systems. For example, Proportional Fair (PF) scheduling algorithm allocates resources accordingly in order to achieve fairness among users [8][9]. However, PF can neither provide the best throughput nor provide the best fairness. The basic scheduling algorithm Round Robin (RR) was also studied in trying to optimize the capabilities of LTE [5][6][10]. PF and RR are not able to handle the priority for type of traffics [6]. Next, Priority Set Scheduler (PSS) was studied. PSS was designed by combining frequency domain (FD) and time domain (TD) scheduler that aims to provide a defined target bit rate to all users [11]. PSS is able to achieve part of QoS requirement by using guaranteed bit rate (GBR) as the observable matrix. However, it does not consider delay requirements, making it perform less efficiently. Another scheduler CQA was proposed by [12], whereby it is designed to improve the resource allocation for real-time voice traffic. CQA sorts the traffic's priority according to the channel condition, head of line (HOL) delay, and GBR. Thus, VoIP through LTE structure improved significantly using CQA scheduler [12].

III. DOWNLINK SCHEDULERS

A. Proportional Fairness

PF in NS3 works by allocating resources to User Equipment's (UE) when the UEs' channel quality is instantaneously high even though the average channel condition over time is low.

Using the PF scheduler, UEs are allocated to different Resource Block Group (RBG). Channel conditions and the throughput value of previous transmission is used to calculate the metrics value [9][13].

B. Round Robin

RR can be considered as the simplest scheduler. The scheduler works by dividing the resource blocks (RB) between the flows with non-empty queues. The scheduler method is to divide all the available resources to active traffic flows if the resource quantity is able to serve all the incoming traffic flows [8][9].

RR will not be able to allocate the resource to all flows if the traffic flow is bigger than the RB. RR will only allocate resources based on time to interval (TTI) and continue to allocate resources in the next sub frame and start from the last unallocated flow.

C. Priority Set Scheduler

PSS is a scheduler that combines FD and TD. It targets to provide fairness to UEs by using a specified Target Bit Rate (TBR). PSS works by selecting UEs which Radio Link Control (RLC) buffer is not empty [1]. The UEs are then divided into two according to their TBR. By dividing the UEs according to their TBR, the scheduler is able to decide based on the priority of serving those UEs [11].

UEs with the highest priority metric are forwarded to the FD scheduler. Then, the RB for each UE is allocated by using PF scheduler to calculate the metric. In the case of having a minimal number of UEs, FD scheduler provides a weight metric to control the fairness.

D. Channel-QoS Aware (CQA)

The CQA works by taking HOL delay, channel quality, and GBR parameter into consideration. CQA also utilizes TD and FD scheduler where it depends on channel quality and QoS requirements to allocate resources. This allows the attainability of a higher amount of spectral efficiency while satisfying the traffic delay requirements [12].

Similarly to PSS, CQA scheduler divides UEs according to their priority in the TD scheduling. The CQA scheduler then groups the UEs into flows and ensures that the FD scheduler allocates resources starting with flows which consists the highest computed metric. The metric is calculated using HOL, channel quality, and GBR.

The channel quality can be calculated using two methods which are PF or frequency selective fading. The GBR is specified in EPS bearer when simulating the CQA. In other words, CQA is aware of its channel conditions as well as its QoS parameters. Hence, a minor requirement on the networks performance is made available which can be in the form of guaranteed amount of data, etc. [14].

IV. LTE IN NS3

A. Voice Traffic

Simulation of VoIP traffic in NS3 is characterized by two periods; ON and OFF. ON is for the time when the users spend on talking whereby constant packets are transmitted at regular intervals. The OFF time is the time where the user stops from talking and packets are not transmitted.

Parameters and details for the traffic simulation are illustrated in TABLE I. ON and OFF time are given as 0.352 and 0.650 seconds respectively [15]. Actual bit rate depends on the codec used and the packetization time. In [15], G.711 codec is used. G.711 does not perform any compression and ensures the best voice quality. Due to the absence of compression rules, bandwidth requirements will be high.

TABLE I. VOICE TRAFFIC CONFIGURATION

Voice Traffic	Details
ON	0.352 seconds
OFF	0.650 seconds
Codec	G.711
Output	64 kbps
Bandwidth per voice call	200 bytes
Bandwidth at IP layer	80 kbps per call
Average bandwidth	28.1 kbps

 ISBN: 978-1-61208-422-0

OnOffHelper is used to generate the traffic. In Figure 2, it is an example of the script to simulate voice traffic with OnOffHelper module.

```
OnOffHelper P1dlVoIPonoff=OnOffHelper("ns3::UdpSocketFactory",InetSocketAddress (ueIpIface.GetAddress (u), pdlPort));
P1dlVoIPonoff.SetAttribute ("OnTime", StringValue("ns3::ConstantRandomVariable[Constant=0.352]"));
P1dlVoIPonoff.SetAttribute ("OffTime", StringValue("ns3::ConstantRandomVariable[Constant=0.65]"));
P1dlVoIPonoff.SetAttribute("PacketSize", UintegerValue(VoIPdlDataSize));
P1dlVoIPonoff.SetAttribute("DataRate",DataRateValue(VoIPdlDataRate));
ApplicationContainer P1dlVoIPapp = P1dlVoIPonoff.Install (PiremoteHost);
P1dlVoIPapp.Start (Seconds(1.0));
P1dlVoIPapp.Stop(Seconds(simTime));
```

Figure 2. Voice ON OFF Helper.

B. *Video Traffic*

The simulation of video traffic requires Evalvid module in NS3. The Evalvid module simulates video by tracing the frame of the video. The behavior of this simulation is according to real time services such as video conferencing. In this simulation, the module uses st_highway_cif.st as the trace file for video traffic.

C. *Best Effort Traffic*

To simulate best effort application in NS3, a UDP echo module is used. The UDP echo is setup using echo server and echo helper. There are 3 attributes that are maximum packet transmitted, the time interval between packets, and the packet size.

D. *Simulation Setup*

The number of UEs starts from 5 UEs to 80 UEs with the interval of 5. The parameter is configured using the command line attribute. It is set within the script to enable the arguments to be parsed in command line. An example of parsing the command line argument is presented in Figure 3. The node for eNodeB and UE is configured with LTE stack protocol. All UEs are attached to the eNodeB and Radio Resource Control (RRC) connection is created between them.

SrsPeriodicity sets the maximum number of allowable UEs to be attached to eNodeB. Radio Network Temporary Identifier (RNTI) is generated to address the UEs. Collision requests from UE to eNodeB cause some RNTIs to be unused. Therefore, to enable the simulation to run more than 30 users, the srsPeriodicity needs to be configured to 160. This will enable the simulation to run without error. Figure 4 is the script to set the srsPeriodicity. The Internet is then created using PointToPointHelper which the attribute of the DataRate is set to 100Gb/s. The mobility model is also setup to position the UEs in different areas. The scheduler is setup using SetSchedulerType and the attribute of path loss model is set to Friis Spectrum Propagation Loss Model. LTE device is installed to the nodes and all the traces are enabled. The simulator is configured to run for 50ms before it stops and it is destroyed.

```
// Command line arguments
CommandLine cmd;
cmd.AddValue("numberOfNodes", "Number of eNodeBs + UE pairs", numberOfNodes);
cmd.AddValue("simTime", "Total duration of the simulation [s])", simTime);
cmd.AddValue("distance", "Distance between eNBs [m]", distance);
cmd.AddValue("interPacketInterval", "Inter packet interval [ms])", interPacketInterval);
cmd.Parse(argc, argv);
```

Figure 3. Parsing Command Line Argument.

```
static ns3::GlobalValue g_srsPeriodicity ("srsPeriodicity",
                                           "SRS Periodicity (has to be at least "
                                           "greater than the number of UEs per eNB)
                                           ns3::UintegerValue (160),
                                           ns3::MakeUintegerChecker<uint16_t> ());
```

Figure 4. Setting SrsPeriodicity.

To simulate traffic from the Internet to eNodeB and to UEs, multiple remote hosts are setup. It acts as the host to each type of traffic. Each remote host is given an IP address and route using routing helper to the default gateway for eNodeB.

This simulation runs using Waf, which is a python-based framework for configuring, compiling and installing applications. A shell script is created to execute the Waf program in order to run the simulation. The simulation is performed multiple times by looping the command in the shell script.

The simulation is monitored by a module called flow monitor. The flow monitor is installed in UE and remote host. While monitoring the flow of the packet, the flow monitor records the data which can be called using a pointer. The flow monitor automatically maps the packets with all the data using flow ID. The throughput and delay can be calculated using the flow monitor. The trace file is loaded into Octave to compute the fairness index, packet loss ratio, throughput, and delay then plot the graph, as in Figure 5.

```
Logger << "Flow ID: " << iter->first << " Src Addr " << t.sourceAddress << " Dst Addr " << t.destinationAddress << std::endl;

Logger << "Tx Packets = " << iter->second.txPackets << std::endl;

Logger << "Rx Packets = " << iter->second.rxPackets << std::endl;
Logger << "Throughput: " << iter->second.rxBytes * 8.0 / simTime / 1024 / 1024  << " Mbps" << std::endl;
Logger << "Packets Loss Ratio: " << ((  iter->second.lostPackets * 100) /iter->second.txPackets) << "%" << std::endl;
plr << ((  iter->second.lostPackets * 100) /iter->second.txPackets) << std::endl;
```

Figure 5. Flow Monitor

The entire flow diagram for configuring the simulation for NS3 is shown in Figure 6.

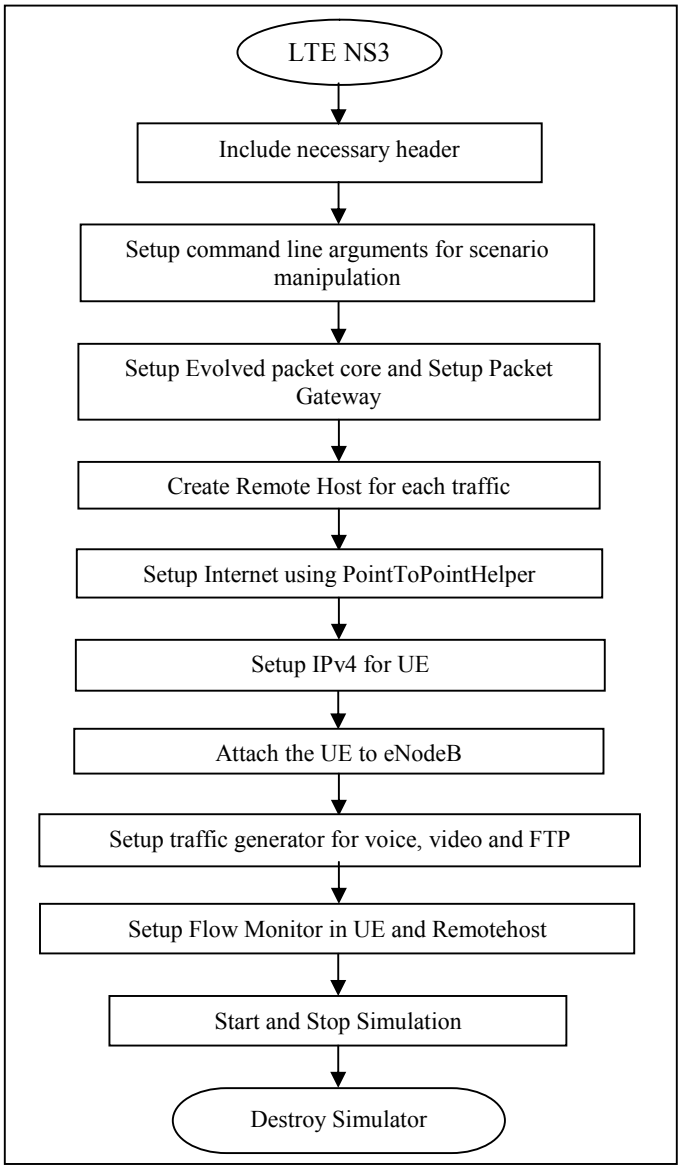


Figure 6. Flow diagram of NS3 LTE Configuration

V. PERFORMANCE EVALUATION

After simulating the VoIP traffic and video traffic of the LTE network in NS3, the graphs are plot using Octave. The outputs from the simulation are throughput, delay, packet loss ratio and fairness index.

A. VoIP Result

The CQA throughput is better than PSS where the throughput is higher starting from 30 UEs until 80 UEs. This can be observed in Figure 7. Compared to RR, throughput of VoIP is higher than CQA. This is the result of RR allocating resources without considering the channel condition of the UE. UEs with poor channel condition are able to get resources but are unable to transmit resources efficiently on time.

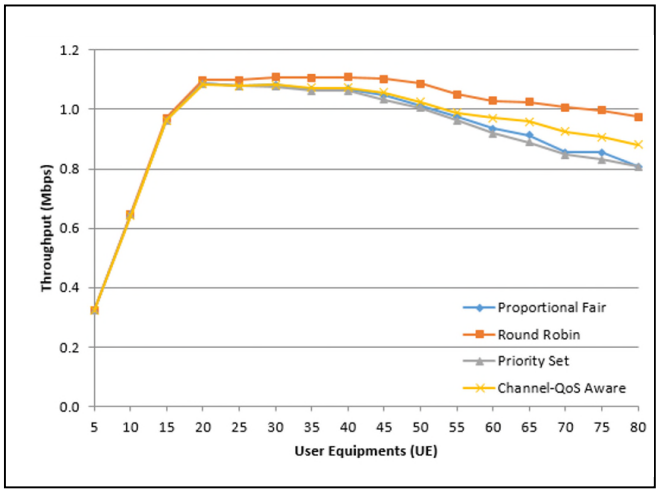


Figure 7. VoIP Downlink Throughput.

Delay for VoIP in CQA scheduler is 28.85% quicker compared to all other scheduler. This means that the voice packets are able to be transmitted in a short amount of time. The delay result is presented in Figure 8.

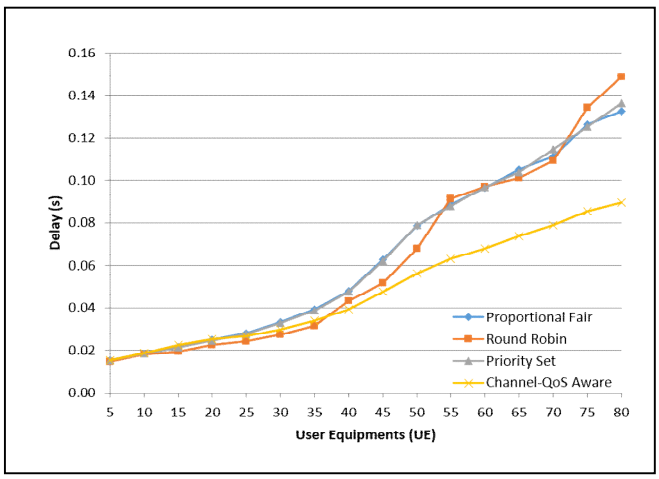


Figure 8. VoIP Downlink Delay.

B. Video Result

As shown in Figure 9, the video throughput for CQA scheduler is 7% lower compared to the other schedulers. This is because CQA scheduler allocates RB while considering the GBR of 64kbps for the UE to transmit the video data.

The CQA scheduler delay is 14.97% lower compared to other schedulers. Based on the results given in Figure 10, it shows that by monitoring the GBR and HOL delay, CQA could optimize the resource allocation and allocate the resource block efficiently.

 ISBN: 978-1-61208-422-0

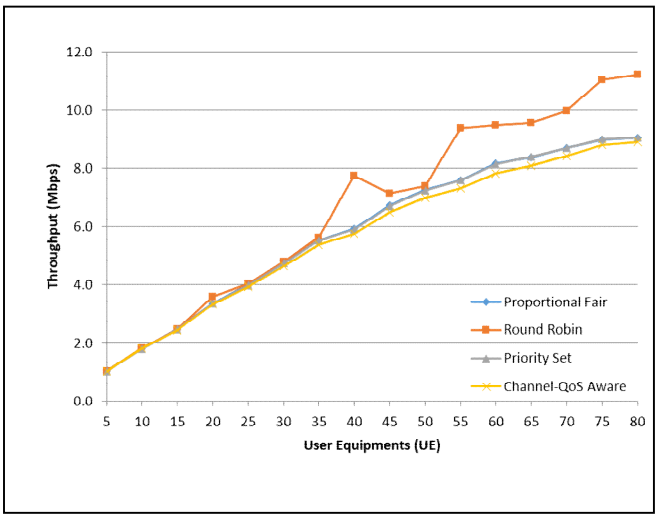


Figure 9. Video Downlink Throughput.

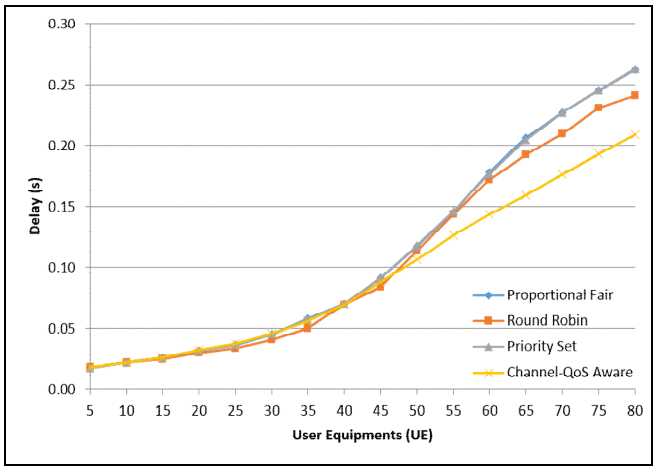


Figure 10. Video Downlink Delay.

C. *Best Effort Result*

The graph for UDP traffic that utilizes best effort technique is displayed in Figure 11. The throughput is 11.38% higher than PSS and RR. Since non-real-time traffic does not have strict delay requirement, the resource allocation prioritize real-time traffic while not jeopardizing the non-real time traffic. Therefore, the CQA is still able to allocate resources to enable the best effort traffic to be transmitted optimally.

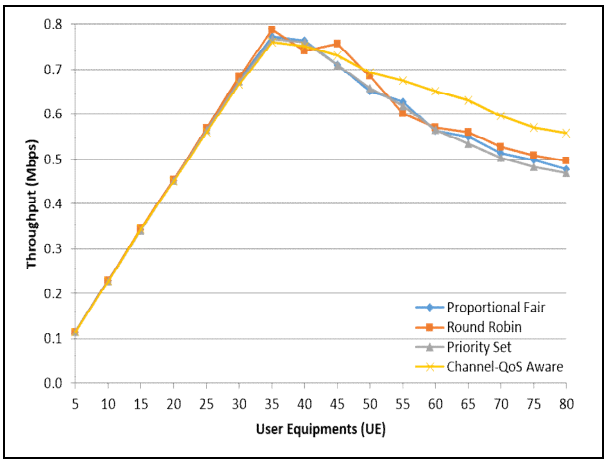


Figure 11. BE Throughput.

The delay for CQA scheduler is 14.96% higher. However, considering this is not real-time traffic, the CQA scheduler seems to fulfill its purpose. Figure 12 illustrates the best effort results for UDP traffic.

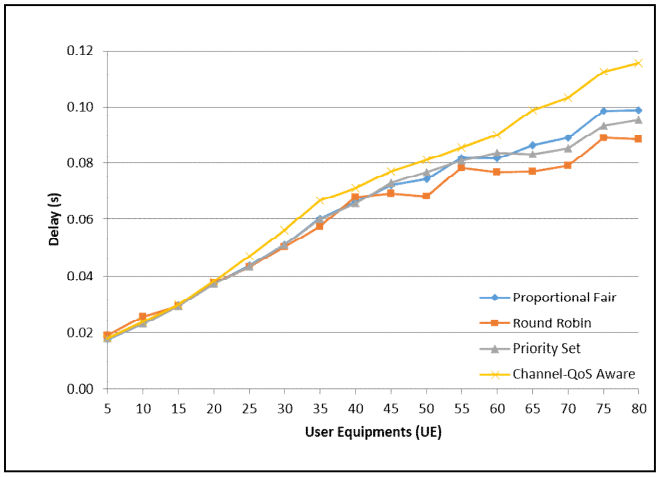


Figure 12. BE Delay.

D. *Fairness Index & Packet Loss Ratio*

The fairness index checks for the fairness in terms of resource allocation between UE. Figure 13 presents a histogram which shows that CQA scheduler performs best starting at 55 users. It is calculated that CQA is 1.94% fairer from 55 users to 80 users.

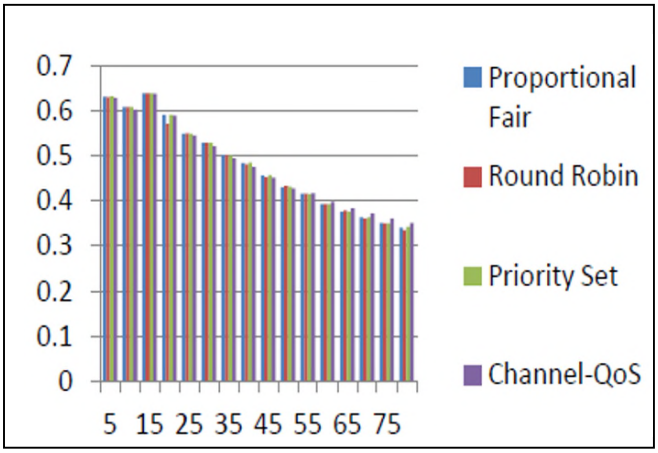


Figure 13. Fairness Index

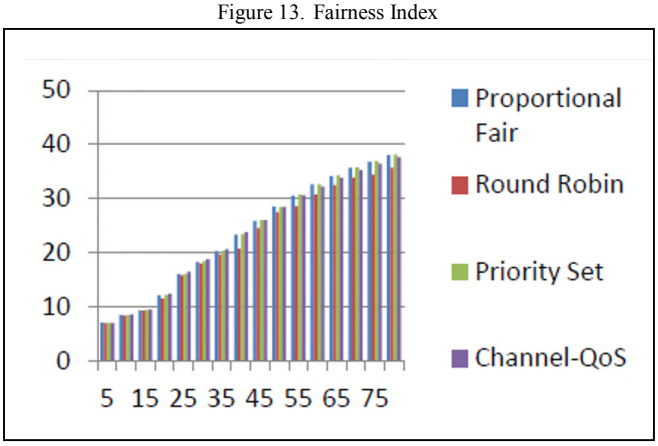


Figure 14. Packet Loss Ratio

Packet loss ratio defines the total of number for packet loss during the transmission of the packets. CQA scheduler has less packet loss ratio compared to PSS at around 0.21% lesser. This can be seen in Figure 14.

VI. CONCLUSION

Based on our study, PF and PSS have similar results. CQA scheduler is more suitable for real-time application compared to RR where delay of packets and fairness of resource allocation is an issue for RR. The delay for CQA is 15.23% better than PSS in video and 25.06% better than PSS in VoIP. Overall, we can conclude that CQA scheduler is more suitable for real-time application compared to PSS.

Further studies will be conducted on how to improve the throughput for video application while having optimum results for voice and best effort traffic.

REFERENCES

[1] "Evolved Universal Terrestrial Radio Access (E-UTRA): Radio Link Control (RLC) protocol specification 3GPP TS 36.322." [Retrieved Online: March, 2015]. Available: http://www.3gpp.org/dynareport/36322.htm.

[2] "Evolved Universal Terrestrial Radio Access (E-UTRA): Radio Resource Control (RRC) protocol specification 3GPP TS 36.331." [Retrieved Online: March, 2015]. Available: http://www.3gpp.org/dynareport/36331.htm.

[3] I. Koutsopoulos and L. Tassiulas, "Channel state-adaptive techniques for throughput enhancement in wireless broadband networks," Proc. IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No.01CH37213), vol. 2, 2001, pp. 757-766, doi: 10.1109/INFCOM.2001.916266.

[4] S. Sesia, I. Toufik, and M. Baker. LTE - The UMTS Long Term Evolution: From Theory to Practice, 2nd ed., John Wiley & Sons Ltd., 2011, pp.1–794.

[5] D. Zhou, N. Baldo, and M. Miozzo, "Implementation and Validation of LTE Downlink Schedulers for ns-3," Proc. Sixth International Conference on Simulation Tools and Techniques, 2013, pp. 211-218, doi: 10.4108/simutools.2013.251608.

[6] G. Piro, N. Baldo, and M. Miozzo, "An LTE module for the ns-3 network simulator," Proc. 4th International ICST Conference on Simulation Tools and Techniques, 2011, pp. 415–422, doi: 10.4108/icst.simutools.2011.245571.

[7] "ns-3" [Retrieved Online: November 2014]. Available: http://www.nsnam.org/.

[8] Y. Barayan and I. Kostanic, "Performance Evaluation of Proportional Fairness Scheduling in LTE," Proc. World Congress on Engineering and Computer Science, vol. 2, 2013, pp. 23–25.

[9] H. J. Zhu and R. H. M. Hafez, "Scheduling schemes for multimedia service in wireless OFDM systems," IEEE Wireless Communications, vol. 14, no. 15, 2007, pp. 99–105, doi: 10.1109/MWC.2007.4396949.

[10] N. Ruangchaijatupon and Y. J. Y. Ji, "Simple Proportional Fairness Scheduling for OFDMA Frame-Based Wireless Systems," Proc. IEEE Wireless Communications and Networking Conference, 2008, pp. 1593–1597, doi: 10.1109/WCNC.2008.285.

[11] G. Monghal, K. I. Pedersen, I. Z. Kovacs, and P. E. Mogensen, "QoS Oriented Time and Frequency Domain Packet Schedulers for The UTRAN Long Term Evolution," Proc. VTC Spring 2008 - IEEE Vehicular Technology Conference, 2008, pp. 2532–2536, doi: 10.1109/VETECS.2008.557.

[12] B. Bojovic and N. Baldo, "A new channel and QoS aware scheduler to enhance the capacity of voice over LTE systems," Proc. IEEE 11th Int. Multi-Conference Syst. Signals Devices, SSD, 2014, pp. 1–6, doi: 10.1109/SSD.2014.6808890.

[13] M. Kawser, H. M. A. B. Farid, A. M. J. Sadik, and I. K. Razu, "Performance Comparison between Round Robin and Proportional Fair Scheduling Methods for LTE," International Journal Information and Electronics Engineering, vol. 2, no. 5, 2012, pp. 678-681, doi: 10.7763/IJIEE.2012.V2.186.

[14] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink Packet Scheduling in LTE Cellular Networks : Key Design Issues and a Survey," IEEE Transaction on Vehicular Technology, vol. 15, no. 2, 2012, pp. 678–700, doi: 10.1109/SURV.2012.060912.00100.

[15] T. Janevski, "Traffic Analysis and Design of Wireless IP Networks," Proc. Global Telecommunications Conference, 2000, pp. 1–6.

 ISBN: 978-1-61208-422-0

Dynamic Node Movement Control in a Mobile Medium Ad hoc Network

Hanin Almutairi, John DeDourek, Przemyslaw Pochec
Faculty of Computer Science
University of New Brunswick
Fredericton, Canada
e-mail: {hanin.almutairi, dedourek, pochec}@unb.ca

***Abstract*—A Mobile Ad hoc Network (MANET) is a network of wireless mobile devices capable of communicating with one another without any reliance on a fixed infrastructure. A Mobile Medium Network is a set of mobile forwarding nodes functioning as relays for facilitating communication between the users of this Mobile Medium. The performance of the Mobile Medium depends on the Mobile Medium node density, distribution and movement. In the proposed dynamic node movement, the movement is determined based on whether the node is on a forwarding path for a data flow or not. Simulation results show that slowing down the speed of mobile nodes when they are forwarding significantly affects the delivery rates in Mobile Medium networks. For networks with a few forwarding nodes dispersed in a large region reducing the mode movement speed by 50% results in an approximately 20% improvement in the delivery ratio, with even higher improvements possible at lower speeds.**

Keywords-mobility models; Mobile Medium; self-organizing mobile network; M2ANET; SMMANET; MANET; AODV; DSDV; ns2 simulation

I. INTRODUCTION

A Mobile Ad hoc Network (MANET) is a set of mobile devices that cooperate with each other by exchanging messages and forwarding data [1][2]. A Mobile Medium Ad hoc Network (M2ANET) proposed in [3] is a particular configuration of a typical MANET where all mobile nodes are divided into two categories: (i) the forwarding only nodes (shown in black in Figure 1) forming the so called Mobile Medium, and (ii) the communicating nodes (shown in red in Figure 1), mobile or otherwise, that send data and use this Mobile Medium for communication. The advantage of this M2ANET model is that the performance of such a network is based on how well the Mobile Medium can carry the messages between the communicating nodes and not based on whether all mobile nodes form a fully connected network. An example of a M2ANET is a cloud of autonomous drones released over an area of interest facilitating communication in this area. The movement of nodes in a M2ANET can be predefined by the user, selected at random or purposefully controlled for the best performance. When the mobile nodes select themselves their movement we refer to such a network as a Self-organizing Mobile Medium Ad hoc Network (SMMANET). Recently, a number of projects that match the M2ANET model have been announced; they include Google Loon stratospheric balloons [4] and Facebook high altitude solar powered planes [5] for providing Internet services to remote areas, and the Swarming Micro Air Vehicle Network (SMAVNET) project where remote controlled planes are used for create an emergency network [6].

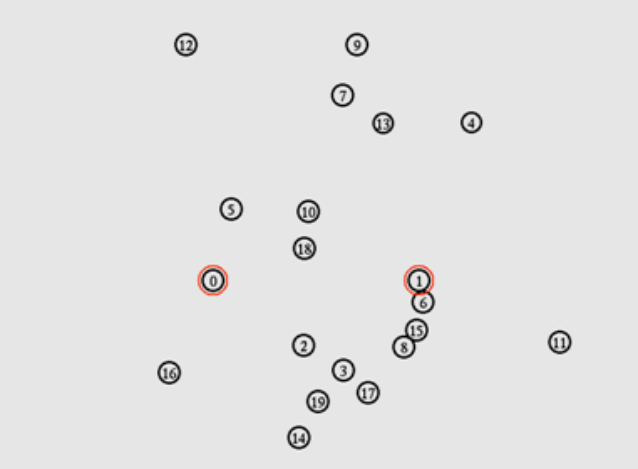


Figure 1. ns2 simulation screen of a M2ANET

Controlling the movement of the forwarding nodes forming the Mobile Medium is a problem in deploying M2ANETs. Random movement, while easy to implement, suffers from the difficulty of how to keep nodes in a sufficient density over the area of interest where the communication infrastructure is to be supplied by means of a Mobile Medium. A Mobile Medium with low mobile node density suffers from frequent disconnections and rerouting, resulting in a low delivery ratio. In this paper, we propose a solution for controlling the movement of the Mobile Medium nodes for M2ANET (SMMANET) deployments. The solution is based on the following observation: In a typical M2ANET only a few (if any) mobile nodes are actively forwarding the network traffic for a given flow at any given time. The longer the forwarding path is maintained the better the delivery ratio. We therefore propose to slow down the movement of the nodes that are actively forwarding data in the M2ANET. The new approach is then compared against the standard Random Way Point (RWP) movement.

In Section II, we present background on Mobile Medium networks and mobility patterns. The new movement pattern based on changing the speed of mobile nodes that happen to be on the forwarding path for a flow is discussed in Section

 ISBN: 978-1-61208-422-0

III. Simulation experiments of this movement under different scenarios are in Section IV. Finally, we present the experimental results in Section V, followed by the conclusion and future work.

II. STATE OF THE ART

A MANET is comprised of interconnected mobile nodes, which make use of wireless communication links for multi-hop transmission of data. They offer distinct advantages over infrastructure based networks and are versatile for some particular applications and environments. There are no fixed or prerequisite base stations or infrastructures; therefore, their set up is not time consuming and can be done at any time and in any place. MANETs exhibit a fault-resilient nature, given that they are not operating a single point of failure and are very flexible. The deletion and addition of new nodes, forming new links are a normal part of operation of a MANET [1][7][8]. A group of nodes can facilitate communication between distant stations by forming a Mobile Medium, as introduced in [3].

Many mobility models have been proposed for recreating the real world application scenarios of MANETs. A mobility model attempts to mimic the movement of real mobile nodes that change speed and direction with time. There are two main types of mobility models currently used in simulation of MANETs [2][9]: trace and synthetic. A trace uses actual node movements that have been observed in a real system. In the absence of traces, synthetic mobility models can be used. The synthetic models attempt to realistically mimic the movements of mobile nodes in mobile networks [2]. The categorization of synthetic models is based on interactions between the nodes and the environment in a mobile network [2]: we can distinguish between individual node movements and group node movements. Based on specific mobility characteristics these models can be further classified into four categories: models with temporal dependency, models with spatial dependency, models with geographic restriction, and random models [2]. In the mobility model with temporal dependency the movement of a mobile node is affected by its movement history. A node's current movement is affected by past movement such as in the Gauss Markov Model and the Smooth Random Mobility model [2]. In mobility models with spatial dependency, the mobile nodes tend to travel into a group and are interdependent one on another. The movement of a node is affected by surrounding nodes in group mobility such as in the Reference Point Group Model [2]. One implementation of a group mobility model uses the attraction/repulsion principle to maintain the connections between the mobile nodes: when nodes get too close (e.g., signal strength is high) they select the next move with a preference in the opposite direction to the incoming signal and conversely, when the signal gets weak, they turn back [10]. Another class is the mobility models with geographic restriction. There, the mobile node movement is limited to certain geographical areas, such as streets or freeways, as for example in the Pathway Mobility Model and the Obstacle Mobility Model [2].

In simulation, a random mobility is often used as a reference case scenario, mostly because of the relative ease of implementing it in a simulator. One of these popular models is the Random Way Point (RWP) model available in ns2 [11]. Nodes are moved in a piecewise linear fashion, with each linear segment pointing to a randomly selected destination and the node moving at a constant, but randomly selected speed.

III. DYNAMIC MOVEMENT CONTROL IN MOBILE MEDIUM

Recall that the Mobile Medium is a particular type of a mobile network that is used for providing communication services to mobile or stationary users. It consists of a cloud of mobile nodes whose sole function is to forward data. Users of this Mobile Medium connect to it by establishing a link to one of its nodes and then send data to other users connected through this medium (Figure 2). Our interest here is to assure the best delivery ratio for the data sent form one user to another through the Mobile Medium.

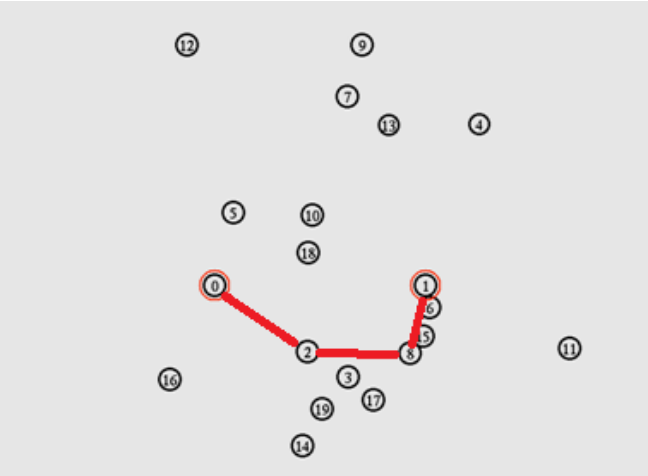


Figure 2. Nodes on the forwarding path

We start with a common reference scenario where a number of Mobile Medium nodes move in a restricted region (a rectangle) in random directions according to the RWP model. The two user stations (nodes) that wish to communicate are placed in the same region, remain stationary and send data one to another through the Mobile Medium (Figure 2). Because the Mobile Medium nodes move in and out of range of the communicating stations the connections break and the routing path changes, affecting the delivery ratio.

The Dynamic Movement Control in Mobile Medium is an attempt aimed at reducing the (number of) changes in the routing path during the data transfer between the users. Normally, the nodes in the RWP model move in random directions and at random speeds independent of whether they are forwarding packets or not. The faster they move the quicker they get out of range and the sooner the connection path breaks. In the proposed Dynamic Movement Control for

 ISBN: 978-1-61208-422-0

the Mobile Medium, the nodes that forward data are slowed down. Note that this decision to slow down does not require any global information or location data; it is solely based on the forwarding status of the node. The easiest way to determine if the node is in the forwarding state is to check if it had forwarded a data packet recently: for example, for the scenario with a Constant Bit Rate (CBR) traffic source a node would maintain its forwarding status, and move at a reduced speed, for at least the time defined by the inter packet interval of the CBR source. For other scenarios, like the exponential traffic sources, multiples of the average inter packet interval can be used for determining the forwarding status maintenance period for a node. In the ns2 simulation we used the actual path determined by the simulator to determine which nodes are in the forwarding state.

The evaluation of the proposed Dynamic Movement Control for Mobile Medium was performed in the ns2 simulator. The simulator is implemented in c++ and each simulation is controlled by an Object Oriented Tool Command Language (OTCL) script. Unfortunately, the OTCL scripting language does not provide any access to the routing information used in the simulation. To access the routing information used by the Ad hoc On-Demand Distance Vector (AODV) protocol we modified the

```
void forward (aodv_rt_entry *rt,
     Packet *p, double delay);
```

function in `aodv.cc`, and to get routing for the Destination-Sequenced Distance Vector (DSDV) the function

```
void forwardPacket (Packet *p);
```

in `dsdv.cc` was modified.

IV. SIMULATION EXPERIMENT

A set of simulation experiments was conducted to evaluate the proposed Dynamic Movement Control for the Mobile Medium networks. Each simulation of a network consists of a different number of nodes roaming (RWP movement, average speed 4 *m/s*) in a square 1000 x 1000 meters. The node transmission range is 250m. The link data rate is 1 Mbps. Every packet has a size of 512 bytes. The buffer size at each node is 50 packets. Data packets are generated following a Constant Bit Rate process [11]. The source and destination nodes are stationary and located at coordinates (50, 400) and (950, 600). The summary of the simulation parameters used in ns2 is shown in Table 1. In each experiment, the designated source node transmits to one designated destination node for 500 seconds. For each node density and speed, the experiment is run three times and the average delivery ratio is reported.

TABLE I. SIMULATION PARAMETERS

Parameters	
Simulator	NS-2.34
Channel Type	Channel / Wireless Channel
Network Interface Type	Phy/WirelessPhy
Mac Type	Mac/802.11
Radio-Propagation Type	Propagation/Two-ray ground
Interface Queue Type	Queue/Drop Tail
Link Layer Type	LL
Antenna	Antenna/Omni Antenna
Maximum Packet in ifq	50
Area (n * n)	1000 x 1000m
Source node location	(50, 400)
Destination node location	(950, 600)
Source Type	CBR over UDP packetSize_ 512 interval_ 0.05
Simulation Time	500 s
Routing Protocol	AODV and DSDV

A. AODV performance

Reducing the speed of the forwarding nodes running the AODV routing protocol results in an improved delivery ratio for all but the very low node density (5 nodes over 1000x1000m region) scenario, Fig 3. The networks with a very few nodes in the Mobile Medium, i.e., with 5 and 10 nodes, as expected show very low delivery ratio. Under normal operating conditions for the Mobile Medium, with 20 or more mobile forwarding nodes, the performance improves gradually. Networks with a large number of nodes, more than 30 in a 1000x1000m area, as expected work well at any node speed with delivery ration ranging from 75% to almost 100%. The most interesting case is the network with a moderate number of nodes, 20 nodes over the area 1000x1000m. This scenario demonstrates best the benefits of the Dynamic Movement Control for Mobile Medium networks. In this scenario the delivery ration is very low 30% for nodes moving at the original speeds, but the performance improves significantly to 85% when the speed is reduced to 10% of the original speed.

The benefits are the highest in the low to medium range of node densities, 10 to 20 nodes moving in a 1000x1000m region, as depicted in Figure 4.

B. DSDV performance

DSDV performance improvement is similar to AODV: low improvements at low node densities (5 to 10 nodes) and a gradual improvement with the reduction of the speed of the forwarding nodes at higher node densities, as shown in Figures 5 and 6. In our experiments, the overall delivery ratio is lower for DSDV than for AODV. This is a common situation for the DSDV protocol when used in the ad hoc networks with a dynamically changing topology [12] and can be attributed to the proactive nature of the DSDV protocol where distributing the routing information to all the nodes and detecting a valid route can take a considerable time.

 ISBN: 978-1-61208-422-0

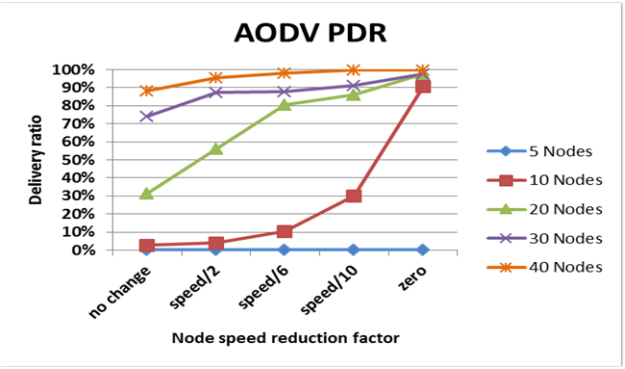


Figure 3. AODV delivery ratio.

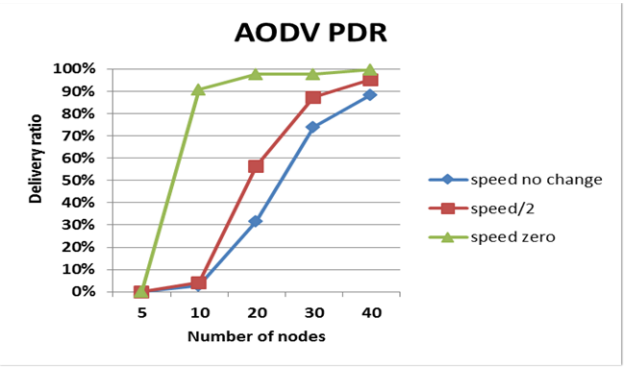


Figure 4. Delivery ratio for AODV at three speeds: no change, with 50% reduction and zero.

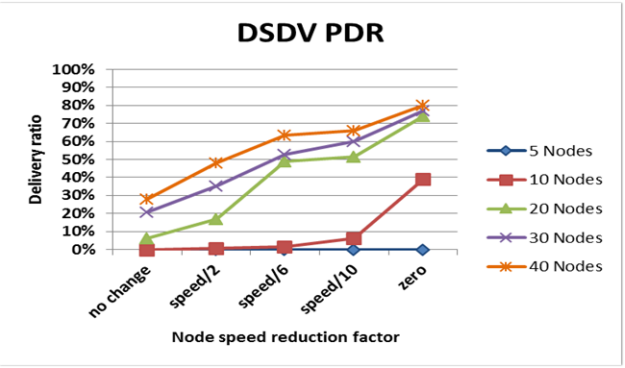


Figure 5. DSDV delivery ratio

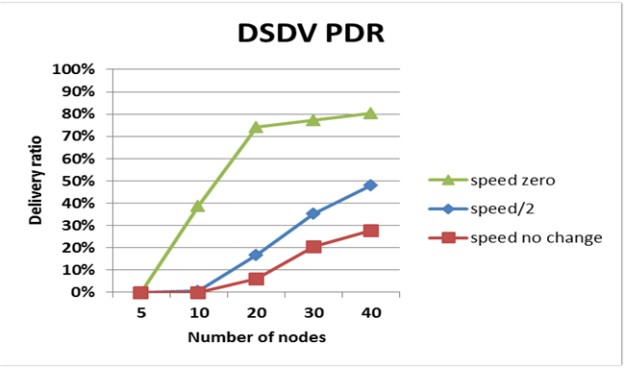


Figure 6. Delivery ratio for DSDV at three speeds: no change, with 50% reduction and zero.

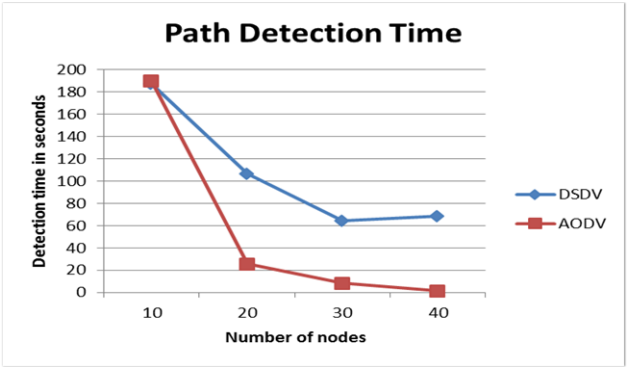


Figure 7. Time to detect the first path, for two different protocols.

In our experiments, no routes were detected in networks with 5 nodes and it took approximately 80 seconds to detect the first route in the networks with 20 or more nodes running DSDV, which is considerably slower than the path detection delay observed for the AODV protocol, see Figure 7.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a new node movement control paradigm for a self-organizing MANET network. The approach is particularly attractive for M2ANETs where the goal is to create a Mobile Medium out of mobile forwarding nodes, and use this Mobile Medium to facilitate data communication between other users. The new mobility control mechanism is based on slowing down the nodes that are on the forwarding path for a flow. The performance of the M2ANET/SMMANET increased significantly with the lowering of the speed of the nodes actively forwarding the data, with a higher relative improvement for the networks with moderate node densities. The improvements were observed in the networks running two different routing protocols AODV and DSDV.

Based on our results, we suggest further testing self-organizing M2ANET/SMMANET networks using different experimental scenarios. In particular while our experiments involved two stationary communicating nodes it might be more realistic for some application scenarios to model them as mobile. Combining the attraction/repulsion mechanism proposed in [10] with the new Dynamic Control may also result in further improvement of the performance of a SMMANET.

ACKNOWLEDGMENT

This work is sponsored and funded by the Ministry of Higher Education of Saudi Arabia through the Saudi Arabian Cultural Bureau in Canada.

REFERENCES

[1] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic (Eds.), Mobile Ad Hoc Networking, New York, Wiley-IEEE Press, 2001.
[2] F. Bei and A. Helmy, A survey of mobility models in wireless Ad hoc Networks, University of California, USA,2004.
[3] J. DeDourek and P. Pochec, "M2ANET: a Mobile Medium Ad Hoc Network", Wireless Sensor Networks: Theory and Practice, WSN 2011, Paris, France, Feb. 2011, pp. 1-4.

 ISBN: 978-1-61208-422-0

[4] H. Hodson, "Google's Project Loon to float the internet on balloons", New Scientist, October 2013.
[5] J. Brustein, "Facebook's Flying Internet Service, Brought to You by Drones", Bloomberg Businessweek, March 4, 2014.
[6] A. Jimenez Pacheco, et al., "Implementation of a Wireless Mesh Network of Ultra Light MAVs with Dynamic Routing", IEEE GLOBECOM 2012, 3rd International IEEE Workshop on Wireless Networking & Control for Unmanned Autonomous Vehicles 2012, Anaheim, California, USA, 2012.
[7] D. P. Agrawal and Q. A. Zeng, Introduction to Wireless and Mobile Systems, Thomson Engineering, 2010, 1591-1596.
[8] S. K. Sarkar, T. G. Basavaraju, and C. Puttamadappa, Ad Hoc Mobile Wireless Networks, Principles, Protocols, and Applications, Auerbach Publications Taylor & Francis Group, 2007.
[9] N. Aschenbruck, E. G. Padilla, and P. Martini, "A survey on mobility models for performance analysis in tactical mobile networks", Journal of Telecommunications and Information Technology, vol. 2, 2008, pp. 54-61.
[10] N. Alsalmi, J. DeDourek, and P. Pochec, "Self-organizing Mobile Medium Ad hoc Network", The Fourth International Conference on Mobile Services, Resources, and Users MOBILITY 2014, July 20 - 24, 2014 - Paris, France, pp. 38-42.
[11] H. Ekram and T. Issariyakul, Introduction to Network Simulator NS2, Springer, 2009.
[12] S. Mohapatra and P. Kanungo, "Performance analysis of AODV, DSR, OLSR and DSDV Routing Protocols using NS2 Simulator", Procedia Engineering (International Conference on Communication Technology and System Design 2011), v. 30, 2012, pp. 69-76.

 ISBN: 978-1-61208-422-0

E-business Collaboration Club

A Case of Public Relation and Information Technology

Yuh-Wen Chen, Jung-Cheng Hsieh
Institute of Industrial Engineering and Management
Da-Yeh University
Chang-Hwa, Taiwan
e-mail: profchen@mail.dyu.edu.tw

Steve Hsieh
Duck Image
Taichung, Taiwan
http://www.duckimage.com.tw/
e-mail: steve@duckimage.com.tw

***Abstract*—Public Relation (PR) is an under-valued management tool for the small and median businesses (SMEs); especially, when PR is supported online. To many SMEs, PR is another form of advertising while others dismiss it as dealing with journalists and sending out press releases. However, PR is more than these. As the information technology becomes popular and mature, many communities like Facebook, Twitter and Line, play important roles in PR for fast communications in order to launch international cooperation. The traditional SMEs face the biggest challenge from the product innovation and the market circuit; interestingly, their collaborated ability often comes from the PR. Our research group including the Taiwanese and Polish professionals expects to assist the SMEs to enhance the competitiveness and international exposure by establishing the international platform for SMEs: E-Business Collaboration Club (EBCC). This research is supported by Ministry of Science and Technology (MOST) of Taiwan, Polish Academy of Science (PAS). The mutual cooperation leads to our initial success of SMEs to form social knowledge groups/communities such that we can observe the SME behavior in the platform. In addition, an actual bike company is used to validate the value of EBCC.**

Keywords- Small and Medium Enterprise (SME); E-Business; Information Technology; Public Relation; Collaboration.

I. INTRODUCTION

Public Relations (PRs) are concerned with improving mutual understanding within an organization as well as between two organizations [1]. It directs a deliberate effort towards improving communication between the people and organizations to broaden its sphere of influence through appropriate advertising, publicity and other form of communication to create a good impression [3][4]. As the competition becomes severe among small and medium businesses (SMEs), and they often lack the innovation capacity and the marketing channels: setting up a collaborated platform for promoting PR of SMEs seems to be a good idea [10][11][12]. Although many scholars mentioned the value of cooperation among SMEs, they seldom talk about and practically observe the relationship between PR and collaboration. This study is focused on using Information Technology (IT) as a power tool to set up the E-Business Collaboration Club (EBCC) for SMEs by the following new features: (a) the company of industrial design: Duck Image is the supporter and integrator of the platform, this platform integrates three main resources and encourages mutual dialogue: SMEs which need new product design, the design company, and the professors in Da-Yeh University, and (b) we try to practically observe that if a poor PR leads to the low collaborated ability of SME, and (c) we conduct the experiment for cloud prototyping in this study, which means we accept the orders of new product design online, and use EBCC intensively for online discussion and 3D-printing [13] to produce the product prototype for SMEs.

Management Master Prof. Drucker in 2001 [6] pointed out that, the business competition already changes from management of the traditional tangible assets: human affairs, to management of knowledge (intangible asset). In addition, the new knowledge: the experience or the technology is tacitly stored in staff's brains. Once the staff leaves the company, this results in draining or losing this precious knowledge. Now the development of the social group/community website as a result, online discussion either the exchange of new ideas has become simpler than before. We attempt to establish the EBCC here in order to help Taiwan and Poland's SMEs to share knowledge, to exchange information and to develop new products rapidly. The international cooperation project of two year grant is supported by Ministry of Science and Technology (MOST) and Polish Academy of Science (PAS): the first year of 2013 was focused on setting up the enterprise cooperation platform. The second year of 2014 is devoted to inviting all the partners to join in; thus, we have been able to analyze and appraise the knowledge sharing process in reality by these registered members.

In order to help the SMEs discussing the innovation concept completely, launch the new product development rapidly, and enter the market in time: the product innovation process should not consist of unplanned/random operations; on the contrary, it should be a sophisticated process from design to market. If we can launch the new ideas rapidly into product prototype, then the competitiveness of the SMEs will be greatly promoted. The design process includes the function, the outlook, the color, the material utilization, and fast prototyping of new products. In this study, each registered enterprise may express/show themselves with advantageous services/products by vivid pictures or videos

 ISBN: 978-1-61208-422-0

via EBCC (explicit knowledge), and seek for the collaboration and dialogue internationally (tacit knowledge). This EBCC is different from the traditional non-vivid, non-interactive websites.

The paper is arranged as follows: in Section II, we briefly introduce the PR and IT. In Section III, the detailed content of EBCC is presented. In Section IV, an actual observation and an experimental example are proposed to show how EBCC works, and validate its value in reality. Finally, the conclusion is available in Section V.

II. PUBLIC RELATION AND INFORMATION TECHNOLOGY

PRs is much more than advertising [2]. It is the practice of managing the delivery of information between two individuals or two organizations [8]. Public relations may include an organization or individual gaining exposure to their audiences using topics of public interest and news items that do not require direct payment [9]. This differentiates it from advertising as a form of marketing communications. The aim of public relations is to inform the public, prospective customers, investors, partners, employees, and other stakeholders and ultimately persuade them have a good view fort the organization, its leadership, products, or of business decisions. According to Daramola in 2003 [7], the purpose of public relations is "to create goodwill, understanding and awareness…of an organization or institution by using the PR techniques of persuasion, information and education to project the organization to its public…".

Professionals of PR serve in government organizations as well as private sector as a significant field communication. Serving its role of management function within and outside these organizations in order to meet the demands of the rapidly changing world, PR needs to employ an effective tool: information technology, that will enable it to achieve its objectives, and spread its messages/information globally. Therefore, application of information technology is a good support of PR [5].

Information Technology defined by the Information Technology Association of America (ITAA) is: design, development, implementation, support or management of computer-based information systems, particularly software applications and computer hardware. It deals with the use of electronic computers to convert, store, protect, process, transmit and securely retrieve information. Nowadays, numerous users rely on Facebook, Twitter, and Line, etc., supported by information technology, to achieve fast communication online. The individual impressions are now well spread by the social communities above, but the SMEs still lack a good tool for international dialogue. The gap between the SMEs which are able to or not able to effectively access the information technology and international resources, will dominate the successes or failure of SMEs. Furthermore, those SMEs without or with limited access to information technology are referred to "digital divide." It includes the lack of resources and expertise needed to effectively utilize the available information technology.

In this study, we use EBCC to improve the collaboration and PR of SMEs.

III. CONTENT OF EBCC

Our experimental idea is simple. We set up an IT platform: EBCC for promoting the PR, and observe the behavior of SME and validate the value of EBCC. That is, we see how they cooperate together. In this study, we simply define the PR of a SME is its used history in EBCC, which could be quantitatively tracked by registered members.

Since the application of information technology is so important today, using such a technology to empower the PR is valuable [2][5]. The project achievements by our efforts comprise:

A. Establishment of Information Technology (IT) platform

This is set up for the interior/exterior knowledge management, which provides the internal discussion and idea exchange for registered members; e.g., the blog, the new knowledge sharing, the talented person recruits, the immediate messenger in order to promote the team work efficiency and to deliver the creativity, thoughts and new product files, etc. Furthermore, we also provide the space of show case for each registered member such that interested customer can track the latest news of SME easily.

B. Establishment of the exterior IT platform for promotion of services/products of SMEs.

Our platform may provide the demonstration spaces for SMEs, and they may interact with exterior customers directly. We define this part as the marketing knowledge, which is explicit.

C. Establishment of the interior IT platform for idea exchanging of SMEs.

This platform preserves the dialogue history among SMEs when they discuss the new product development. Real time messages with sketches, pictures or files could be intensively exchanged here. In addition, we also provide the service of cloud prototyping by 3D printers. We define this part as the collaborated knowledge, which is tacit.

Figure 1. Portal Entrance of SMEs for EBCC [14]

The service goal of EBCC enables the SMEs to have the primary IT competition power as soon as possible such that their impressions could be improved. The portal entrance is shown in Figure 1. On the top area of the webpage pictures

 ISBN: 978-1-61208-422-0

in Figure 1, which may be automatically changed for the registered members for multimedia (video, pictures) of news, this area is available for the browsing of multi-pictures within one webpage.

This platform could be extended by general computer languages. The platform is divided into the following parts for the development:

A. *Discussion Space*

The internal discussion space only allows registered companies to propose their own preliminary ideas, discuss, and exchange ideas with each other. This part could be open publicly or privately by user option. The discussion is set up by different industries via blogs; for example, bike blog, electrical blog, agriculture blog, ribbon blog, mechanical blogs, etc. which is shown in Figure 2.

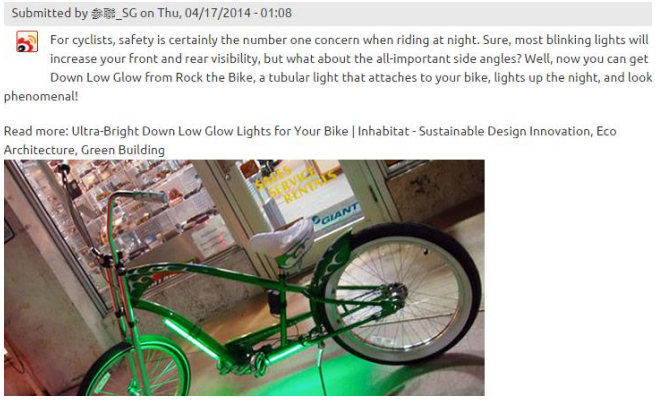


Figure 2. Individual Forum for each Enterprise [14]

B. *Sharing Space*

The sharing space is designed for industrial designers in order to show their crafts/sketches, these designers may have a demonstration space.

Figure 3. News of SMEs in EBCC [14]

This part is mainly used to release the latest issues for each registered member, which is shown in Figure 3. In addition, the latest news could be easily spread by sharing buttons of social networks.

C. *Individual Space*

The individual space allows the individual designer to work independently and privately. We provide the cloud prototyping service online. Simply speaking, we accept the orders by designers or SMEs for prototyping. This means we integrate the manufacturing capacity with the designer's idea in order to practically propose the prototype/service in time for SMEs. This service is shown in Figure 4.

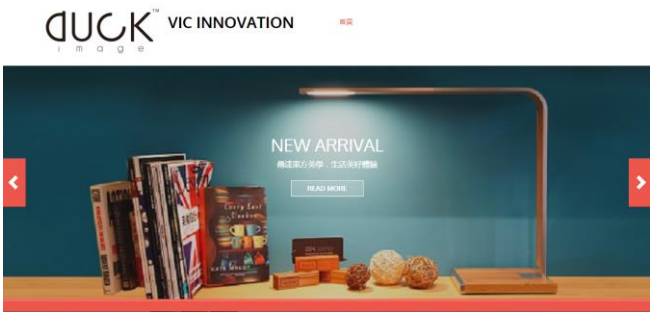


Figure 4. Individual Service in EBCC platform [14]

We induct the cloud concept in the system structure, the private cloud represents individual/private space; the social group cloud represents the discussion space; the blending cloud represents the sharing space to the public. The sharing space is similar to social groups/communities, such as well known models of Facebook or YouTube. The success of Facebook and YouTube comes from the user's self-awareness and the desire of personal showcase.

IV. EXAMPLE

We had collected about thirty registered members in EBCC, such that we can observe how they cooperate together by their explicit image (PR). We found that if a SME is good at PR by EBCC, then it is easy to get more supports/collaboration in EBCC. The PR here is simply defined as the tracking number of the specified SME in EBCC. This behavior is somewhat similar to the phenomena in Facebook.

Figure 5. Bike Frame Example

We also demonstrate a bike example in Figure 5 here to show the EBCC value from design to market. This company now enters the platform to seek a new design for a bike frame, and expects to see the frame prototype within 2 months resulting from the European Bike Exhibition.

First of all, we accept the order online, and clarify that this bike company has the time pressure to finish the job within two months.

Second, the Duck Image Company joins to prepare the sketches for candidates. An example of this process is shown in Figure 6. The selection process considers multiple attributes among five candidates; for example, outlook, strength, the time of 3D-printing (time to market), etc. At the same time, the bike company and the design company communicate mutually and intensively online via EBCC. The discussion content mainly focuses on the issues of progress and strength of this new product.

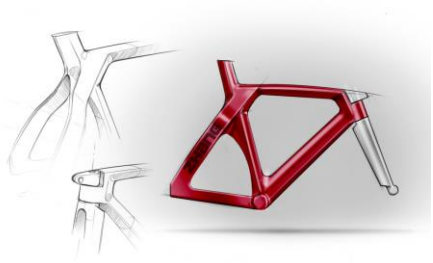


Figure 6. The Sketch of Bike Frame

Third, once the sketch is chosen, professors in Da-Yeh University receive the final sketch with dimensions online, and use the software: SOLIDWORKS[15] and ANSYS[16] to construct the 3D-model, and analyze the frame to see if it meets the European safety requirements. After it passes the requirements, professors separated the 3D frame model into many small parts for 3D-printing. The 3D printer and the research team are shown in Figure 7.

Figure 7. The Research Team and 3D Printer

Finally, we join all the small parts from 3D printing to form the whole frame, and send the new product (prototype) back to the bike company. The company received new orders because of the fast-manufacturing prototype via EBCC. The prototype is shown in Figure 8. The bike company and the design company are both satisfied with the test run via EBCC.

To summarize, we evaluate the EBCC that has the following new features which facilitate the success of cloud prototyping for the bike company. First, the knowledge management system within EBCC is quite huge, which contains a series of steps: the knowledge collection, the reorganization, the dispersion, the application, the renewal, the creation of values, and so on. Taking the bike company as an example, we may collect the bike design cases via internet: these data include the image, the writing or the multimedia. And these data are stored in the interior databases for reference by the registered members.

Figure 8 The Printed Bike Frame

Second, registered members are able to read the interior references. After that, in view of the new product development, they may discuss everything in the platform by the forum. In addition, if the discussion process is stopped for a short run, we allow the user to save the wisdom from discussion to the platform for internal databases, the discussers may continue the discussion in the next time by the support above. This function is just like the discussion history in Facebook.

Third, in the exterior platform, each SME is able to summarize, arrange and show their past history, present development, and the future vision one by one, and is able to contact the external customer by gathering fans via EBCC. The feedbacks from exterior customers may enter the interior social group/community for additional discussion of new products or services.

Finally, we had already and successfully connected the on-line 3D files to the 3D printer, which means if we receive the 3D data online, then we are able to control the 3D printer online. Furthermore, the company delivering orders is also able to track the progress online by project management.

V. CONCLUSION

The most common challenging problems for SMEs are the innovation capacity and the marketing channels. Our approach is not different from the exited approaches of social communities; on the contrary, we integrate the existed approaches and provide a simple and integrated solution for SMEs. Interestingly, we observe that a SME lacks the collaboration ability resulting from the poor PR in EBCC. Simply speaking, if a company PR is poor, then it is not easy to find a good partner to cooperate with it, and vice versa.

 ISBN: 978-1-61208-422-0

Since the IT power dominates the PR of SMEs today, each SME should do it best to show itself and to seek the supports/resources internationally for survival.

The poor PR could come from various aspects: this SME may not update its news rapidly, this SME may not be enthusiastic to use IT for PR, this SME could be afraid of the secret exposed to its competitor, etc. We need further study on this issue later.

The challenging issue of EBCC is the business secret security, which is left for resolution. In the phase of business secret management, the traditional method is secured by the paper copies, which can be read with interior approval of company in tradition, but the shortcoming problems are on the retrieval of complexity, large consumption of man-power for data management (big data), and the data search is not easy. Although we had successfully set up the IT framework; however, EBCC still needs more protections for business secret security. If more products/services and various members could be displayed and invited into EBCC, this platform could be expanded for PR of SMEs. We now focus on the local partners in Taiwan and Poland for customization. In the long run, we hope the EBCC could attract more international members to join in, and our final goal is encourage SMEs stand locally and win globally. Since an actual B2B platform to encourage international cooperation is still not easy to be observed nowadays; thus, we think the experiments here are valuable.

ACKNOWLEDGMENT

We appreciate the supports from the University of Economics in Katowice：Dr. hab. Jerzy MICHNIK, Prof. dr hab. Jerzy GOŁUCHOWSKI, Dr. hab. Krzysztof KANIA, Dr. Barbara FILIPCZYK and Dr. Tomasz STAŚ.

REFERENCES

[1] A. Theaker, The Public Relations Handbook, Psychology Press, 2004.

[2] C. H. Botan and V.Hazleton, Public Relations Theory II, Routledge, 2006.

[3] L. Curtis, C. Edwards, K. L. Fraser, S. Gudelsky, J. Holmquist, K. Thornton, and K. D. Sweetser, "Adoption of social media for public relations by nonprofit organizations," Public Relations Review, vol. 36, iss. 1, 2010, pp. 90–92.

[4] M. A. Johnson, "Public Relations and Technology: Practitioner Perspectives," Journal of Public Relations Research, vol. 9, iss. 3, 1997, pp. 213-236.

[5] V. Lance, L.V. Porter, and L. M. Sallot, "The Internet and Public Relations: Investigating Practitioners' Roles and World Wide Web Use," Journalism & Mass Communication Quarterly, vol. 80, no. 3, 2003, pp. 603-622.

[6] P. F. Drucker, Management Challenges for the 21st Century, Harpercollins, 2001.

[7] I. Daramola, Introduction to Mass Communication, Rothan Press Lagos, 2003.

[8] J. E. Grunig and T. Hunt, Managing Public Relations (6th ed.), Orlando, FL: Harcourt Brace Jovanovich, 1984.

[9] F. P. Seitel, The Practice of Public Relations. (10th ed.), Upper Saddle River, NJ: Pearson Prentice Hall, 2007.

[10] R. Vidgen, D. Francis, P. Powell, and M. Woerndl, "Web service business transformation: collaborative commerce opportunities in SMEs," Journal of Enterprise Information Management, vol. 17, iss. 5, 2004, pp. 372 – 381.

[11] L. M. Camarinha-Matos, "Collaborative networked organizations: Status and trends in manufacturing," Annual Reviews in Control, vol. 33, 2009, pp. 199–208.

[12] A. Sharma, "Collaborative product innovation: integrating elements of CPI via PLM framework," Computer-Aided Design, vol. 37, iss. 13, 2005, pp. 1425–1434.

[13] T. Rayna and L. Striukova, "The Impact of 3D Printing Technologies on Business Model Innovation," Digital Enterprise Design & Management, Advances in Intelligent Systems and Computing vol. 261, 2014, pp. 119-132.

[14] EBCC, http://iem.ietm.dyu.edu.tw/drupal/, accessed July 2015.

[15] SOLIDWORKS, http://www.solidworks.com/, accessed June 2015.

[16] ANSYS, http://www.ansys.com/, accessed June 2015.

 ISBN: 978-1-61208-422-0

Profiling Users in the Smart Grid

Carl Chalmers, William Hurst, Michael Mackay, Paul Fergus
School of Computing and Mathematical Sciences,
Liverpool John Moores University,
Byrom Street
Liverpool, UK
Email: C.Chalmers@ljmu.ac.uk, W.Hurst@ljmu.ac.uk, M.I.Mackay@ljmu.ac.uk, P.Fergus@ljmu.ac.uk

***Abstract*— The implementation of the smart grid brings with it many new components that are fundamentally different to traditional power grid infrastructures. The most important addition brought by the smart grid is the application of the Advanced Metering Infrastructure (AMI). As part of the AMI, the smart meter device provides real time energy usage about the consumer to all of the smart grid's stakeholders. Detailed statistics about a consumer's energy usage can be accessed by the end user, utility companies and other parties. The problem, however, is in how to analyse, present and make best use of the data. This paper focuses on the data collected from the smart grid and how it can be used to detect abnormal user behaviour for energy monitoring applications. The proposed system employs a data classification technique to identify irregular energy usage in patterns generated by smart meters. The results show that it is possible to detect abnormal behaviour with an overall accuracy of 99.45% with 0.100 for sensitivity, 0.989 for specificity and an error of 0.006 using the Linear Discriminant (LDC) classifier.**

Keywords*— *Smart Meter, Profiling, Advance Metering Infrastructure, Data Classification, Critical Infrastructure.

I. INTRODUCTION

Smart meters are one of the most significant components of the smart grid [1] and they are seen as the foundation of any future smart electricity network. Their introduction brings with it the ability for consumers to accurately monitor energy usage in real-time. By showing consumers their current energy usage, along with its associated cost, a more informed decision can be taken on when electricity consumption would be at its most cost effective. In addition, smart meters provide the gateway and monitoring for allowing consumers to generate their own electricity, selling any excess electricity back into the smart grid. This removes the traditional top down distribution, implementing a bidirectional energy flow [2]. The result is a bidirectional communication and power exchange between suppliers and consumers, transforming the traditionally passive end-users into active players.

The deployment and use of the smart meter generates large amounts of valuable data. For example, the amount of energy being consumed at set times and intervals. Additionally, smart meters can securely communicate many different real-time consumption values including: voltage, phase angle, frequency and any home generated energy. This produces accurate information about the amount of energy used so customers can be better advised on billing and consumption. In addition, it allows utility companies to plan future energy requirements based on past usage trends. This is achieved by analysing the data and identifying reoccurring patters of usage. As the roll-out proceeds, devices, including smart appliances, can be automatically controlled in order to better balance grid demand.

Smart meters have already been trialled in a number of countries, such as the USA, Australia, Netherlands, Italy and the UK, with future planned expansion [3]. In 2008, less than 4% of the electricity meters in the world were smart meters. By 2012, the percentage had grown to over 18% and it is expected to rise to 55% by the end of 2020.

In this paper, we present a case study into the data generated by three households. The households where selected from 78,720 individual smart meters in Australia. Using the data an approach for profiling consumers is put forward. We address how the development of a system, using advanced data analysis techniques, can be employed to assess consumers and detect abnormal behaviour. A discussion is also put forward on the benefits and challenges that smart meters bring with their implementation. The remainder of the paper is as follows. Section 2 presents a background on the Advanced Metering Infrastructure (AMI) and smart meters, section 3 defines the data collected from our smart meter case study. Section 4 discusses the methodology and techniques used for profiling users. The paper is concluded in Section 5.

II. BACKGROUND RESEARCH

The implementation of the smart grid has brought with it significant developments in technology such as the ability to remote read meters, allow bidirectional generation and distribution etc. Its introduction has enriched the way in which electricity usage data is produced and collected. This comprehensive data access is fundamentally different when compared to traditional grid infrastructures [4] and, for the first time, operators are able to gather highly detailed information about how their individual customers use electricity in real-time [5]. In this section, the focus is on the technologies which enables access to the data.

A. Advanced Metering Infrastructure

The AMI offers bidirectional communication between the consumer and the rest of the smart grid stake holders and replaces the traditional need for energy usage readings to be

 ISBN: 978-1-61208-422-0

collected manually [6]. There are a number of advantages associated with Automated Meter Readings (AMR) these include:

- Reduced costs for meter readings.
- The possibility to access meters otherwise difficult to attend due to position or security reasons.
- Support for real-time pricing.
- Increased fraud detection.
- Reduced read-to-bill time allowing utility companies to learn more about consumer power consumption.
- More informed choices about energy usage based through the use of in home displays and smart devices [7].

Specifically, the AMI can be broken down into three main areas: The Home Area Network (HAN), Wide Area Network (WAN) and the utility companies. Figure 1 shows the interaction between the different components that comprise the AMI. Each of the layers is subsequently explained.

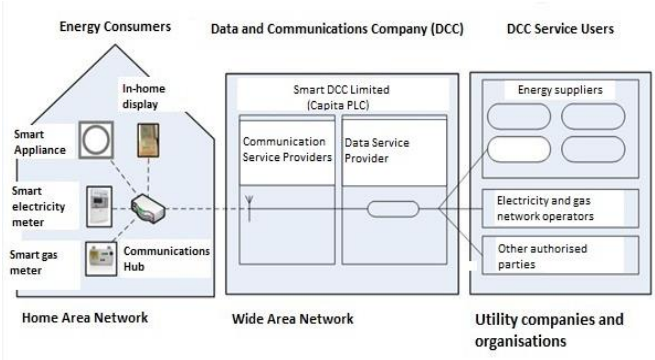


Figure 1. End To End Smart Meter Infrastructure

The Home Area Network (HAN) is housed inside the consumer's premise and is made up of a collection of devices. The in-home display unit (IHD), is the most visible and accessible part of the AMI. It provides the consumer with up-to-date information on electricity usage, as well as the units of energy being consumed. Secondly, the smart meter provides real-time energy usage to both the consumer and all of the stakeholders. This builds detailed energy usage profiles of its consumers. Smart appliances also react to peak events within the smart grid eco system. Adapting to these events enables a reduction in demand on the grid and energy costs for the consumer [8]. These smart appliances can effectively contribute to load management in future energy systems.

The Wide Area Network (WAN) handles the communication between the HAN and the utility companies. The sending of polled meter data to the utility, using a robust backhaul network such as: Ethernet, GSM, CDMA or 3G, is controlled by the WAN.

Organisations and utility companies have access to the data for analysis purposes this may include energy suppliers or energy networks. Enabling them to have detailed and accurate information. This improves management and planning of activities. In addition, grid reliance and performance is enhanced, as demand can be better provisioned.

B. Smart Meter Benefits

As previously discussed, the smart meter is an advancement on existing traditional meters [9]. Many of the advantages outlined below can be attributed to smart metering. These include lower metering costs; energy savings for customers; more reliability of supply through detailed monitoring and easier detection of supply problems and fraud. Its communication and data gathering capacities offer vast improvements and advancements over the traditional meters.

Electrical data, such as voltage, frequency and energy consumption information is recorded in real-time and reported every 30 minutes. Additionally, consumers are able to interact with their smart meters in order to interpret the data that is collected. Overall, smart meters can perform a wide variety of tasks including:

- Accurately record and store information for defined time periods (to a minimum of 30 minutes). This enables remote, accurate meter-readings with no need for estimates [10]. As these meters become more sophisticated, they are able to measure household power consumption at ever finer time-scales.
- Offer two way communications to and from the meter so that, for example, suppliers can read meters and update tariffs remotely [11].
- Allow customers to collect and use consumption data by creating a home area network to which they can securely connect data access devices [12].
- Enable other devices to be linked to the home area network, allowing customers to improve their control of energy consumption [13].
- Support time-of-use tariffs, under which the price varies depending on the time of day at which electricity is used [14]. Energy prices are more expensive during peak times. Billing consumers by time, as well as usage, will encourage them to change their consumption habits.
- Support future management of energy supply to help distribution companies manage supply and demand across their networks [15]. This is achieved automatically through previously agreed Demand Response (DR) actions.
- Allow remote enabling/disabling of supply by energy suppliers [16].
- Measure electricity exported from micro generation equipment to the network [17]. Having sensors capable of measuring a multitude of consumption generation parameters, rewards the consumer for adopting green technologies.
- Communicate with micro generation, home appliances and equipment within the property [18]. Smart meters will be able to control smart home appliances and communicate with other smart meters within reach. This

 ISBN: 978-1-61208-422-0

allows devices to be switched on when grid demand is low and turned off when demand is high.

The implementation of smart meters undoubtedly brings many benefits. However, there are also many challenges which need to be addressed.

C. New Technology Challenges

One of the main concerns is for the privacy of the consumer. Smart meters enable detailed profiling of consumers' energy usage and household activities. Patterns can be identified without prior knowledge of the consumer. It is possible to extract detailed usage patterns and consumer habits from the data as readings are at granular intervals.

Work undertaken by Andr´es [20] highlights the privacy concerns associated with smart meters. Their research demonstrates how complex usage patterns, can be obtained from smart meter data using off-the-shelf statistical methods. Specifically, their investigations focus on identifying trends in energy usage. However, none of the data was taken from actual smart meters. While energy readings were taken every second, in order to identify usage patterns, this does not accurately reflect real world smart meter usage. Typically smart meters report energy usage at 30 minute intervals. This is unlikely to change to due to the size of the data that is generated.

Storing, processing and analysing all of the data generated from smart meters, and the wider smart grid is also a challenge, due to its size and complexity. This is due to the variety of data and parameters from consumer usage to power generation some of which are shown in table 1.

TABLE 1. SMART METER DATA PARAMETERS

Reading	Description
Generated interval data kW	Half hourly interval held on meter for 13 months – average kW demand over half hour period.
Generated Kilovolt-Ampere-Reactance (kVAr)	Reactive power measurement in half hourly interval held on meter for 13 months – average kVAr demand over half hour period.
Generation Technology Type	e.g. Solar PV, micro CHP, wind, hydro, Anaerobic Digestion.
Import demand kW	Load being drawn from grid.
Export kW:	KW being exported to grid.
Total consumption today (kWh)	Import + Generated –Export.
Cost of energy imported (£/hr) and £ today	Net cost of imported energy less value of exported energy. Pushed to the IHD via SMS for the consumer.

The implementation of the smart grid represents a huge technical challenge. Aspects such as, networking, security, communication and data management require careful planning. For example, the integration of Supervisory Control and Data Acquisition (SCADA) networks with other commercial networks has made control systems vulnerable to various threats. Possible refinement, or creation of new and existing technology standards, has now become a necessity [21].

D. Profiling Consumers

Creating detailed energy profiles with associated data has a variety of benefits to the grid stakeholders. Some of which include: predicting future energy requirements based on historical data; establishing a detailed correlation between energy usage, weather conditions and social events; the refinement of more accurate demand and response systems using historical data; anomaly detection within the smart grid and, as our research focusses on, the detection of abnormal user behaviour.

By observing and analysing readings from the AMI and smart meters it is possible to detect anomalies, cyber-intrusions and factors that affect energy distribution. This requires the incorporation of data sets which include: weather/environmental; household demographics; social events; home generation; distribution of Plug-In Hybrid Electric Vehicles and home plug readings data.

E. Discussion and Summary

There are a number of large scale smart meter deployments in various countries. These include: France, Italy, Netherlands, Norway, Australia and the United States. Each of these counties have different policies and regulations governing their smart meter implementation programs. Such as: if smart meters are voluntary or compulsory; the frequency of readings and whether they have full function of automatic meter readings. The United States has one of the largest smart meter deployments [22]. In 2012, 533 U.S. electric utilities had 43,165,185 AMI installations. About 89% were residential customer installations. In the UK, the objective is to install smart meters for both gas and electricity by the end of 2020. The UK government estimates that the installation of smart meters will provide £6.2 billion net benefits to the United Kingdom.

The detail and granularity of the data collected can be used to address many of current and future challenges faced by the grid. One of the main challenges is being able to meet future energy demands in an efficient and environmentally safe way. The International Energy Agency expects worldwide energy demands to increase at an annual rate of 2.2 percent, eventually doubling the global energy demand. Analysing historical data can help plan the provision of future energy needs. It is noteworthy that the introduction of new green technologies, such as plugin electric vehicles, will put additional strain on the grid [23].

There has been little research effort to investigate the benefits of user profiling in the smart grid. The level of granularity in the datasets allows for accurate modelling and prediction of an individual's behaviour. This is beneficial both for predicting how the power-grid can accommodate the integration of new technologies, such as plug-in vehicles or assessing an individual's well-being.

 ISBN: 978-1-61208-422-0

III. APPROACH

By analysing the rich dataset provided by smart meters, the research in the paper aims to identify patterns and behaviours which can be used to profile users; predict future energy requirements; detect faults in the grid and identify compromised meters. In this section, a system framework which is able to process the significantly large datasets generated by smart meters is presented. A case study into three different consumers, and a demonstration on how an individual's behaviour can be profiled through their electricity usage, are also put forward.

A. System Framework

The proposed system shown in Figure 2 is adaptable and can be applied to a variety of functions some of which are shown in table 2.

TABLE 2. POTENTIAL SYSTEM FUNCTIONS.

Functions
Predicting future energy requirements based on historical data.
Establishing a detailed correlation between energy usage and weather conditions.
Aid in the detection of illegal activates through energy usage monitoring.
Accurately plan for the use of Plug-In Hybrid Electric Vehicles (PHEVs) based on historical gird load fluctuations.
Analyse what effects social events such as football matches, concerts etc. have on energy usage to better plan grid requirements for large events.

The system operates autonomously and identifies anomalies, energy usage, distribution patterns and trends. Figure 2 illustrates the proposed system along with the dynamics and correlations that the system monitors.

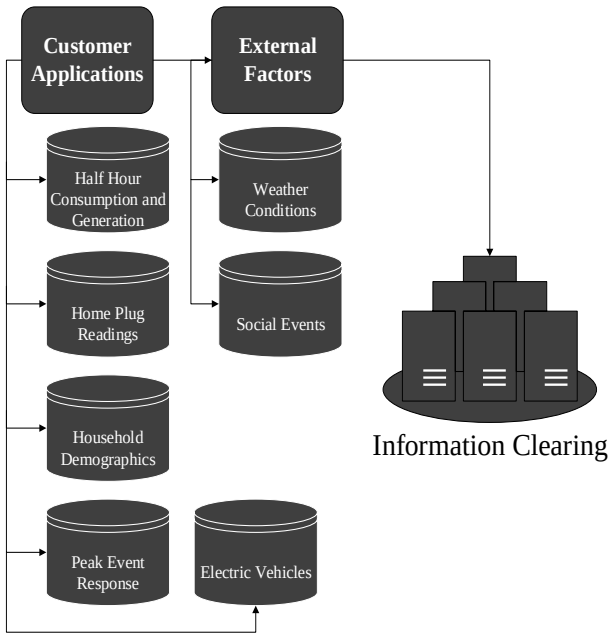


Figure 2. Proposed System Along With Associated Factors

The system provides detailed information about consumer consumption, which aids in the provisioning of future energy requirements and improving grid design and reliability.

B. Data Sample

As a demonstration of the functionality of the system, a case study is presented which focuses on an individual user taken from a large dataset during an extensive smart meter trial. Specifically, the data set generated by the trial contains smart meter readings taken from the 78,720 individual consumers over a six year period from 2008 to 2014.

For example, an individual user, selected at random from the data set, has 52560 rows of data. A typical sample of the data collected over a three hour period is displayed in Table 3.

TABLE 3. INDIVIDUAL SMART METER DATA SAMPLE (IN KWH)

Time (05/03/2013)	General Supply	Off Peak	Total Usage
07:29	0.014	No Usage	0.014
07:59	0.064	No Usage	0.064
08:29	0.107	1.132	1.239
08:59	0.155	0.169	0.324
09:29	0.092	0.563	0.655

The data shows the amount of electricity being used for each 30 minute interval (starting on the hour or half-past the hour) at each service location in the customer trial. Power consumption is reported as either normal domestic load or controlled load (i.e. off-peak - switched in or out by network control).

IV. USER PROFILING

To highlight how this data can be used to profile individuals, in this section we present a case study of the behaviour of two individual smart meter users. By analysing the dataset it is possible to develop a pattern of behaviour over one year and profile the behaviour of an individual. To enrich the dataset, household demographics taken from national census records for each consumer in the trial are included as above. This includes information about working patterns, household income and electrical devices that reside in the premise. For the initial profiling, the focus is on one years' behaviour data.

A. Profiles

In order to present a case study into profiling users in the smart grid, two individuals are randomly selected from the dataset as a sample, one with a normal energy pattern and one with an abnormal pattern. However, both users are within the same energy usage range. The census records detail that there is one occupant living in each of the premises selected for analysis. The residents are also known to be absent from home during the day.

Figure 3 and Figure 4 show a scatter plot of the daily max energy readings over 365 days of two individual consumers.

 ISBN: 978-1-61208-422-0

The energy usage is displayed along the x-axis. The y-axis refers to the day of the year.

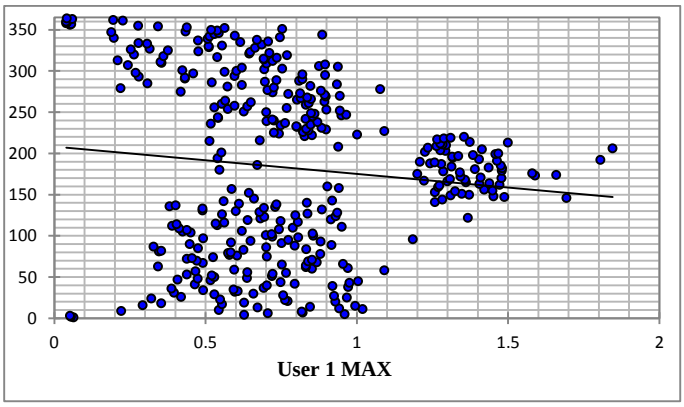


Figure 3. User 1 Max Energy Usage Kwh

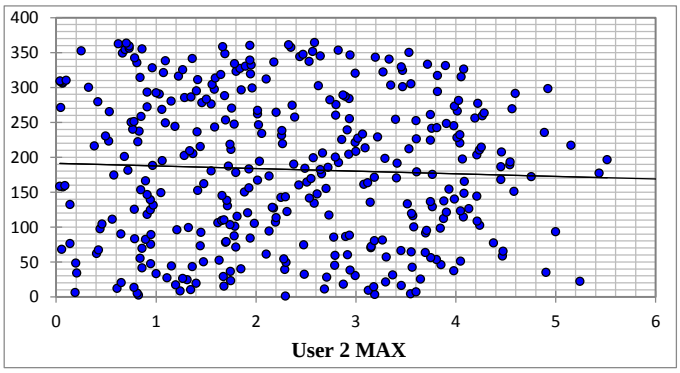


Figure 4. User 2 Max Energy Usage

The full profile of both consumers was evaluated over a one year period. This was undertaken to discover if any reoccurring patterns or habits could be identified based on their energy usage. Both figures also show the linear division of the data. As the division shows, the consumers have a tenancy to use more energy overall. Each consumer's energy readings were taken every half hour equating to 48 readings per 24 hour period, totalling 17520 individual readings per consumer per year. As shown in Figure 3 and Figure 4, clear divisions in the usage patterns can be clearly identified. Both households in this experiment have 1 resident who is out during the day.

B. Methodology

Figure 5 shows two users, comparing one with normal behaviour against a single individual expressing abnormal electricity usage patterns. Normal behaviour is represented by blue crosses, while abnormal is displayed as red dots. There is a clear visible difference in behaviour when comparing the usage.

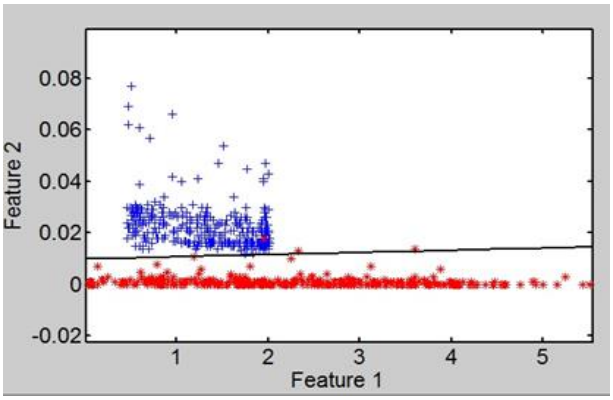


Figure 5. LDC Total usage per consumer over 1 year period

However, differences in behaviour are problematic to detect in the extensive datasets generated by smart meters. For this reason, our approach employs six different supervised machine learning classifiers: Uncorrelated Normal Density based Classifier (UDC), Quadratic Discriminant Classifier (QDC), Linear Discriminant Classifier (LDC), Polynomial Classifier (PLOYC), k-Nearest Neighbour (KNNC) and Support Vector Classifier (SVC). Each of these classifiers is chosen because they have the ability to learn how to recognise abnormal values in a dataset. They also employ a supervised learning approach, which is a key part of the system design. Table 4 presents the results of the classification.

TABLE 4. CLASSIFICATION RESULTS COMPARISON

Classifiers	AUC (%)	Sensitivity	Specificity	Error
LDC	99.45	0.100	0.989	0.006
UDC	98.90	0.100	0.978	0.011
QDC	98.90	0.100	0.978	0.011
SVC	78.29	0.100	0.565	0.217
POLYC	99.45	0.100	0.989	0.006
KNNC	65.93	0.598	0.719	0.341

LDC and UDC were the most accurate, with both classifying over 98.90% of the data accurately. The LDC, is able to categorize the dataset with high accuracy (99.176%) with an error rate of 0.006.

LDC is shown in Figure 5 where the classifier is able to predict 99.45% of the data accurately. The KNNC classifier as shown in Figure 6 was the least accurate with an accuracy of 65.93%. As the graph displays, samples of the data from the abnormal set (red dots) is misclassified inside the KNNC contour grouping.

 ISBN: 978-1-61208-422-0

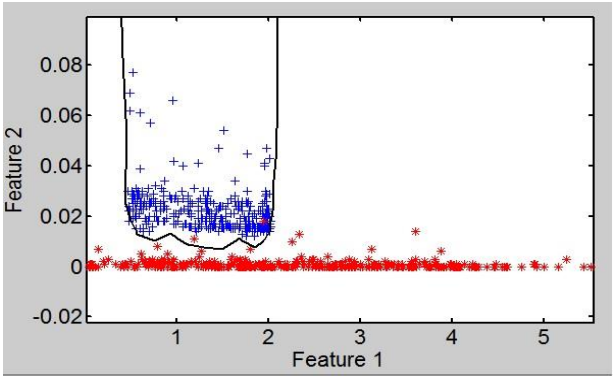


Figure 6. KNNC Total usage per consumer over 1 year period

C. Discussion

Despite selecting two users within the same energy usage range, clear deviations can be seen in the data sets. This change in usage patterns reflects an individual's unique behavioural characteristics. However, Figure 5 shows how the second consumer's behavioural patterns are not as clearly identifiable and there is no obvious visible trend in the usage pattern.

Profiling consumers with fine grained measurements brings many applications and benefits. For example, when comparing a third consumer's energy usage to the initial case study is it apparent that behavioural patterns are unique to the individual consumer. This data can therefore be analysed to give an accurate account of the customers' habits, characteristics and expenditure.

V. Conclusion and Future Work

The implementation of the AMI and, in particular, smart meters enables the analysis of energy usage with a high degree of accuracy and granularity. Being able to utilise the collected data brings countless benefits to the grid's stakeholders and consumers alike. In order to meet future energy needs and be able to incorporate home generation and distribution, it is clear that changes in the current grid model are needed. The smart grid addresses the constraints imposed by the current power infrastructure by allowing detailed grid monitoring and involving the consumer. Being able to collect and analyse sufficient amounts of usage data makes it possible to identify reoccurring patterns and trends which can be used to address gird problems by profiling users of current and future grid implementations.

The results have shown that it is possible to identify and categorise individual user patterns generated by smart meters. Using the classification techniques presented in this research, it is possible to establish both normal and abnormal consumer behaviour based on granular energy usage data. The classifiers achieved high results. However, the main errors in classification are seen in the misclassification of normal user behaviour (sensitivity). This results in the false positive identification of abnormal user behaviour. This will be addressed in the future work by including the inclusion of more case studies to expand the dataset. Our future work will involve incorporating additional datasets, including home plug readings showing how much energy each electrical device has used at 30 minute intervals. The use of the additional data will create a more detailed profile of a user and allow us to make accurate assumptions about an individual's behaviour patterns.

References

[1] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet and D. Irwin, "Private Memoirs of a Smart Meter," Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, 2010, pp 61-66.

[2] A. Mengolini and J. Vasiljevska, "The social dimension of Smart Grids," Joint Research Centre of the European Commission, 2013

[3] J. Zheng, D. Wenzhong Gao and L. Lin, "Smart Meters in Smart Grid: An Overview," Proceedings of IEEE Conference on Green Technologies Conference, 2013, Pages 57 -64.

[4] D. J. Kerrigan, L. Gamberini, A. Spagnolli and G. Jacucci, "Smart meters: A users' view," PsychNology Journal, 2011

[5] S. O. Hurtado and P. Parmenter, "The Green Grid Energy Savings and Carbon Emissions Reductions Enabled by a Smart Grid," Electric Power Research Institute 2008.

[6] M. Popa, "Data Collecting from Smart Meters in an Advanced Metering Infrastructure," Proceedings of 15th International Conference on Intelligent Engineering Systems, 2011

[7] A. Faruqui and L. Wood, IEE Releases: "The Benefits of Smart Meters," NARUC Winter Meetings. Institute for Electric Efficiency, Washington, D.C. 2011.

[8] R. Stamminger, "Synergy Potential of Smart Appliances, Project report within the Smart-A project," v2, November 2008

[9] R. van Gerwen., S. Jaarsma and R. Wilhite., "Smart Metering," IDC-Technologies, 2006

[10] F. Benzi, N. Anglani, E. Bassi and Lucia Frosini "Electricity Smart Meters Interfacing the Households," IEE transactions on Industrial Electronics, Vol 58, No10, 2011VOL. 58, NO. 10.

[11] C. Bennett, D. Highfill, "Networking AMI Smart Meters," November 2008, IEEE Energy2030.

[12] T. Krishnamurti, D. Schwartz, A. Davis, B. Fischhoff, W¨and i Bruine de Bruin, L. Lave and J. Wang, "Preparing for smart grid technologies: A behavioral decision research approach to understanding consumer expectations about smart meters," Energy Policyc41, 2011.

[13] M. Venables, "Smart meters make smart consumers," Engineering & Technology, Volume: 2, Issue: 4, 2007.

[14] L. Olmos, S. Reuster, S. J. Liong and J. M. Glachant, "Energy efficiency actions related to the rollout of smart meters for small consumers, application to the Austrian system," Energy Volume 36, Issue 7, July 2011.

[15] M. Anas, N. Javaid, A. Mahmood, S. M. Raza, U. Qasim and Z. A. Khan, "Minimizing Theft using Smart Meters in AMI," Seventh International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 2012.

[16] H. Farhangi, "The path of the smart grid," Power and Energy Magazine, IEEE (Volume:8 , Issue: 1), February 2010.

[17] J. Zheng., D. Wenzhong Gao and L. Lin., "Smart Meters in Smart Grid: An Overview," Proceedings of IEEE Conference on Green Technologies Conference, 2013, Pages 57 -64.

[18] S. S. Sreenadh, R. Depuru, L. Wang and V. Devabhaktuni, "Smart meters for power grid: Challenges, issues, advantages and status, Renewable and Sustainable Energy Reviews 15," February 2011.

[19] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private Memoirs of a Smart Meter," BuildSys '10, 2010.

[20] Á. MacDermott, Q. Shi, M. Merabti and K. Kifayat: "Intrusion Detection for Critical Infrastructure Protection," June 2009, Neural Networks, 2009. IJCNN 2009, Pages 14-19.

[21] U.S. Energy Information Administration, March, 2015 [Accessed online at: http://www.eia.gov/]

F. Bouhafs., M. Mackay and M. Merabti "Links to the Future," IEEE Power & Energy Magazine, 2012.

 ISBN: 978-1-61208-422-0

Towards Integrated Engineering of Adaptive Resilient Systems

Elena Troubitsyna
Åbo Akademi University,
Turku, Finland
e-mail: Elena.Troubitsyna@abo.fi

***Abstract*— Resilience is an ability of a system to deliver trustworthy services despite changes. It is a much sought after property in a wide range of applications. However, currently, development of resilient adaptive systems constitutes a major engineering challenge due to a diversity of methods and tools used in the development and a lack of support for efficient information engineering. In this paper, we discuss the challenges engineering resilient adaptive systems. We propose the Problem-Design-Exploration framework as a model of the adaptive service development process and define the key concepts supporting multi-view engineering. Moreover, we discuss the advantages of the Open Services for Lifecycle Collaboration (OSLC) as a technology enabling integrated information engineering for resilient adaptive systems.**

Keywords-resilience; adaptability; changes; evolution; integrated engineering environment.

I. INTRODUCTION

Resilience is an ability of a system to deliver trustworthy services despite changes [1]. It is a much sought after property in a wide range of applications. Resilience is an evolution of the dependability concept [2] that puts an emphasis on the ability of a system to adapt to changes. However, currently, development of resilient adaptive systems constitutes a major engineering challenge [3]. Firstly, the existing development methods are unable to efficiently and confidently cope with the overwhelming system complexity and deliver required assurance of system trustworthiness. Secondly, they do not provide efficient platform for integrating changes in the system design. Finally, they give a rather limited support for innovation and experimentation, i.e., do not allow the designers to assess the impact of changes on system behavior with high productivity and confidence.

In this paper, we discuss the challenges in creating an efficient environment for engineering resilient adaptive systems. We explore the challenges in adapting to changes of different nature and introduce the Problem-Design-Exploration framework [4] - [8] as a model for the development process of resilient adaptive systems. We propose the fitness criteria that can be used to assess how adaptation to a change impacts different resilience attributes.

Since resilience is a multi-facet characteristic, diverse engineering tools are used for such an assessment [1]. We discuss the problem of tool integration and demonstrate how Open Services for Lifecycle Collaboration framework [9] can facilitate creation of an integrated engineering environment.

We believe that the problems discussed in this paper constitute the important challenges in the area of adaptive resilient systems engineering.

The paper is structured as follows: in Section II, we introduce the concept of resilience. We show its connection with the dependability concept and discuss the role of changes. In Section III, we introduce the Problem-Design-Exploration framework and define the fitness criteria relevant for the design of resilient system. In Section IV, we discuss the tool integration problem. In Section V, we outline the benefit of OSLC as a technological enabler of the integrated development. Finally, in Section VI, we overview the related work and conclude.

II. RESILIENCE AND ADAPTABILITY

Resilience is an ability of the system to persistently deliver its services in a dependable way despite changes [1]. The concept of resilience is an evolution of the concept of dependability – a system property to deliver services that can be justifiably trusted [2]. Dependability is a multi-facet system characteristic that includes the following attributes:

- *availability* is the ability of the system to provide service at any given instance of time;
- *reliability* is the ability of the system to continuously provide correct service over a given period of time;
- *safety* is the ability of a system to deliver service under given conditions without catastrophic consequences to its user(s) or environment;
- *integrity* is the absence of improper system alteration;
- *maintainability* is the ability of a system to be restored to a state, in which it can deliver correct service;
- *confidentiality* is the absence of unauthorized disclosure of information.

Faults of different nature might jeopardizes dependability by propagating to the system or service interface level and, as a result, introduce undesirable deviations in service provisioning.

Engineering of resilient systems relies on four main techniques: *fault prevention*, *fault removal*, *fault forecasting* and *fault tolerance* [1] [2]. *Fault prevention* is a set of

 ISBN: 978-1-61208-422-0

techniques aimed at preventing introduction of faults during the development process. It relies on formal and structured techniques aiming at ensuring high quality of the system and spotting problems in the design before the system becomes operational.

Fault removal techniques are used to identify and remove errors in the system. The activities of fault removal process include system verification as well as corrective and preventive maintenance of the system. *Fault forecasting* aims at predicting and evaluating the impact of fault on the system behaviour. It might be performed qualitatively or quantitatively. The qualitative assessment aims at identifying and classifying failures as well as defining combinations of faults that may lead to a system failure. The quantitative analysis is performed to assess the degree of satisfaction of the different attributes of dependability.

Finally, *fault tolerance* techniques aim at ensuring that the system continues to deliver its services or behaves predictably even in presence of faults.

Several decades of research have resulted in creating a solid body of techniques for engineering dependable systems. Majority of these techniques rely on assumption of exhaustive knowledge of system and its environment behavior, i.e., are static by nature. However, currently it is widely recognized that changes are inevitable and hence, the systems should be able to adapt to them while remaining dependable, i.e., be resilient.

The changes with which the system should be able to cope might be external, i.e., in the operating environment of the system or internal ones. In general, the changes can be classified according to their character as follows:

- *nature:* functional, environmental or technological;
- *prospect:* foreseen, foreseeable, unforeseen (or drastic) changes;
- *timing:* short term (e.g., seconds to hours), medium term (e.g., hours to months) and long term changes (e.g., months to years).

The changes cause continuous system evolution. The evolutionary development approach is supported by agile development model, Scrum development approach as well as DevOps. All these models emphasis the need for iterative development and continuous experimentation with the system under construction. Therefore, engineering of adaptive resilient systems provide a powerful support for change management, continuous evolution and experimentation. That requires an integration of current approaches to engineering dependable systems into a highly dynamic engineering environment facilitating modelling, design and assessment of resilient systems as well as supporting a novel iterative model of development process.

In the next section, we propose the Problem-Design-Exploration framework as a model for the development of adaptive resilient systems and demonstrate how to tailor to address various aspects of resilience.

III. PROCESS OF ENGINEERING ADAPTIVE RESILIENT SYSTEMS

The Problem-Design Exploration Model [4] proposes to model design process as two interacting evolutionary domains – problem space *P* and solution space *S*, as shown in Figure 1. The clear distinction between problem and solution spaces is supported in analytical [5], empirical [6] and prescriptive [7] research. The problem space contains mental representations of the developer's interpretation of the requirements" and "the design space" contains mental representations of the developer's specific solutions [8].

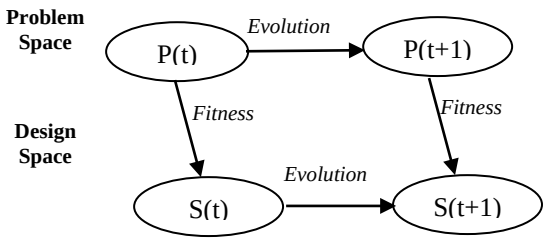


Figure 1. Problem-Design-Exploration Process.

The exploration process shown in Figure 1 has the following characteristics:

1. It is carried in two distinct search spaces: Problem Space and Design Space.
2. These state spaces interact in real time.
3. The horizontal movement represents is an evolutionary process such that
   a. Problem space P(t) evolves to P(t+1), P(t+2) etc.,
   b. Solution space S(t) evolves to S(t+1), S(t+2), etc.
4. The diagonal movement represents a process where goals lead to solution. It exists in two variants: *"Problem leads to Solution"* (downward arrow) or *"Solution refocusses the Problem"* (upward arrow).

The problem space P(t) is the design goal at time *t* and S(t) is the solution space, which defines the current space for the design solutions. The solution space S(t) provides not only a state space where a design solution can be found, but it also prompts new requirements for P(t+1), which were not in the original problem space, P(t). This is represented by the dashed upward arrow from design space S(t) to problem space P(t+1). The upward arrow is opposite: S(t) becomes the goal and a "*search*" is carried out in the problem space, P(t+1), for a "*solution*". This iterative relationship between problem space and design space evolves over time.

The Problem-Design-Evolution framework fits the main requirement for a development model of resilient adaptive systems because it explicitly supports evolution. Indeed, both the problem space and the solution space co-evolve simultaneously as a result of exploration. The basis for co-evolution is to consider the representation and application of

 ISBN: 978-1-61208-422-0

the fitness function so that the problem definition can change in response to the current solution space.

To tailor the Problem-Design-Evolution model to engineering of resilient adaptive systems, we need to understand how changes affect the main design objectives. The design objectives are defined by the dependability attributes. Therefore, we should devise the guidelines to be followed while assessing the impact of changes and creating a corresponding artefact in the design space. These guidelines are defined as the fitness criteria.

Table 1 presents the examples of the fitness criteria that might be evaluated while introducing changes of different nature in the design of resilient services.

The Problem-Design-Evolution paradigm provides us with a suitable general model of the process of adaptive system development. However, we also need to "zoom into" the development process and address the problem of integration. Indeed, a variety of methods and tools are used to achieve different design objectives. Therefore, to support the process of designing resilient adaptive services, we need to create an integrated development environment that establishes and information continuum between diverse methods and tools. The problem is traditionally addressed by the Application Lifecycle Management – the concept that we study next.

IV. INTEGRATED ENGINEERING ENVIRONMENT

Application Lifecycle Management (ALM) is a concept that aims at studying "The coordination of development lifecycle activities, including requirements, modelling, development, build and testing, through:

1. enforcement of processes that span these activities;
2. management of relationships between development artefacts used or produced by these activities; and
3. reporting on progress of the development effort as a whole" [10].

The term artifact broadly refers to any item (requirement, code, model, test case) produced during the development of software. ALM often seen as a concept that tries to syncronise all the lifecycle activities instead of focusing on any specific lifecycle activity [10].

The concept of ALM is still rather new and lacks well-established definition. In this paper, we focus on the technological aspects of ALM – the tool integration.

Tool integration is a rapidly growing interdisciplinary research area. It is a cross-road between Software engineering, Systems engineering, Human-Machine interactions and Economics. The tool integration discussion was originated in STONEMAN report [11] where among the other Buxton introduces the notion of integrating tools throughout a software project life-cycle.

The essence of tool integration was defined in the seminal paper by Wasserman [12]. He introduced the following 5 types of tool integration: Control, Data, Platform, Presentation and Process Integration.

Control Integration is the ability of tools to notify each other of events and activate each other under program control.

Data Integration is the ability of tools to share data with each other and manage the relationships among data objects produces by each other.

Platform Integration is a set of system services that provide network and operating systems transparency to tools and tool frameworks.

Presentation Integration refers to the set of services and guidelines that allow tools to achieve a common representation from the user's perspective.

Process integration defines linkage between tool usage and the software development process. Usually it tries to tie process integration to the definition and integration of process models.

From the ALM point of view, tool integration should therefore produce integrated environments that support the entire software development lifecycle. According to Pederson [13] an integrated environment allows the users easily move from one function to another without having to work with multiple, disconnected tools and manually integrate data between these tools.

It is easy to observe that the evolutionary aspect will result in creating various dynamically changing interdependencies and data. Now, we discuss the technological platform enabling creation and maintenance of such an integrated environment.

We can now define the requirements to integrated environment for engineering resilient adaptive systems as follows:

1. The integrated engineering environment should be non obtrusive and support heterogeneous design space.
2. The environment should allows the designers to continue to use their native tools regardless whether they are open source or proprietary
3. The environment should enable global traceability and querying of information by different engineering teams. The new information, introduced as a result of changes, should be easy to incorporate and link.

In the next section, we introduce OSLC and show that it satisfies the abovementioned criteria.

 ISBN: 978-1-61208-422-0

TABLE I. EXAMPLES OF FITNESS CRITERIA.

	Functional	Environmental	Technological
availability	Can new functionality result in an interruption of a service?	Can the system cope with peak loads?	How new platforms affect performance?
reliability	Does new functionality reduce the level of redundancy?	Can the system maintain reliable operation under the stress conditions?	Does the changed platform increase redundancy?
safety	Does new functionality expand safety kernel?	Which safety mechanisms are affected the by change?	Does new technological platform allow for the use of existing safety mechanisms?
integrity	Does new functionality requires weakening access policy?	Does the new environment introduce different data handling mechanisms/policy?	Does the new platform allows for the same degree of data protection?
maintainability	Can the relationships between the new and existing functions be properly documented and observed?	Which maintainability requirement will be introduced in the new environment?	How the existing maintainability routine will be affected?
confidentiality	Does new functionality increases openness of the system?	How can confidentiality be preserved if new access channels are introduced?	Does new platform introduced any additional vulnerabilities?

V. INTEGRATED INFORMATION ENGINEERING

OSLC [9] is an open community, whose main goal is to create specifications for integrating tools, their data and workflows in support of end-to-end lifecycle processes. OSLC is organised into workgroups that address integration scenarios for individual topics such a change management, test management, requirements management and configuration management. Such topics are called *OSLC domains*. Each workgroup explores integration scenarios for a given domain and specifies a common vocabulary for the lifecycle artefacts needed to support the scenarios. OSLC has received a notable industrial uptake.

Essentially, OSLC specifications focus on defining how the external resources of a particular tool can be accessed, browsed over, and specific change requests can be made.

OSLC does not aim at standardising the behaviour or capability of any tool. Instead, OSLC specifies a minimum amount of protocol and a small number of resource types to allow two different tools to work together in a collaborative way.

To ensure coherence and integration across these domains, each workgroup builds on the concepts and rules defined in the OSLC Core specification. The OSLC Core specifies the primary integration techniques for integrating lifecycle tools. This consists mostly of standard rules and patterns for using HTTP and RDF that all the domain workgroups must adopt in their specifications.

OSLC is based on the W3C Linked Data. The four rules of linked data introduced by Berners-Lee [14] are as follows:

- Use URIs as names for things.
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).
- Include links to other URIs, so that they can discover more things.

In OSLC, each artefact in the lifecycle -- for example, a requirement, defect, test case, source file, or development plan and so on -- is an HTTP resource that is manipulated using the standard methods of the HTTP specification (GET, PUT, POST, DELETE).

According to the third rule of linked data, each resource has an RDF representation. OSLC mandates RDF/XML,

 ISBN: 978-1-61208-422-0

which is the most widely adopted RDF notation, but can have representations in other formats, like JSON or HTML.

The OSLC Core specification defines a number of simple usage patterns of HTTP and RDF and a small number of resource types that help tools integrate and make the lifecycle work. The OSLC domain workgroups specify additional resource types specific to their lifecycle domain, but do not add new protocol.

OSLC defines the concept of ServiceProvider to allow applications to expose their external resources for integration scenarios. ServiceProviders answer two basic questions, which are:

- To which URLs should one POST to create new resources?
- Where can one GET a list of existing resources?

A ServiceProvider is intended to represent a "container" of resources that is hosted by a tool, not the tool itself. A single instance of a tool will typically host multiple ServiceProviders, for example one for each "project" or "product".

ServiceProvider is the central organising concept of OSLC, enabling tools to expose resources and allowing consumers to navigate to all of the resources, and create new ones.

Two fundamental properties of a ServiceProvider are given below:

- oslc:creation: the URL of a resource to which you can POST representations to create new resources.
- oslc:queryBase: the URL of a resource that you can GET to obtain a list of existing resources in the ServiceProvider.

ServiceProviders have a third important property -- dialog -- that is the foundation of the second major OSLC integration technique based on invocation of HTML web user interface dialogs of one tool by another.

There are three different approaches to implementing an OSLC provider for software:

- Native Support approach is to add OSLC support directly into the application, modifying whatever code is necessary to implement the corresponding OSLC specification.
- Plugin approach is add OSLC support to the application by developing code that plugs-in to the application and uses its add-on API.
- Adapter approach is to create new web application that acts as an OSLC Adapter, runs along-side of the application, provides OSLC support and "under the hood" makes calls to the application web APIs to create, retrieve, update and delete resources.

The Native approach allows tool vendors to add the OSLC support to their own products. The Plugin and Adapter approaches are best suited for adding OSLC support to the tools that have been bought from a tool vendor or obtained from an open source project.

Eclipse Lyo is an SDK to help the Eclipse community adopt OSLC specifications and build OSLC-compliant tools.

Lyo OSLC4J is a Java toolkit for building Open Services for Lifecycle Collaboration providers and consumers. It includes:

- annotations to decorate Java objects with OSLC attributes;
- annotations to assist with resource preview UIs;
- built-in support for service provider and resource shape documents;
- libraries to simplify service provider and consumer development;
- Tests for the sample applications to complement the Lyo OSLC Test Suite.

We argue that OSLC satisfies the criteria defined in Section V for an integrated engineering environment for designing resilient adaptive systems. Firstly, it is non-obtrusive because it does not enforce any standards on the engineering tools. Secondly, it allows the designers to continue to use their native tools and smoothly introduce the facilities for linked data. Thirdly, it support highly dynamic information creation and management via support of linked data.

VI. Related Work and Conclusions

Tool integration facilitates a productive development environment by allowing the user to launch tools and transfer information easily between different tools. Booch and Brown [15] introduced an interesting vision of a ‘frictionless surface’ provided by Collaborative Development Environments. They argue that such environments can remove the points of friction in the daily life of the developer that hinder effective operation. These friction points relate to issues, such as insufficient work product collaboration and problems maintaining effective group communication, including knowledge and experience, project status and project memory.

In his recent paper [16], Ralph introduces a further development of the Problem-Design-Exploration framework -- The Sensemaking-Coevolution-Implementation Theory of software design. In this theory, he aims at blending the boundaries between the problem and design space. It is an interesting theory that fits the novel trends in software development, such as DevOps. We are planning to investigate the use of this theory in the development of resilient adaptive systems in our future work.

The notion of resilience is a subject of active research discussions. Among the most prominent initiatives that contributed to defining the concept of resilience and taxonomy of related terms are the projects ReSIST [17] and Resilinets [18]. In our paper, we rely on the definitions introduced in these projects.

The concept of resilience addresses a wide variety of issues in system design [18]. Therefore, an integration various design methods and tools is especially interesting for resilient systems engineering. The problem of integration has been explored in the context of formal modelling and

 ISBN: 978-1-61208-422-0

verification of safety-critical and fault tolerant systems. In particular, [19], [20] and [21] address an integration of safety analysis into formal system model.

In this paper, we discussed the problems in establishing an integrated environment for engineering resilient adaptive services. We introduced the Problem-Design-Exploration framework as the model of engineering for resilience. The model is targeted towards supporting continuous experimentation and introducing changes in the design. We defined the fitness criteria, which serve as guidelines while assessing the impact of changes on resilience and devising the suitable design solution.

Resilience introduces multiple, sometimes conflicting objectives in service design. Since diverse methods and tools are used to achieve them, it is important to provide the designers with a powerful platform for integrated engineering. We discussed the problem of tool integration and identified the need to support dynamic data as the main requirement for the technological support. We argued that OSLC provide us with an adequate technological support for creating and integrated engineering environment and enables non-intrusive integration that supports experimentation.

ACKNOWLEDGMENT

This work is supported by the Finnish National Program Need for Speed http://www.n4s.fi/en/.

REFERENCES

[1] J. C. Laprie, “From Dependability to Resilience,” In 38th IEEE/IFIP Conference On Dependable Systems and Networks, IEEE Computer Press, pp. G8-G9, 2008.

[2] J. C. Laprie, Dependability: Basic Concepts and Terminology. New York, Springer-Verlag, 1991.

[3] Top Challenging Issues for Software Development. [Online]. Available from http://www.iaria.org/conferences2013/filesICSEA13/. 01.05.2015.

[4] M. Maher, J. Poon, and S. Boulanger, “Formalising design exploration as co-evolution: a combined gene approach”, Preprints of the Second IFIP WG5.2 Workshop on Advances in Formal Design Methods for CAD, 1995, pp. 1–28, doi=10.1.1.56.4459&rep=rep1&type=pdf.

[5] K. Dorst and N. Cross. “Creativity in the design process: co-evolution of problem–solution,” Design Studies, 22(5), pp. 425–437, Elsevier 2001.

[6] J. S. Gero and T. McNeill, “An approach to the analysis of design protocols”, Design Studies. 19 (1), pp. 21–61, Elsevier, 1998.

[7] P. Checkland, Systems Thinking, Systems Practice, Wiley, 1999.

[8] S. Purao, M. Rossi, and A. Bush, “Towards an understanding of problem and design spaces during object-oriented systems development”, Information and Oranisation, 12 (4) , pp. 249-281, Pergamon, 2002.

[9] OSLC: (Open Services for Lifecycle Collaboration.) [Online] Available from http://open-services.net/ 01.05.2015.

[10] C. Schwaber, “The Changing Face of Application Life-Cycle Management”, Forrester Research Inc., White Paper, August 2006. [Online] Available from www.serena.com /docs/repository/alm/changing-face-applic.pdf 01.05.2015.

[11] J. N. Buxton. “STONEMAN, Requirements for Ada Programming Support Environments,” Technical Report. Department of Defense. 1980.

[12] A. I. Wasserman, “Tool Integration in Software Engineering Environments”. In Software Engineering Environments: International Workshop on Environments, Chinon, France, September 1989, ISBN:3-540-53452-0.

[13] J. Pederson, “Creating a tool independent system engineering environment”, IEEE Aerospace Conference, March 2006.

[14] W3C web site. [Online] Available from http://www.w3.org/DesignIssues/LinkedData.html 01.05.2015.

[15] G. Booch and A. Brown, “Collaborative development environments”, Advances in Computers, Vol. 59, Academic Press. 2003. doi=10.1.1.84.3292

[16] P. Ralph, “The Sensemaking-Coevolution-Implementation Theory of software design,” Science of Computer Programming, v.101, pp.21-41, April 2015.

[17] ReSIST project. [Online]. Available from http://www.resist-noe.org/ 01.05.2015.

[18] RESILINETS project [Online]. Available from https://wiki.ittc.ku.edu/resilinets/ Accessed 01.05.2015.

[19] E. Troubitsyna, “Elicitation and Specification of Safety Requirements”, The Third International Conference on Systems (ICONS’08), IEEE Computer Society, April 2008, ISBN978-0-7695-3105-2.

[20] K. Sere, K. and E. Troubitsyna, “Safety Analysis in Formal Specification”, World Congress on Formal Methods 1999, LNCS 1709, pp. 1564-1583, Springer, 1999, ISBN:3-540-66588-9.

[21] I. Lopatkin, A. Iliasov, A. Romanovsky, Y. Prokhorova, and E. Troubitsyna, “Patterns for Representing FMEA in Formal Specification of Control Systems”. Proceedings of the 13th IEEE International High Assurance Systems Engineering Symposium (HASE 2011), pp. 146-151, IEEE, 2011, ISBN 978-1-4673-0107-7.

[22] A. Iliasov, A.Romanovsky, L.Laibinis, E.Troubitsyna, and T.Latvala, “Augmenting Event-B modelling with real-time verification,” FormSERA@ICSE 2012, pp. 51-57, IEEE Computer, 2012.

[23] E. Troubitsyna, “Reliability assessment through probabilistic refinement,” Nordic Journal of Computing 6 (3), 320-342, 1999.

 ISBN: 978-1-61208-422-0

Analyze OSPF Convergence Time in the Presence of Single and Multiple Failures

Cristina-Loredana Duta, Laura Gheorghe, Nicolae Tapus
Department of Computer Science and Engineering
University Politehnica of Bucharest, Bucharest, Romania
Email: cristina.duta.mapn@outlook.com, laura.gheorghe@cs.pub.ro, nicolae.tapus@cs.pub.ro

***Abstract*—Open Shortest Path First (OSPF) is a widely used link-state routing protocol in IP networks. Processing delays in OSPF implementations have an effect on the time necessary for inter-domain and intra-domain routing to re-converge after a topology change. OSPF implements different timers in order to reduce the protocol overhead. These timers ensure that the OSPF network takes several tens of seconds to recover from a failure. The delay that appears in the convergence time is due to failure detection, more specifically, is due to the value of timers and of routing calculation scheduling. In this paper, we evaluate OSPF convergence time in the presence of single or multiple failures using Quagga software routing engine and Mininet simulated network environment. The purpose is to understand the impact of failures on convergence, to observe their effects on end-to-end traffic and to determine what components should be taken into consideration in order to reduce the convergence time in a network topology based on OSPF.**

Keywords-OSPF; failure; convergence; routing software Quagga; Mininet.

I. INTRODUCTION

Nowadays, the popularity of information and communication technologies is increasing as new high bandwidth applications and services based on streaming are emerging. Because of this rapid technological advance, there is a growing demand for high-performance switching and transmission equipment. Compared with Personal Computers (PCs), where standards for development have been defined since the beginning, the field of networking equipment (that of packet switching more specifically), has always supported the development of proprietary architectures.

Routers represent the key component of the Internet infrastructure because they are interconnected through networks or links to form a backbone network which can guarantee communications between Internet users. In general, routing protocols are used in dynamic environments [1] where they have the purpose to constantly monitor any changes of the network or any events that appear [2, 3]. These functions are usually implemented at local level, in routers.

An essential characteristic of a routing protocol, which impacts end-to-end performance, is how fast it converges when topology changes happen. Convergence is when all routers have their routing tables in a state of consistency [4]. The key factor that distinguishes different routing protocols is the convergence time. Based on the speed of convergence, the routing protocols can be evaluated: the faster the convergence is, the better the routing protocol is [5].

Issues regarding convergence have been identified and analyzed at the beginning for Border Gateway Protocol (BGP) [6], but nowadays, OSPF has become widely used in the Internet infrastructure.

The purpose of this paper is to measure and analyze OSPF convergence in the presence of single and multiple failures and their impact on end-to-end traffic. We have created a simple topology and we have investigated the routing convergence under five different situations: two single link failures and three multiple link failures between different routers according to topology which will be presented further on. All the experiments were performed 10 times and each time 50 ping packets were sent.

We present the fundamental concepts regarding OSPF convergence, we analyze the impact of single and multiple failures on convergence dynamics and we describe some methods that can be useful for improving network convergence.

The rest of the paper is organized as follows. The necessary background for our work is presented in Section 2. Related work is presented in Section 3. Section 4 gives an overview of the test scenario created and it also includes a briefly outline about the convergence process and the related timers of OSPF. Section 5 offers details about our implementation, about the analysis we performed using different tests scenarios as well as the experimental results obtained and includes the adjustments of the parameters which we have done with the purpose to minimize the convergence time. Section 6 draws the conclusions for analysis of OSPF convergence behavior in the presence of single or multiple failures.

II. BACKGROUND

In this section, we present some details about OSPF, Mininet, which is the network emulator we have used in our scenarios and Quagga, which is the routing software we have selected for this evaluation.

A. Open Shortest Path First (OSPF)

OSPF [7] is a non-proprietary routing protocol which was developed in 1998 and is widely used in intra-domain Internet Service Provider (ISP) networks. OSPF is a link-state protocol that has the purpose to manage the routing table in order to use the best path to reach destination during packet forwarding.

Link-state feature is related to the functioning mode of OSPF: each OSPF router describes its topological situation,

 ISBN: 978-1-61208-422-0

its active links to all the connected counterparts so that every router knows exactly the entire topology. Link State Advertisements (LSAs) are crucial to OSPF and building the topology. LSAs represent the mean through which routers know about each other's links and who connects to whom across an area.

When the network topology changes (an event was produced – e.g., a link is down), the router communicates with the neighboring routers to determine the state of all adjacencies. The protocol used is the Hello protocol, in order to detect the failure and then generate new LSAs. LSA dissemination is done through a flooding mechanism: when a router receives a new LSA (this notifies a topological change), the LSA is sent through all of the router's interfaces, except the one it has received the new LSA from.

After the LSAs are synchronized through the mechanism previously described, the routers can correctly calculate the routing table for packet forwarding. To compute the shortest path to all destinations the Dijkstra algorithm [8] is applied, having the router as root node – in this way every router will calculate a different shortest path sub-graph.

B. Mininet network emulator

Mininet [9] is a widely used open source network emulator that can simulate a number of end-hosts, switches, routers, and links on a Linux kernel. Mininet offers several advantages such as: speed (a simple network takes only a few seconds to start up), creation of custom topologies, running real programs (anything that runs on Linux can be executed by the user of Mininet too), customization of packet forwarding, ease of use and active development.

Because it is easy to interact with the created network using Mininet Command Line Interface (CLI), to customize it, to deploy it on real hardware and to share it with others, Mininet is very useful for teaching, development and research. Mininet can be very helpful to develop and experiment with OpenFlow and Software-Defined Networking (SDN) systems.

C. Quagga routing software

Quagga is a fork of GNU Zebra Project [10] which started in 1996 by an idea of Kunihiro Ishiguro. It is a routing software package that manages TCP/IP based routing services with routing protocols support such as: BGP, Routing Information Protocol (RIP) v1, RIPv2, RIPng, OSPFv2, and OSPFv3. Quagga, allows the machine of the user to exchange routing information with other routers through specific protocols. The information gathered is used to update the kernel routing table in order to ensure the correct placement of data. Quagga can setup interface's address, flags, static routes and others. There are two modes available: normal mode and enable mode. The first one allows the user to view only the system status and the second one allows him to change the system's configuration.

It is composed of a collection, including different daemons that interact in order to build together the routing table: *RIPD* – which handles RIP protocol; *OSPFD* – which supports OSPF version2; *BGPD* – which hands BGP-4 protocol; *ZEBRA* – allows establishing communication between underlying Linux kernel and the other routing protocol daemons. For instance, if it is necessary to change the kernel routing table and to redistribute the routes between different routing protocols, *ZEBRA* sends a specific message to the kernel; *VTY* – is an additional daemon which allows configuring different routing protocols through a network accessible CLI, which accepts commands similar with the ones used on Cisco devices.

III. RELATED WORK

This section presents various methods and techniques for analyzing and improving OSPF convergence, which are described by researchers in other articles.

When dealing with protocol design, the main goal is to limit the processing power or bandwidth requirements of the protocol, while the time necessary to recover from a failure in the network topology is of secondary importance. The trade-off between efficiency and overhead can be adjusted using protocol timers. For example, Hello packet is sent periodically between neighboring routers with the frequency established by *HelloInterval* (this limits the number of hello packets).

Considering real-time applications, researchers have focused to achieve fast convergence to ensure uninterrupted traffic delivery. For instance, Francois et al. [11] tried to obtain sub-second Interior Gateway Protocol (IGP) convergence in large IP networks. Their implementation is highly dependent on the existing network resources and the frequency of failures. To achieve sub-second convergence they decreased the OSPF timers, which have an important impact on network stability.

Basu and Riecke [12] have been studying stability issues. Stability is necessary if there is a change in the network topology, all the nodes are guaranteed to converge to the new network topology in finite time, in the absence of other events. Hence, the controversy between fast convergence and protocol stability requires continuous study and research.

Referring to the topic of improving OSPF convergence is of high interest for the network research domain. Some articles propose algorithms and schemes in order to avoid the convergence process. For instance, IETF IP Fast Reroute (IPFRR) framework [13] proposes the use of pre-computed backup paths in order to reroute around the failures in the network. In [14], the authors propose a new routing scheme, which has the purpose to eliminate the convergence process completely. They present a new technique which allows packets to autonomously discover a working path. In [15], the authors present a solution that involves using network graphs and the corresponding link weights to produce a set of backup network configurations. The disadvantage of the previously mentioned approaches is that they are similar to patches, which means that they need to be added to the protocol and that they assume complex configurations.

Most of the papers that study OSPF convergence are focused on the assumption that a single failure that affected the network topology has occurred. This is good because it is widely known that network failures are of the type single link failure [16], but we have to keep in mind the fact that sometimes multiple failures can occur. In this context, we

 ISBN: 978-1-61208-422-0

decided to analyze the behavior of OSPF when dealing with single and multiple failures and to compare the results.

The researchers show that multiple failures can occur due to electromagnetic pulse (EMP) attacks [17], to natural disasters such as floods, hurricanes and earthquakes. The contribution in this area implies the presentation of guidelines for topology design and for maintenance.

In this paper, we aim to understand and analyze the behavior of OSPF convergence in the presence of single or multiple failures. Also, we present some ideas that can improve network convergence.

Due to the fact that the test scenarios vary from other similar studies (we have taken into consideration single and multiple failure for a specific network topology, performed each experiment 10 times for measurements and moreover transmitted each time 50 ping packets) it is difficult to compare our results with the ones obtained by other researchers.

IV. OSPF CONVERGENCE AND TIMERS

Network convergence represents the process of synchronizing network forwarding tables after a topology change. A network has converged when none of the forwarding table is changing for a "reasonable" amount of time. The amount of time can be defined as an interval, based on the expected maximum time to stabilize after a single topology change is produced. A diagram which represents the process of OSPF convergence is presented in Figure 1.

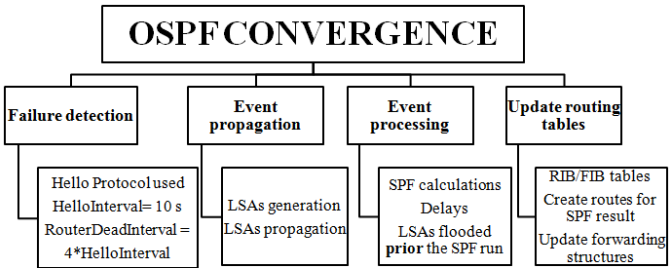


Figure 1. OSPF convergence diagram

Network restoration, which repairs the lost connections, is actually network convergence based on native IGP mechanisms. An important observation regarding IGP-based restoration techniques is the following: during the time of re-convergence, temporary micro-loops may appear in the topology (due to inconsistency of Forwarding Information Base Tables of different routers). This is very important for algorithms, because routers closer to the failure tend to update their forwarding database before the other routers.

$$Convergenc\ \ e = Failure_{\ DetectionT\ \ ime} + Event_{\ \Pr opagationT\ \ ime} + SPF_{\ RunTime} + RIB\ /\ FIB_{\ UpdateTime} \quad (1)$$

As it can be seen in formula (1), the convergence time for a link-state protocol is represented by the sum of the next components: the time necessary to detect a network failure (for instance, an interface down condition); the time it takes to propagate the event (for example, flooding the LSA, across the topology); the time necessary to perform Shortest Path First (SPF) calculations on all routers when new information is received; the time it takes to update the forwarding tables for all the routers in the area.

In this section, we describe failure detection and routing calculation related timers that are important components of convergence delay.

The top priority for fast convergence is to detect link and node failures very quickly. The primary goal is to minimize the detection/indication timers. An advantage of using point-to-point links is the fact that OSPF becomes adjacent very fast, due to the fact that Designated Routers (DRs) are no longer needed. Moreover, type 2 LSAs are not generated for point-to-point links, which reduces a little the Link-State Database (LSDB) of OSPF and also the topology complexity.

We take into consideration the fact that OSPF uses Hello protocol to detect the failure. This means that it enables routers to periodically exchange Hello packets to establish adjacency with a frequency determined by *HelloInterval*.

If Hello packets are not received by one router during *RouterDeadInterval* which is typically 4 *HelloIntervals*, the adjacency is considered down. The router that detects the failure generates new LSAs and will propagate them through the network. The default value for *HelloInterval* is considered to be 10 seconds [1]. This means that a network failure can be detected in 30 to 40 seconds after its occurrence. As it can be observed, achieving faster failure detection will significantly accelerate convergence. However, reducing the *HelloInterval* has a significant drawback: all Hello packets are processed by the router's main CPU, and if there are hundreds or more OSPF neighbors, this may have a significant impact on the router's control plane performance. The chance of false alarm increases as *HelloInterval* becomes smaller. This means that it is not recommended to reduce *HelloInterval* to the millisecond range [18].

In OSPF, topology changes are advertised using LSA/LSP (Link State Packet) flooding mechanism. To ensure that a network completely converges, a LSA/LSP must reach every router within its flooding scope.

The throttling process is controlled by three parameters: *initial* interval, *hold time*, and *max_wait time* using the command: *timers throttle lsa initial hold max_wait*.

Initial LSA generation delay has a significant impact on network convergence time, so it is important to be configured properly. The *initial delay* should be set to minimum, for instance to 5-10 milliseconds. It is not recommended to set it to zero because multiple link failure may occur synchronously.

The *hold interval* should be set so that the next LSA is sent only after the network has converged in response to the first event that occurred. In general, a single link failure results in at least two LSAs being generated, by every attached router.

Processing delay represents the time needed by the router to put the LSA on the outgoing flood lists and it is significant if the SPF process will start before flooding the LSA. Even though there are also other components that contribute to the

 ISBN: 978-1-61208-422-0

processing delay, the SPF is the most important one and we can have control over it. To ensure fast convergence, it is necessary that the LSAs are always flooded *prior* the SPF run which means that we must properly tune SPF runtime delays.

When a new LSA reaches the routers, routing calculation is scheduled. It is not recommended for the router to start executing the routing calculations immediately after receiving a LSA because more LSAs may be received and it will have to do many routing table updates in this situation. To avoid keeping the CPU busy in the case previously mentioned, OSPF uses a timer called *spfDelay* that has the purpose to delay the first routing calculation when the router receives a new LSA, so that the calculations will be performed on the entire collection of generated LSAs by the topology change. The main goal of SPF throttling is to avoid excessive calculations when the network is very unstable, but still keep the SPF reaction fast for stable networks.

V. EXPERIMENTAL EVALUATION

In this section, we perform experiments on a network topology created using an emulation system and routing software to measure and analyze OSPF convergence in the presence of single or multiple failures.

A. Test scenario

The topology is created using Mininet emulation network and in order to work with routing protocols, such as OSPF we have used Quagga routing engine. There are three applications involved.

The first application, the client, runs as a daemon in the Virtual Machine and has the role to detect changes in the Linux ARP and also in the routing tables.

The second is a standalone application, the server, which has the role to manage VMs running the client daemons. It has the role to keep the mapping between the client VMs instances and interfaces and the corresponding switches and ports. The server is responsible for deciding what to do with the packets that arrive at the controller, so it handles the protocol packets generated by Quagga and sends them out through the datapath switches.

The third application, the proxy, is responsible for interacting with the OpenFlow switches via OpenFlow protocol. The setup of the experiment is shown in Figure 2. It involves four routers each connected to a host. All of the routers will route traffic from different networks in the topology by using OSPF.

We investigate five convergence behaviors after link failure. In the first scenario, we disconnect the link between router A and router B (interface eth2) and examine the convergence time.

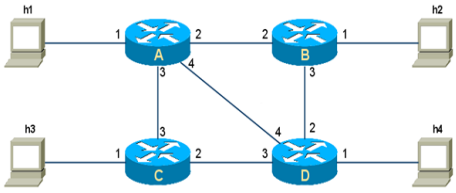


Figure 2. Evaluation Topology

In the second scenario, we remove at the same time with the link between router A and router B (interface eth2), the link between router A and router D (interface eth4) and measure the convergence time.

In the third scenario, we measure the convergence time when the link between router A and router B (interface eth2) and the link between router A and router C (interface eth3) fail at the same time.

In the fourth scenario, we determine the convergence time when the link between router B and router D (interface eth2) has failed.

The fifth experiment assumes that the link between router B and router D (interface eth2) fails at the same time with the failure of the link between router B and router A (interface eth2).

At the end of the experiment we modify OSPF timers and *HelloInterval* in order to examine any improvement in the convergence time.

B. Experimental results

In this section, we present the experimental results that we have obtained for each of the scenarios previously mentioned.

1. Single link failure (link $Router_A$–$Router_B$)

The normal traffic flows from R_A to R_B directly. In order to verify this traffic pattern, the command *traceroute* can be used. We disconnect R_B's Ethernet interface module (eth2 which connects to R_A) to simulate a broken Ethernet link.

In this case, the traffic is shifted through R_D in order to reach the host connected to R_B. We transmit fifty ping packets from R_A to R_B using ping command. We disconnect R_B's eth2 sometime during the ping command is issued.

For instance, the first 12 ping packets travel the normal path and then when R_B's eth2 is disconnected, during the transient time when the routing protocol is converging, three packets are lost. After the convergence of OSPF, the rest of the packets are sent through the backup path between R_A and R_B. In general, if there are three missing packets during the transient state of the network this indicates that OSPF needs six seconds to converge in this topology.

This ping experiment has been done ten times and the results are presented in Table I.

TABLE I. SINGLE FAILURE-RESULTS FOR OSPF CONVERGENCE

Number of experiments	Packets received	Packets lost	Convergence time (in seconds)
1	47	3	6
2	47	3	6
3	47	3	6
4	47	3	6
5	47	3	6
6	48	2	4
7	48	2	4
8	48	2	4
9	49	1	2
10	49	1	2
Average	48	2.3	4.6

 ISBN: 978-1-61208-422-0

2. Multiple link failure (link $Router_A$–$Router_B$ and link $Router_A$–$Router_D$)

The experiment setup is similar to the previous one, however this time when we transmit the fifty ping packets from R_A to R_B using ping command, we will disconnect R_B's eth2 and R_D's eth4 sometime during the ping command is issued. This ping experiment has been done ten times and the results can be seen in Table 2.

TABLE II. MULTIPLE FAILURE - RESULTS FOR OSPF CONVERGENCE

Number of experiments	Packets received	Packets lost	Convergence time (in seconds)
1	42	8	16
2	42	8	16
3	42	8	16
4	43	7	14
5	43	7	14
6	43	7	14
7	43	7	14
8	44	6	12
9	44	6	12
10	44	6	12
Average	43	7	14

As it can be observed from the results in Table I and Table II, when dealing with single link failures, the convergence time is smaller (which means faster convergence). When multiple failures occur, more delay is introduced and the convergence time rapidly increases, reaching an average value of 14 seconds for two link failures.

3. Multiple link failure (link $Router_A$–$Router_B$ and link $Router_A$–$Router_C$)

The experiment setup is similar to the previous one. The main difference is that this time when we transmit the fifty ping packets from R_A to R_B using ping command, we will disconnect R_B's eth2 and R_C's eth3 sometime during the ping command is issued. After performing the experiment 10 times, we obtained the results presented in Table III.

TABLE III. MULTIPLE FAILURE - RESULTS FOR OSPF CONVERGENCE

Number of experiments	Packets received	Packets lost	Convergence time (in seconds)
1	37	13	26
2	37	13	26
3	38	12	24
4	38	12	24
5	38	12	24
6	38	12	24
7	39	11	22
8	39	11	22
9	39	11	22
10	39	11	22
Average	38	11.8	23.6

According to Figure 3, which shows a comparison of the convergence time when we deal with single and multiple failures, it can be observed that multiple failures have a larger impact on network connectivity and protocol reaction behavior. The last values represent the average convergence time for each case.

The existence of multiple failures means multiple routing calculations which will certainly introduce more delay to convergence.

Single failure represent the situation when the link between $Router_A$ and $Router_B$ fails, the multiple failure 1 is the case when links between $Router_A$–$Router_B$ and $Router_A$–$Router_D$ fail, and the multiple failure 2 is the case when links between $Router_A$–$Router_B$ and $Router_A$–$Router_C$ fail.

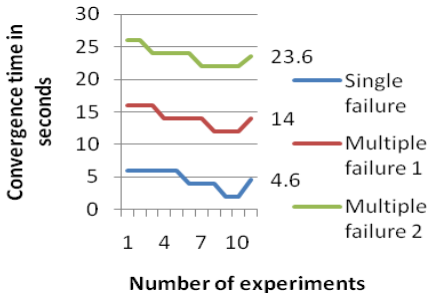


Figure 3. Comparison between convergence time when single or multiple failure occur.

4. Single link failure (link $Router_B$–$Router_D$)

The experiment setup is the same with the first, except the fact that this time when we transmit the fifty ping packets from R_B to R_D using ping command, we will disconnect R_D's eth2 sometime during the ping command is issued. This ping experiment has been done ten times and the results are presented in Table IV.

TABLE IV. SINGLE FAILURE - RESULTS FOR OSPF CONVERGENCE

Number of experiments	packets received	packets lost	Convergence time (in seconds)
1	42	8	16
2	42	8	16
3	43	7	14
4	43	7	14
5	43	7	14
6	43	7	14
7	43	7	14
8	44	6	12
9	44	6	12
10	44	6	12
Average	43	6.9	13.8

5. Multiple link failure (link $Router_B$–$Router_D$ and link $Router_B$–$Router_A$)

The experiment setup is slightly different from the previous one. This time when we transmit the fifty ping packets from R_B to R_D using ping command, we will disconnect R_D's eth2 and R_A's eth2 sometime during the ping command is issued. After performing the experiment 10 times, the results obtained can be observed in Table V.

TABLE V. MULTIPLE FAILURE - RESULTS FOR OSPF CONVERGENCE

Number of experiments	Packets received	Packets lost	Convergence time (in seconds)
1	40	10	20
2	40	10	20
3	40	10	20
4	40	10	20
5	41	9	18
6	41	9	18
7	41	9	18
8	41	9	18
9	42	8	16
10	42	8	16
Average	41	9.2	18.4

C. Options to improve OSPF convergence

Gathering all of the information above, we tried to find an optimum convergence profile based on the fact that we have different information from each router. We modify the initial *spfDelay* time, the minimum and the maximum hold time between consecutive SPF's using the command *timers throttle spf 10 100 1000* in the router's OSPF interface. After this change and taking into consideration the last case of multiple link failure (link $Router_B$–$Router_D$ and link $Router_B$–$Router_A$ fail), we obtained the values showed in Table VI for OSPF convergence time.

TABLE VI. MULTIPLE FAILURE - RESULTS FOR OSPF CONVERGENCE

Number of experiments	Packets received	Packets lost	Convergence time (in seconds)
1	43	7	14
2	43	7	14
3	44	6	12
4	44	6	12
5	44	6	12
6	44	6	12
7	44	6	12
8	45	5	10
9	45	5	10
10	45	5	10
Average	44	5.9	11.8

Another experiment that we performed is to set the *HelloInterval* to 5 seconds and to compare it to the results obtained when *HelloInterval* has the default value, which is 10 seconds. For this experiment, we have considered the case when the link between $Router_A$–$Router_B$ and the link between $Router_A$–$Router_D$ fail. The results of the convergence time are shown in Table VII and a chart with the comparison between the times can be seen in Figure 4.

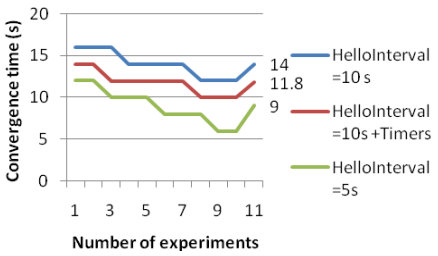


Figure 4. Comparison between convergence times with different HelloInterval values.

TABLE VII. MULTIPLE FAILURE - RESULTS FOR OSPF CONVERGENCE

Number of experiments	hellointerval=10s (no timer modifications)	hellointerval=10s (with timer modifications)	hellointerval=5s (with timer modifications)
1	16	14	12
2	16	14	12
3	16	12	10
4	14	12	10
5	14	12	10
6	14	12	8
7	14	12	8
8	12	10	8
9	12	10	6
10	12	10	6
Average	14	11.8	9

We can clearly observe from the figure that the convergence delay is increased when *HelloInterval* is larger. For instance, if the *HelloInterval* is 5 seconds, then the convergence time has an average value of 9 and if the *HelloInterval* is 10 seconds, then the convergence time reaches an average of 11.8 seconds. This test scenario involves the existence of multiple failures which introduce even more delay into the convergence process. This is because multiple failures can partition the network into many isolated parts. When *HelloInterval* has the value 5 seconds, the detection time variation will not exceed the *spfDelay*. In our experiments, the convergence time reaches an average of 8 seconds. Due to the fact that the *HelloInterval* is 10 seconds, the chance that both *spfDelay* and *spfHold* will delay successive routing calculations is higher. Also, because of the value of *HelloInterval*, in some partitions of the network, detecting the failures is much slower. Therefore, the convergence time takes approximately 14 seconds in our test scenarios. All these results demonstrate that the convergence can be delayed by timers because of protocol reaction to single and multiple failures.

VI. CONCLUSION

The aim of this paper is to measure and analyze OSPF convergence in presence of single and multiple failures and their impact on end-to-end traffic. We investigated the routing convergence under five different situations and we can conclude that OSPF converges in about 10 seconds when there is a broken Ethernet connection, and in about 19-20 seconds when there are two broken Ethernet connections. This means that the convergence is greatly delayed when multiple failures occur in the network topology.

According also to the experimental results, the convergence time is influenced by the values of OSPF timers. Larger timer values cause a slower convergence, while smaller timer values ensure a fast convergence. It is recommended to set the timers to smaller values to improve convergence time when dealing with dynamic networks. However, tuning timers require a lot of investigation on specific networks and knowledge about network management. We believe that this is still a wide research area that has just started developing.

 ISBN: 978-1-61208-422-0

ACKNOWLEDGMENT

This work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134398 and by program Partnerships in priority areas – PN II carried out by MEN-UEFISCDI, project No. 47/2014.

REFERENCES

[1] Introduction to routing protocols. [Online]. Available from: http://www.cisco.com/networkers/nw00/pres/2204.pdf 2015.06.02

[2] N. Dubois, M. Capelle, S. Chou, and B. Fondeviole, "The benefits of monitoring routing protocols in live networks", in Proc. of IP Operations and Management, 2004, pp. 9-15.

[3] E. Baccelli, and R. Rajan, "Monitoring OSPF Routing", in Proc. of the IEEE/IFIP International Symposium on Integrated Network Management (IM 20001), 2001, pp. 825-838.

[4] D. Sankar and D. Lancaster, "Routing Protocol Convergence using Simulation and Real Equipment", in Advances in Communications, Networks and Security, vol. 10, 2013, pp. 186-194.

[5] G. Lichtwald, U. Walter, and M. Zitterbart, "Improving Convergence Time of Routing Protocols", in Proc. of the 3rd International Conference on Networks (IEEE-ICN2004), 2004.

[6] T. Griffin and G. Wilfong,"An analysis of BGP convergence properties", in Proc. of ACM SIGCOMM, 1999, pp. 277-288.

[7] J. Moy, "OSPF version 2", Internet Engineering Task Force, Request For Comments (Standards Track) RFC 2328, 1998.

[8] Dijkstra's algorithm. [Online]. Available from: http://en.wikipedia.org/wiki/Dijkstra's_algorithm 2015.03.22

[9] Introduction to Mininet. [Online]. Available from: https://github.com/mininet/mininet/wiki/Introduction-to-Mininet 2015.03.22

[10] Zebra software. [Online]. Available from http://www.gnu.org/software/zebra/ 2015.03.22

[11] P. Francois, C. Filsfils, J. Evans, and O. Bonaventure, "Achieving sub-second IGP convergence in large IP network", Computer Commun.Rev., vol. 35, no. 3, 2005, pp. 35-47.

[12] A. Basu and J. Reicke, "Stability issues in OSPF routing", in Proceedings of 2001 Cnference on Applications, technologies, arhitectures, and protocols for computer communications, 2001, pp. 225-236.

[13] M. Shand and S. Bryant, "IP Fast Reroute Framework", Internet Engineering Task Force, Request For Comments (Standards Track) RFC 5714, 2010.

[14] K. Lakshminarayanan, M. Caesar, M. Rangan, and T. Anderson, "Achieving Convergence-Free Routing using Failure-Carrying Packets", in Proc. of ACM SIGCOMM, 2007, pp. 241-252.

[15] A. Kvalbein, A. F. Hansen, T. Čičić, S. Gjessing, and O. Lysne, "Fast IP network recovery using multiple routing configurations", in Proc. of IEEE INFOCOM, 2006, pp. 1-11.

[16] A. Markopoulou, G. Iannaccone, S. Bhattacharaya, C. Chuah, and C. Diot, "Characterization of failures in an IP backbone", in Proc. of IEEE INFOCOM, 2004, pp. 749-762.

[17] J. S. Foster Jr., "Report of the commission to assess the threat to the United States from electromagnetic pulse (EMP) attack", vol. I, Executive report, 2004, pp. 1-208.

[18] M. Goyal, K. Ramakrishnan, and W. Feng, "Achieving faster failure detection in OSPF networks", in Proc. IEEE International Conference on Communications (ICC2003), 2003, pp. 296-300.

[19] M. Goyal, "Improving Convergence Speed and Scalability in OSPF: A Survey", IEEE Commun. Surveys & Tutorials, vol. 14, no. 2, 2012, pp. 443-463.

[20] S. Banerjee, S. Shirazipourazad, and A. Sen, "Design and Analysis of Networks with Large Components in Presence of Region-Based Faults", in Proc. of IEEE International Conference on Communications (ICC2011), 2011, pp. 1-6.

 ISBN: 978-1-61208-422-0

The Transmission Protocol of Sensor Ad Hoc Networks

Andrzej Marczak
Gdansk University of Technology,
Faculty of Electronics, Telecommunications and Informatics
Gdansk, Poland
e-mail: amarczak@eti.pg.gda.pl

***Abstract*— This paper presents a secure protocol for a radio Ad Hoc sensor network. This network uses the TDMA multiple access method. The transmission rate on the radio channel is 57.6 kbps. The paper presents the construction of frames, types of packets and procedures for the authentication, assignment of time slots available to the node, releasing assigned slots and slots assignment conflict detection.**

Keywords-Ad Hoc; TDMA; ASAP; OFB Mode.

I. INTRODUCTION

Wireless sensor networks are appealing to researchers due to their wide range of application. The sensor networks can be used for various application areas (e.g., home, health, military). The Medium Access Control (MAC) protocol in the wireless sensor network must achieve two goals. The first is the creation of the network infrastructure. The second objective is to fairly and efficiently share communication resources between wireless sensor nodes [1]. The MAC protocols can use the Self-Organizing Medium Access Control for Sensor Networks (SMACS) protocol, the Eavesdrop-And-Register (EAR) protocol, the Carrier Sense Multiple Access (CSMA) method or the Time Division Multiple Access (TDMA) method [1] [2]. The TDMA method is often used in Ad Hoc sensor networks because of its ability to ensure collision-free transmission of packets regardless of the amount of traffic on the network. Many types of transmission scheduling protocols are used in TDMA Ad Hoc networks. Some of them do not support autonomous behaviors of mobile nodes. They cannot update the slot assignment of each node due to arrival or exit of mobile node [5]. Unifying the Slot Assignment Protocol (USAP) [7] and USAP-MA [8] Protocol, allows the operation of networks whose topology dynamically changes. However, they are characterized by poor channel utilization because of the existence of many conflicting or unassigned slots. The Traffic Adaptive MAC Protocol (TRAMA) is the algorithm proposed to increase the utilization of classical TDMA in an energy-efficient manner [3].

The sensor Ad Hoc network protocol, presented in this paper, uses the TDMA method and the Adaptive Slot Assignment Protocol (ASAP) protocol [4]. The ASAP protocol was chosen because of its ease of implementation in hardware and good properties [4]. This protocol is enhanced with authentication and encryption procedures.

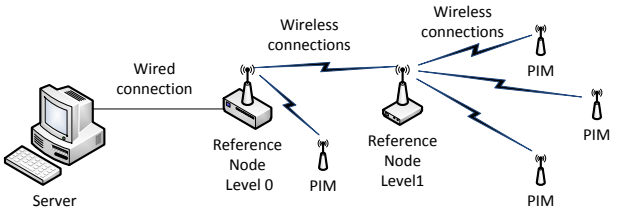


Figure 1. Sensor network.

The remainder of the paper is organized as follows. In Section II, we present the structure of the TDMA frames. In Section III, we describe the formats of the packets. Section IV presents the cipher method used in the protocol. In Section V, we introduce the authentication procedure and, in Section VI, we present the slot assignment procedure. Section VII describes the detection of conflicts in the time slot assignment. Section VIII presents the releasing of time slot assignment. Finally, conclusions are presented in Section IX.

II. TDMA FRAMES STRUCTURE

A sensor network consists of Server, Personal Identification Module nodes (PIM) and Reference Node nodes (RN) (see Figure 1). The hierarchy level of RN indicates the number of radio hops to the server. Hierarchy level 0 means that the RN is connected via a wired connection to the server.

Network nodes transmit seven types of packets:

- Data packet – DATA;
- Request packet – REQ;
- Information packet – INF;
- Hierarchy level packet – LEVEL;
- Suggestion packet – SUG;
- Reply packet – REP;
- Authentication packet – AUTH.

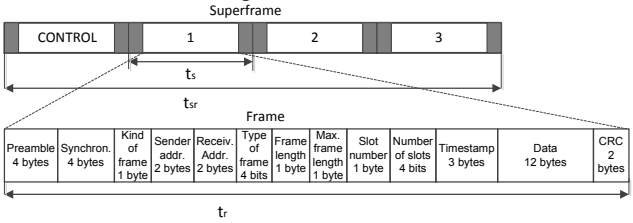


Figure 2. Frame and superframe in sensor network.

 ISBN: 978-1-61208-422-0

Packets are transmitted in the frames. The frames are organized into superframes. The primary (shortest length) superframe has 4 time slots. The duration of the superframe is 20 ms. The ASAP protocol allows the use of long superframes. The lengths of such superframes are multiples of the primary superframe. The length of the long superframe is set as a power of two. The superframes can be composed of 8, 16, 32 and 64 time slots, respectively. Such superframe durations will be then of 40 ms, 80 ms, 160 ms and 320 ms.

The transmission rate at the radio interface is 57.6 kbps, so the duration of the 1 bit is $t_b = 17.36$ µs, and the duration of 1 byte is $t_B = 8 \times t_b = 138.88$ µs. The duration of the 33-byte data packet (DATA) $t_r = 4.583$ ms. The duration of the time slot $t_s = 5$ ms, so guard intervals have 2×208.48 ms (416.96 µs), or duration of 3 bytes. In the case of packet types with fewer bytes, we used the addition of appropriate number of zero bytes (0x00) to align the packet length.

Figure 2 shows the frame and the superframe in a sensor network. The first slot in the superframe (CONTROL) has been reserved for the new node to transmit control Request packets (REQ) or authentication packet (AUTH). This way, no data packets (DATA) are transmitted in this time slot. The data packets can be transmitted over the remaining three time slots (for the superframe with length L = 4). The same is true in the case of superframes with a greater number of time slots. The first time slot is always CONTROL, used for the REQ or AUTH packets to be transmitted, and the remaining slots are used for data transmission.

Each packet type has a fixed part, depending on the structure of the packets sent by used the radio modems. This fixed part has a 4 byte preamble, 4-byte synchronization and 1 byte the kind of frame information. The PIM node addresses will have values between 0 (0x00) to 127 (0x7F), while RN will have address values from 128 (0x80) to 254 (0xFE). Address with a value 255 (0xFF) is for a broadcast transmission. In the case of transmission towards the server in packet address fields will be the source address (PIM or RN) and the address of the next RN node (the destination node). The RN node receiving the packet containing its address as the target, it checks its routing table to the next RN node address in the direction of the server and forwards the received packet in a different time slot. In the case of transmission from the server, the source address is the address of the RN node sending the packet and the destination address is the address PIM or RN node (if the packet is sent to the RN node).

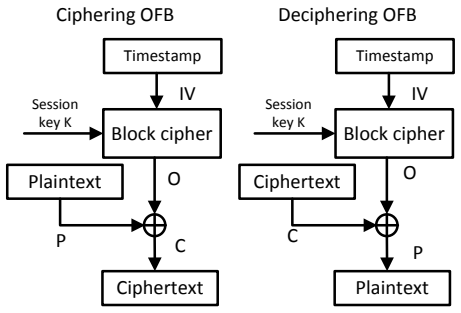


Figure 3. Ciphering and deciphering in sensor network.

III. PACKET FORMATS

- The data packet (DATA) (0x0D) is the first type of packet. It contains information on the frame length and time slots assigned to the sender, and the maximum frame length of the sender and its neighbors [4]. This packet also contains the encrypted data sent by a node.
- Request packet (REQ) (0x0C) is transmitted only by a new node. By sending this packet to neighbors, a new node requests the information on the frame length and assigned time slots of all nodes in contention area [4].
- Information packet (INF) (0x0B) contains the information on the frame length of the sender and time slot assigned to the sender and its neighbors [4].
- Hierarchy level packet (LEVEL) (0x6) is transmitted periodically by a RN node in the CONTROL time slot. The RN node is sending the packet to its neighbors, informing all nodes about its network hierarchy level.
- Suggestion packet (SUG) (0x0A) is transmitted only by a new node. By sending this packet to the neighbors, the new node announces the frame length and its assigned slot [4].
- Reply packet (REP) (0x05) is transmitted for the confirmation of receiving SUG packet [4].
- Authentication packet (AUTH) (0x01) is used in the authentication procedure. This packet sends an encrypted node address and the encrypted session key (in 2 consecutive packets).

IV. DATA SECURITY

The security of the transmission is ensured by the use of the block cipher algorithm (e.g. Advanced Encryption Standard - AES, Data Encryption Standard - DES) working in Output Feedback (OFB) mode (see Figure 3). The OFB mode has the structure of a typical stream cipher, because it generates a stream of bits (O) as a function of the initialization vector IV and session key [6]. In this solution, the IV vector is a timestamp. Its uniqueness is critical. The ciphertext is obtained by the modulo 2 addition of the plaintext bits (P) and block cipher output bits (O). One advantage of the OFB mode is low sensitivity to transmission errors, and more specifically the lack of error propagation [6]. Using the OFB mode we can encrypt data blocks of any length, even shorter than the length of the data block used in encryption algorithm (e.g. 128 bits for AES algorithm).

V. AUTHENTICATION PROCEDURE

The OFB mode is used in the procedure of authentication nodes (PIM, RN), and to encrypt data transmitted in the DATA frames. The authentication procedure takes place after the new node determines the first time slot (CONTROL) in the superframe. All network nodes keep in the memory a pair of numbers (8-bit address) and the master key (128 bits)). The same pair of numbers are stored in the server, which acts as a Key Distribution Center (KDC). Only

 ISBN: 978-1-61208-422-0

the nodes whose data (address and master key) are stored in the memory server can connect to the network. Authentication is performed after the connection to the KDC server. The transmission associated with the authentication is performed in the CONTROL time slot. The slot assignment procedure to the new node occurs only after successful authentication. The result of authentication procedure is to provide the RN or PIM the session key K, which is necessary for the exchange of information with the server. The session key is the same for all nodes in the network. Its validity can range from a few to several hours. The length of the session key depends on the encryption algorithm. For the AES algorithm, the key length is equal 128 bits. The authentication procedure consists of four steps:

- The node that wants to connect to the network transmits in CONTROL time slot, authentication packet (AUTH) containing his encrypted address. The address is encrypted using a master key.
- The server, based on the node address, searches in its memory the master key and decrypts the encrypted address. Then, it compares the two addresses (the decrypted address and the address sent without encryption in the address field of the frame).
- If the comparison result is positive, the server encrypts 128-bit session key K using the master key and sends it to the authenticated node also in the CONTROL time slot. A negative comparison result ends the authentication procedure. The server sends one AUTH packet containing zero.
- The node receives the encrypted session key K and decrypts it. Since then, all transmitted data is secured. From that moment, the entire transmission is secure. In the case of a negative authentication, after receiving the AUTH packet with the content zero, the node, after a few superframes, may initiate a re-authentication procedure.

VI. TIME SLOT ASSIGNMENT PROCEDURE

The time slot assignment procedure is performed after successful authentication of the new node. All RN nodes, passing AUTH packet with a message about the negative authentication (1 packet containing zeros - 8 bytes of 0x00) know that the node is unauthenticated and cannot compete for access to the channel. After successful authentication the new node selects a time slot assigned to itself in four steps.

A. Requesting the information on time slot assignment in the contention area.

When a new node joins the network, it does not know the information on network topology or the time slots assigned to other nodes in its contention area.

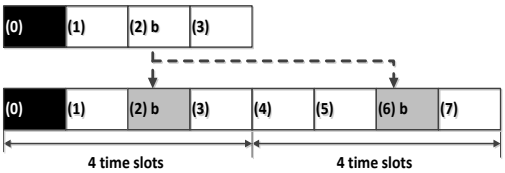


Figure 4. Copying information about node b.

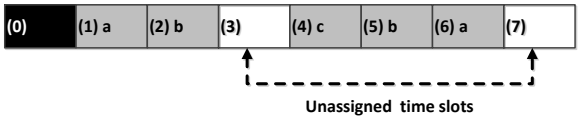


Figure 5. Getting unassigned time slots.

To get this information, the new node listens to the channel and checks packets transmitted from the neighbors. DATA packets from neighbors contain the information on their assigned time slots, superframe length, and maximum superframe length. From these pieces of information, the new node knows the position of the first time slot in a superframe and maximum superframe length among all nodes in its contention area. Then, the new node sends a REQ packet (0x0C) in the first time slot of the next superframe. Neighbors, that have received the REQ packet transmitted from the new node, transit to the control mode. Each neighbor of the new node gives information in its superframe length and time slot assigned to itself and its neighbors by transmitting an INF packet (0x0B) in its assigned time slot. After all neighbors of the new node have transmitted INF packets, all nodes in the contention area of the new node can know its structure [4].

B. Setting the superframe length and time slot assignment.

After receiving INF packets from all neighbors, the new node sets its superframe length. If all nodes in its contention area have the same superframe length, the new node sets its own superframe length to this length.

Otherwise, the new node uses the maximum superframe length among all nodes in the contention area. Then, from the received INF packets, the new node knows the information on slot assignment in this contention area. The new node creates its own time slot assignment information of superframe length, S_0, where S_0 denotes the frame length that is set to the new node. If the superframe length of a neighbor is the same as S_0, the time slot assignment information of the neighbor is copied to that of the new node. Otherwise, if $S_0 = a \cdot S_i$, the time slot assignment information of the neighbor is copied repeatedly to every S_0/a slots. S_i is the superframe length of the neighbor and a is an integer of a power of two. The new node merges the information from all neighbors and creates its own time slot assignment information [4].

For example, when the new node sets its superframe length as 8, the time slot assignment information in the INF packet received from node b whose superframe length is 4 and assigned slot is 2 is copied repeatedly to every 4 time slots in that of the new node (see Figure 4) [4].

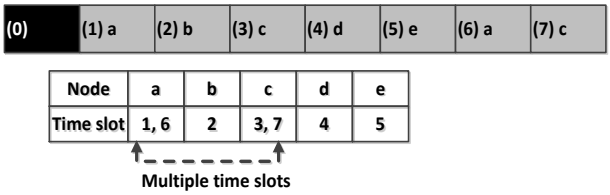


Figure 6. Releasing multiple assigned time slots.

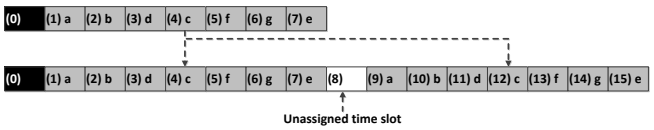


Figure 7. Doubling the superframe.

C. Selecting an assigned time slot.

Based on the time slot assignment information, the new node selects a time slot assigned to itself by three procedures.

1) Getting an unassigned time slot.

If some unassigned time slots are found in the time slot assignment information, the new node assigns one of them to itself. For example, as shown in Figure 5, when unassigned time slots 3 and 7 are found, the new node can assign a time slot either 3 or 7 to itself [4].

2) Releasing multiple assigned time slots.

If no unassigned time slot is found, the new node checks whether some nodes in the contention area are assigned multiple time slots. If such node is found, the new node releases one of these time slots and assigns it to itself. If there are more than one node to which multiple time slots are assigned, the node with the largest number of assigned time slots among them is chosen to release a time slot [5]. For example, as shown in Figure 6, when node a and c are assigned multiple time slots, the new node selects a time slot from time slots 1,3,6, and 7 which are assigned to nodes a and c, and assigns the selected time slot to itself [4].

3) Doubling the superframe.

If no unassigned time slot is found and no node has multiple assigned slots which are able to be assigned to the new node, the new node doubles the superframe length of the slot assignment information and copies the assignment information to both the former half and the latter half of doubled superframe. The first time slot in the superframe is not assigned to any nodes. Therefore, after doubling the superframe length, the first time slot in the latter half becomes unassigned slot. The new node assigns this time slot to itself [5]. For example, when the new node doubles the superframe length, time slot 8 can be assigned to itself (see Figure 7) [4].

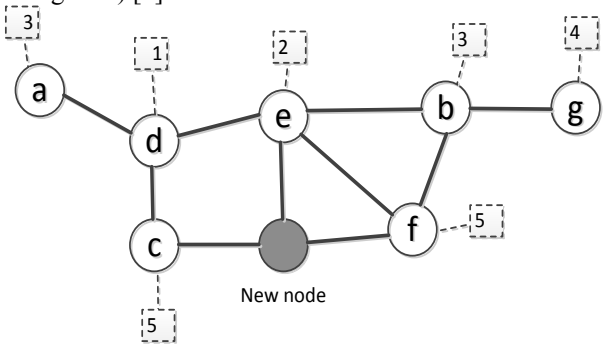


Figure 8. Example of conflict an assigned time slot.

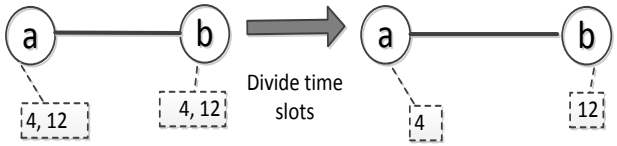


Figure 9. Dividing the assignment of time slots.

D. Announcement of updating the time slot assignment information

After selecting a time slot assignment, the new node in network sends a SUG packet to its neighbors. The SUG packet contains information on the superframe length and the assigned time slot.

When the neighboring nodes receive this packet, they update their time slot assignment information. After updating the information based on the received SUG packet, each neighboring node sends a REP packet to its neighboring nodes. Sending this packet implies the confirmation of the SUG packet for the new node and announcement of updating the time slot assignment information and exiting from the control mode. The sender and receivers of the REP packet adopt the new time slot assignment and can restart data transmission from the next superframe. The new node, after receiving the REP packets from all neighboring nodes, transits to the transmit mode [4].

VII. DETECTION OF CONFLICT

In the protocol, a conflict of slot assignment occurs when a new node connects to two or more nodes to which the same slots are assigned. In the example shown in Figure 8, a conflict occurs at a new node between node c and node f in time slot 5. When a new node detects the conflict, it solves this conflict using the following procedure [4]:

A. Dividing the assignment

If multiple time slots are conflicting at the new node, these time slots are divided to the nodes which have caused the conflict. Figure 9 shows the example when conflicting slot 4 and 12 are divided to nodes a and b [5].

B. Deleting a conflicting slot

If in the network are some un-conflicting time slots assigned to nodes causing the conflict, the conflicting time slot is released from all the nodes except for that with the smallest number of assigned slots as shown in Figure 10 [5].

C. Doubling the superframe and dividing the assignment.

If the conflict occurs among nodes to which only one time slot is assigned, this conflict cannot be solved with the current superframe length.

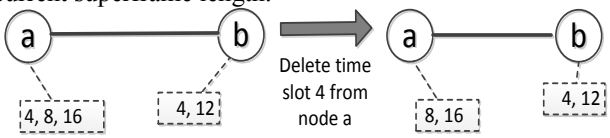


Figure 10. Deleting the conflicting time slot

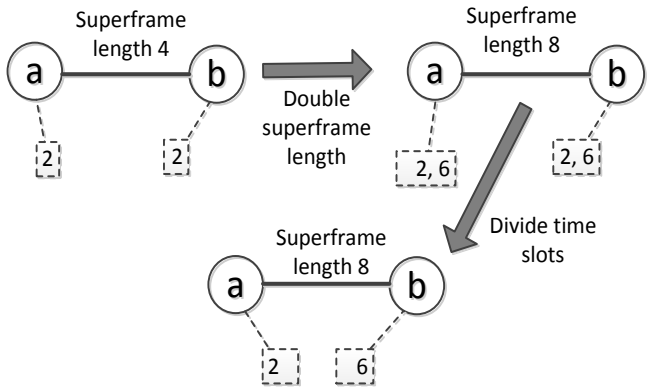


Figure 11. Doubling the superframe and dividing the assignment.

In this case, the superframe length of these nodes is doubled and the time slot assignment is divided in the doubled superframe. In the example in Figure 11, the space for conflicting time slot is doubled by doubling the superframe length. The space can be divided to nodes a and b (see Figure 11) [5].

After reconfiguring the time slot assignment, the new node sends SUG packet with the information on the reconfigured time slot assignment and the selected time slot. Neighboring nodes which have received this SUG packet also reconfigure their time slot assignment and send REP packets with reconfigured information. The new node may fail to collect the information on the time slot assignment correctly due to the collisions of INF packets. Then, the new node sends the information on the time slot in which collisions have occurred to all neighboring nodes instead of the SUG packet. Neighboring nodes of the new node, which have sent the INF packets in the conflicting time slot, retransmit the INF packets after waiting for certain superframes determined at random [4].

VIII. RELEASING TIME SLOT ASSIGNMENT

When a node exits from the network, it stops transmitting DATA packets and releases time slots assigned to itself. Neighboring nodes detect the exit of the node when no packets from exited node have been received during the time of the superframe length of the exited node. Then, they release the time slot assigned to the exited node from their time slot assignment information. They also release the time slots assigned to nodes that have gone out of their contention area due to exit of the node.

After reconfiguring the time slot assignment, neighboring nodes of the exited node send the updated information to their neighboring nodes. The nodes which have received this information reconfigure the time slot assignment by releasing the time slots assigned to the exited node [4].

IX. CONCLUSION AND FUTURE WORK

The paper presents the construction of the Ad Hoc sensor network protocol. The operation of this protocol is based on the ASAP protocol, whose efficiency measured in terms of the radio channel utilization is much larger than the USAP protocol [4]. The protocol described in this paper has been extended with additional functions related to data security (authentication of new nodes and encryption of data transmission), and the determination of the network hierarchy level of the reference nodes. Two new types of packets have been proposed: the authentication packet (AUTH) and the hierarchy level packet (LEVEL). The authentication packet is used during the authentication procedure, before the time slot assignment procedure. It increases network security and prevents connection of unauthorized nodes. Applying the LEVEL packet, that informs about hierarchy of the RN, allows the assignment of a smaller number of time slots to the PIM nodes and improves the network performance. This protocol has been selected for hardware implementation because of the relatively simple structure and principles of operation.

The future tasks of research are as follows.

- Software simulation of this protocol.
- Comparison properties of this protocol with other protocols used in TDMA Ad Hoc networks.
- Implementation of this protocol in hardware.
- Measurement of hardware network in real indoor environment.

ACKNOWLEDGMENT

This work was supported in part by the Project DOBR–BIO4/058/13045/2013.

REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, „Wireless sensor networks: a survey," Computer Networks 38, (2002), pp. 393-422.

[2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, „A Survey on Sensor Networks," IEEE Communications Magazine, August 2002, pp. 102-114.

[3] I. Demirkol, C.Ersoy, and F Alagöz, "MAC Protocols for Wireless Sensor Networks: A Survey," IEEE Communications Magazine, April 2006, pp. 115-121.

[4] A. Kanzaki, T. Uemukai, T. Hara, and S. Nishio, "Dynamic TDMA Slot Assignment in Ad Hoc Networks," Proceedings of the 17 th International Conference on Advanced Information Networking and Applications (AINA'03), March 2003, pp. 330 - 335.

[5] A. Kanzaki, T. Hara, and S. Nishio, "An Adaptive TDMA Slot Assignment in Ad Hoc Sensor Networks," Proceedings of the 2005 ACM symposium on Applied computing (SAC '05), pp. 1160–1165.

[6] W. Stallings, "Cryptography and Network Security Principles and Practice Fifth Edition," Pearson Education Inc., publish as Prentice Hall, 2011.

[7] C. D. Young, "USAP: a unifying dynamic distributed multichannel TDMA slot assignment protocol," in Proc. IEEE MILCOM '96, vol. 1, October 1996, pp. 235-239.

[8] C.D. Young, "USAP multiple access: dynamic resource allocation for mobile multihop multichannel wireless networking," in Proc. IEEE MILCOM '99, vol. 1, November 1999, pp. 271-275.

 ISBN: 978-1-61208-422-0

Securing Indirect Communication for Advanced Metering Infrastructure in Smart Grid

Mustafa Saed
Electrical and Computer Engineering
University of Detroit Mercy
Detroit, USA
email: saedma@udmercy.edu

Kevin Daimi and Nizar Al Holou
College of Engineering and Science
University of Detroit Mercy
Detroit, USA
email: {daimikj, alholoun}@udmercy.edu

***Abstract*— Smart grid will soon be a reality. As a result of connecting the traditional power grid to networks, all the vulnerabilities related to information technology will be inherited by the smart grid. Hence, the smart grid must be protected against various cyber-attacks. An essential component of the smart grid is the Advanced Metering Infrastructure (AMI). In an attempt to protect smart meters' communication with the collector, two security schemes based on PKI are introduced in this paper. The security requirements of confidentiality, integrity, and nonrepudiation are analyzed with respect to these schemes.**

***Keywords*— AM; Direct Connection; Smart grid; Security**

I. INTRODUCTION

Smart grids utilize bidirectional communication with consumers to facilitate an information-driven style to indirect energy control and management. To this extent, they deploy large scale smart meters at consumer's sites for bidirectional real time communication using existing network protocols [17]. The smart grid characterizes the new trends of the current power grid nationally and internationally. It emerged in response to environmental changes, improved energy efficiency, and reduced pollution emissions [15]. The smart grid, which is supported by information technology and intelligent control, relies on six components, namely; power generation, transmission, transformation, distribution, consumption and dispatching [11]. Smart grid refers to the next generation power grid, which upgrades the electricity distribution and management by encompassing a scalable and ubiquitous two-way communication infrastructure to enhance control, efficiency, reliability and safety [19] [24]. It is, therefore, no surprise that many countries are considering it as the future direction of the classical power grid [10] [16] [18].

Incorporating the Internet in the smart grid will widely open the door for various security attacks traditionally associated with the Internet. Undoubtedly, Smart Grid systems will significantly improve efficiency and reliability but at the expense of possibly introducing new vulnerabilities. Hence, smart grid utilization should meet rigorous security requirements [14]. Cyber-security, as a vital challenge of the smart grid transformation, must be enforced right at the beginning and not glued when attacks take place [1]. To reach full customer trust and to ensure excellent permanence of the current power supply, all components of smart grid communication network need to be extremely secure to satisfy confidentiality requirements [22]. Vulnerabilities are expected in power transmission networks, power grid, SCADA system access points and zone management [4] [6] [8]. To eliminate vulnerabilities or at least minimize their impact, strong security measures must be put in place.

Within the smart grid, the AMI plays a major role. It uses bi-directional communications between consumers and the utility, and requires robust communication network to take into account a large number of devices, small data burst transmission, high-level of reliability, and changing propagation conditions [13] [20]. Formerly, Automatic Meter Reading (AMR) was used for automatically collecting energy consumption and status data from metering devices and then transferring that data to a central database system for billing and further analysis. To allow for additional data to be read, stored, and transmitted to servers, and to control the metering devices remotely, The AMI proved to be the solution [12]. Advanced Metering Infrastructure includes the components responsible for measuring, collecting and analyzing energy usage. It consists of the Meter Data Management (MDM) system, communication network, access points, and the end points. The end points connect to smart meters, and other display and control end devices [3] [9] [23]. AMI is the only part of the smart grid in which all line segments and substations are visible [5].

Based on the importance of AMI and the vital role that it plays within the smart grid, it is very demanding that the AMI must be protected from various possible cyber-security attacks. The following security requirements must be enforced: confidentiality, integrity, availability, and nonrepudiation [7]. Consumers do not want others to know how much energy they are consuming or how it is being used (confidentiality). Meter readings and control commands should not be modified while they are being transferred (integrity). The availability of meter reading is critical for utilities and consumers. It is also critical that sending and receiving components and devices cannot deny sending information including readings and commands (nonrepudiation). There are a number of possible attacks on AMI components including denial of service, device

 ISBN: 978-1-61208-422-0

tampering, snooping, impersonation, wormhole, black hole and routing attacks. Therefore, AMI demands a reliable and secure communication approach between the smart meters and consumer equipment [2].

Vaidya et al [21] stressed that many of the available schemes for both single-path and multipath routing are not suitable for meshed AMI network. Consequently, they introduced a security mechanism for multipath routing based on Elliptic Curve Cryptology (ECC), digital signature, and Message Authentication Code (MAC) for such an AMI network. Their approach allows the Certificate Authority to do a lot more work than they should normally do (issuing certificates) including controlling the nodes' creation of public and private key. Nodes (smart meters) are doing a number of computations despite their known limited computing power. This also tends to slow the system. Furthermore, a smart meter sends it information to all the neighboring smart meters. This provides attacker the opportunity for attacking more than one goal (smart meter) as they all have the information of the source meter. The neighboring nodes, acting as as intermediate nodes, will do even more calculations and broadcast the results. This means all other nodes (smart meters) have now the information. Again, there are many nodes that the attacker can try and many nodes will be affected. An interesting security protocol for AMI communications in smart grid where the smart meters are interconnected through wireless network was introduced by Yan et al [25]. The paper indicated that the Public Key Infrastructure (PKI) is not desirable and relied on symmetric key cryptology. However, the number of symmetric keys used is large (2n, where n is the number of nodes) and comparable to the number of keys should the PKI has been followed. Symmetric keys are normally used for large messages. Furthermore, smart meters have limited capabilities, and therefore, verifying the MAC by the successor node is time consuming and should have been left to the collector. The paper did not specify what will happen when the two MAC's are not equal. This implies that the integrity of a meter's reading is not handled correctly.

This paper proposes two schemes for securing the indirect meter-to-collector communications. Both schemes are based on PKI. Unlike the work of Vaidya et al [21], this paper allows each node to send the encrypted, authenticated, and signed reading of a smart meter to its successor only (just one node). The successor cannot tell the reading of the predecessor node. If a node is attacked, readings of other nodes will not be affected. The paper also avoids the need for a certificate authority by allowing the collector node to take care of issuing certificates to all smart meters under its authority. Furthermore, nodes do not waste time performing lengthy calculations. In contrast to the approach of Yan et al [25], PKI provides stronger encryption using public and private keys. It is clear how the keys are created/recreated and exchanged. The messages (readings) are small indicating PKI is the convenient way here. The verification of the hash functions is carried out by the collector, which has more powerful computing capabilities. If the computed hash function is not equal to the received hash function for a smart meter's reading, the collector will reject that reading and inform the substation of a possible attack on that smart meter. Therefore, the integrity of a message (reading) is handled correctly. Furthermore, this paper adds anonymity to the meters by using anonymous IDs, and adds confusion to the order of readings of smart meters using a PRNG.

The AMI architecture used for this scheme will be introduced, and the security of the schemes will be analyzed. The reminder of the paper is organized as follows: Section II introduces the AMI architecture. Section III deals with the process of secure reading collection. The analysis of AMI communication security is presented in Section IV. Finally, the paper is concluded in Section V.

II. AMI ARCHITECTURE

AMI networks are responsible for connecting a substantial number of devices needed to collect readings from smart meters. As this paper is concerned with securing smart meters to collector communication, only this part of the AMI architecture will be introduced.

There are two ways of connecting smart meters to connecters; direct and indirect connections. In direct connection, smart meters directly communicate with collectors to transfer readings and exchange information and commands. For indirect (or indirect) connection, one or more smart meters are directly connected to the collector. The rest are either connected to the nearest smart meters that have direct connection with the collector or through a series of smart meters until the one directly connected to the collector is reached. The collector is responsible for collecting readings from all smart meters within its coverage area (network). Coverage area could include both direct and indirect connection.

An example of an indirect connection is presented in Figure 1 to clarify the connection." In this figure, smart meters SM_0 and SM_6 are directly connected to the collector C. Other smart meters are either directly connected to SM_0 and SM_6 (SM_1 and SM_7), or through other smart meters (for example SM_3, SM_4, SM_5, SM_8, SM_9). Collectors are connected a substation. The substation is extremely important to the efficient functioning of an electric utility since it contains a large quantity of significant information needed for the successful operation and management of the smart grid. Securing the direct smart meter-to-collector connection is easier than the indirect connection because it only needs one level of security connection with collector only. For indirect smart meter-to-collector connection, two levels of security are needed. First, the inter-meter connections must be secured, and then the direct connection with the collector. The collector will collect all the readings. If there is a problem with a reading or a missing reading due to an attack or any physical reason, the collector will report that to the substation, which will inform the management of the utility company.

 ISBN: 978-1-61208-422-0

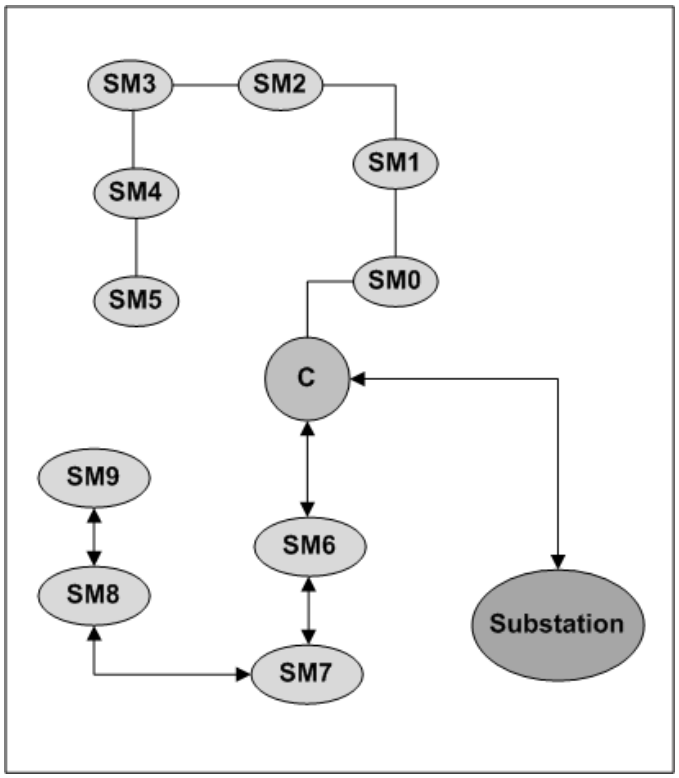


Figure 1. Smart meter-collector indirect connection

III. SECURE READING COLLECTION PROCESS

Two approaches for the indirect communication between smart meters and collector will be introduced below. In both approaches, anonymous ID's (A-ID's) for the smart meters are used. To create anonymous ID's, each smart meter XORs the current ID (real one initially and then anonymous) with the output of a true random number (TRN) generated by a ring oscillator, T_i [26]. Any other true random value can be used instead of or in addition to the one generated by the ring oscillator. In other words, A-ID_i = ID_i XOR T_i for the first A-ID_i, and A-ID_i = Previous A-ID_i XOR T_i for subsequent A-ID_i's. Table I presents the notations used in these approaches.

In the first approach, the collector C should have initially received all the public keys and IDs of the smart meters. On the other hand, the smart meters, SM's, should have the public key of the collector using any secure process. Furthermore, the predecessor and successor nodes for each smart meter are identified during installation and configuration of each smart meter. The node directly connected to the collector has no successor. The nodes at the end of the connection have no predecessors. Note that the scheme will be applied to the upper part of Figure 1 to observe how smart meters SM_0-SM_5 securely send their readings to the collector C. The readings for smart meters SM_6-SM_9 at the lower part of the figure will be collected using the same approach.

Each smart meter, SM_i, replaces its real ID_i with an anonymous one, A-ID_i, appends ID_i to it and encrypts both with the public key of collector, PU_c, before sending the resulting message, E(PUc, A-ID_i || ID_i), to C through the indirect connection (Figure 2). The collector, C, creates certificates for each smart meter, SM_i. It appends A-ID_i to the public key of each smart meter, PU_i, and the period of validity PRV, and then encrypts PU_i||A-ID_i||PRV with its private key, PR_c to get the certificate for each smart meter (CR_i = E(PR_c, PU_i||A-ID_i||PRV) since all smart meters have the public key PU_c of the collector. The CR_i is further encrypted with PU_i. Having done that, C then attaches A-ID_i to the resulting message and forwards E(PU_i, CR_i) || A-ID_i to smart meters via SM_0. Certificate creation is depicted in Figure 3 for both the collector and smart meter.

Every SM_i checks the A-ID_i. If it is its ID, it decrypts E(PU_i, CR_i) with its private key PR_i to get its certificate. Otherwise, it will forward the message to adjacent smart meters to do the same until all smart meters receive their certificates.

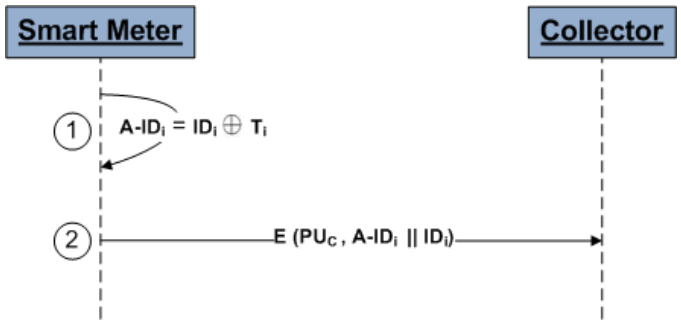


Figure 2. Creating and sending anonymous ID

Each SM_i XORs its reading, R_i, with the TRN produced by the ring oscillator, T_i, concatenates the resulting message with T_i and the hash function of the reading $H(R_i)$. The resulting message will be encrypted with PR_i to get X_i = E [PR_i, M_i || $H(R_i)$ || T_i], where M_i = R_i XOR T_i. To enable the collector to recognize the source meter's reading, A-ID_i is attached to X_i and both encrypted with PU_c to get Y_i = E(PU_c, X_i || A-ID_i). The XOR operation is used to obscure the reading of the meter. T_i is needed to allow the receiver to XOR it with M_i to get R_i. Having done that, R_i will be hashed and compared to $H(R_i)$.

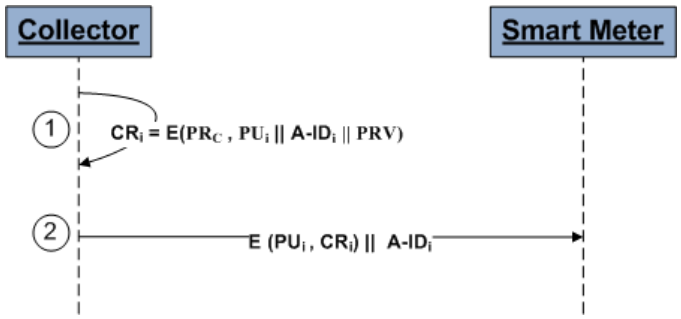


Figure 3. Creating and sending certificates

The predecessor and successor nodes exchange certificates to authenticate each other. On successful authentication, the predecessor smart meter encrypts its Y_i with the public key PU_{i-1} of the successor, and forward E[PU_{i-1}, Yi] to the successor. The receiving successor decrypts the received message with its private key PR_{i-1}, prepends or appends its own Y_{i-1} and encrypts the two (Y_i || Y_{i-1}, or Y_{i-1} || Y_i, for

example) with its successor's public key. This process will continue until all Y_is have been concatenated at SM_0. Using Figure 1 above, we should have $Y = Y_5 \| Y_4 \| Y_3 \| Y_2 \| Y_1 \| Y_0$ or any other ordering. SM_0 sends Y to C. Any missing Yi indicates a problem, possibly an attack, within that meter. If this occurs, the collector will reject the received message and report to the substation to investigate the issue.

The decision on whether to append or prepend Y_i is based on pseudorandom number generator (PRNG), which generates pseudorandom bit stream. Y_{i-1} is prepended if the pseudorandom bit is '0' and appended if the bit is '1'. This will obscure the order of Yi's and make it hard to relate the Yi's to their smart meters. To illustrate this, Figure 4 is provided.

The collector, C, uses its PR_c to decrypt Y. Then, based on the A-ID_i, it uses the appropriate PU_i to decrypt each Y_i to obtain $M_i \| H(R_i) \| T_i$ for each smart meter. It XORs M_i with T_i to get the reading R_i. It later finds the hash function of R_i and ensures it is equal to the received hash function $H(R_i)$ to guarantee the integrity of the reading, R_i. Figure 5 illustrates the meter readings collection process. To simplify Figure 5, $Z = Y_5 \| Y_4 \| Y_3 \| Y_2 \| Y_1$ (order is based on PRNG') is used.

Note that smart meter 5, SM_5, has no predecessor, and therefore, no PRNG' unit exists. Only smart meters SM_4-SM_1 have it because they have predecessors (smart meters connected to them, as depicted in Figure 1). Once the order of Y_i's is decided, the result is encrypted with the public key of the next meter, PU_{i-2}, and forwarded to the next smart meter, SM_{i-2}. The PRNG for SM_0 is not followed by encryption as in Figure 4 because it is forwarding directly to the collector.

TABLE I. NOTATIONS USED

Symbol	Meaning
C	Collector
SM_i	Smart meter i
SM_0, SM_6	Smart meters directly connected to C
PU_C , PR_C	Public & private keys of collector
PU_i , PR_i	Public & private keys of smart meter i
||	Concatenation
E	Encrypt
→	Send to
R_i	Reading of smart meter i
$H(R_i)$	Hash function of reading R_i
T_i	TRN from Ring Oscillator for smart meter i
PRV	Period of validity
ID	Identification
ID_C	ID of collector
ID_i	ID of smart meter i
A-ID_i	Anonymous ID for smart meter i
CR_i	Certificate of smart meter i

After a predefined number of readings or when the validity period PRV of the certificate expires, new keys for both collector and SM's will be generated and exchanged. The collector will use its old PR_c to encrypt the new PU_c and then encrypt the result with the old PU_i and attaches A-ID_i prior to sending it to SM_i. The A-ID_i will allow each smart meter to tell if the message is intended for it. The smart meter in question, SM_i, will decrypt this message to get the new public key of the collector. At the other side, each smart meter generates new A-ID_i, PU_i and PR_i, appends the new A-IDi to the new PU_i, encrypts the resulting message with the old PRi and then with the new public key of the collector, PU_c. Finally, the old A-IDi is attached before sending it to the collector. The collector will apply the required series of decryptions to get the new A-ID_i and PU_i of each smart meter. Note that the old A-ID_i is added to allow the collector to recognize each smart meter. Furthermore, new certificates will be generated and forwarded to the smart meters as mentioned above. This is detailed in Figure 6 below. New keys, certificates, and anonymous IDs are also created and exchanged when an attack is anticipated or has already occurred.

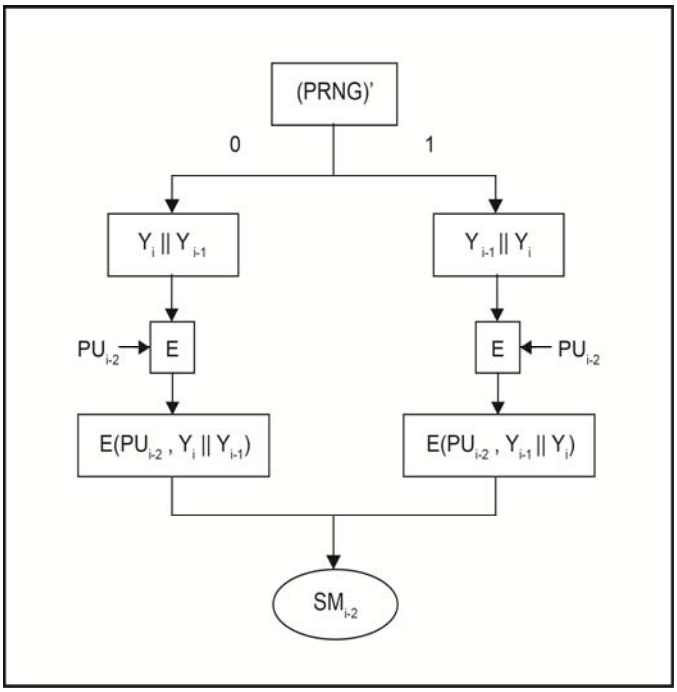


Figure 4. PRNG operation

An alternative approach is used if the creation and storage of certificates are not desirable due to computing power and memory limitations. For each adjacent smart meter pair, the collector sends the predecessor the public key of the successor encrypted with the public key of the predecessor, and sends the successor the public key of the predecessor encrypted with the public key of the successor. In both cases, the A-ID_i is attached to allow smart meters to capture messages belonging to them. Apart from replacing the certificate with the collector providing the public keys for the predecessors and successors, the rest is exactly as in the first approach.

IV. AMI COMMUNICATION SECURITY ANALYSIS

The security of the above schemes is analyzed with respect to confidentiality, integrity, and non-repudiation. Although hash functions can help with intrusion and virus detection, availability cannot be satisfied by cryptology alone (schemes above), and therefore, it will not be part of the analysis.

 ISBN: 978-1-61208-422-0

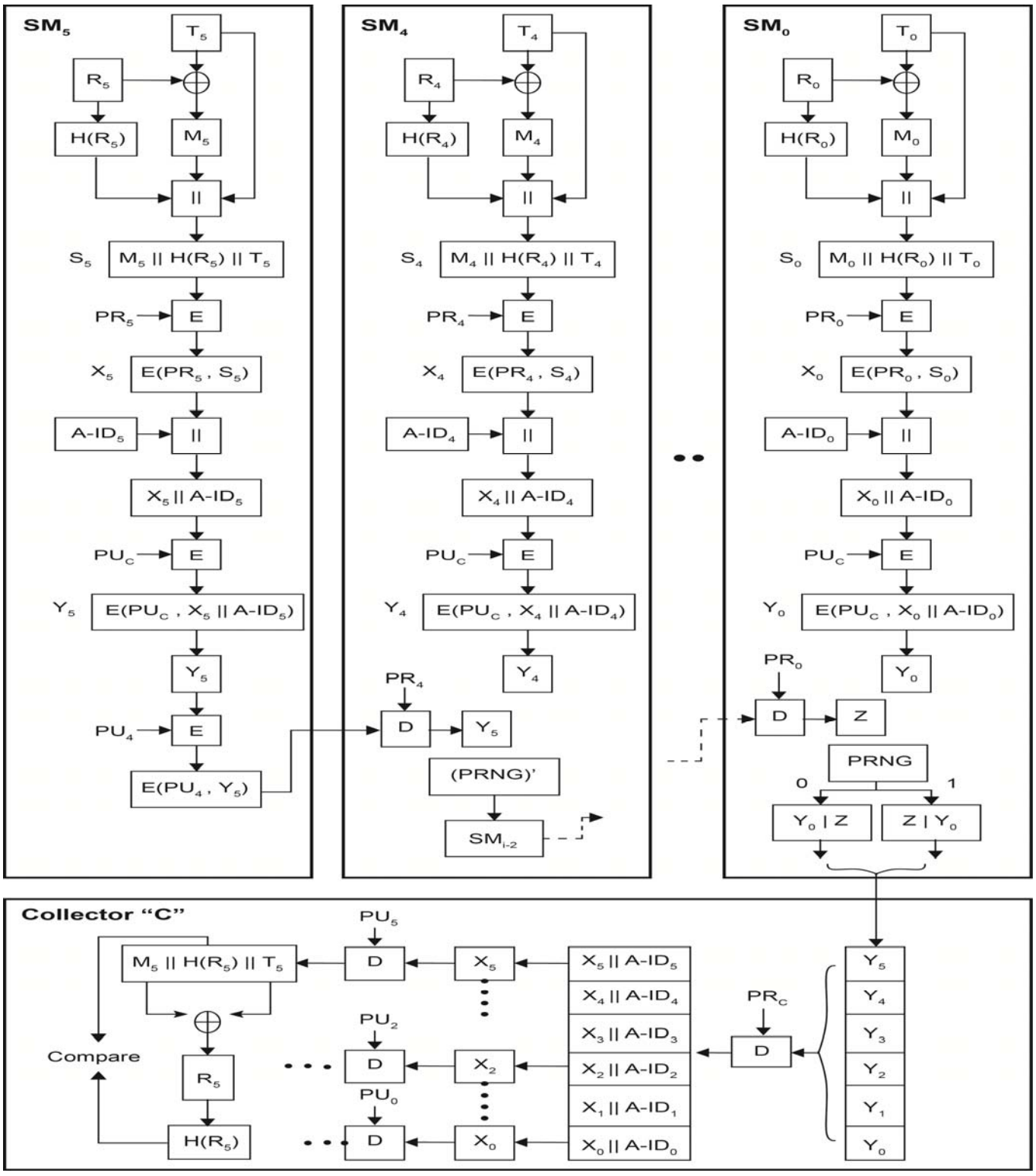


Figure 5. Meter readings collection process

 ISBN: 978-1-61208-422-0

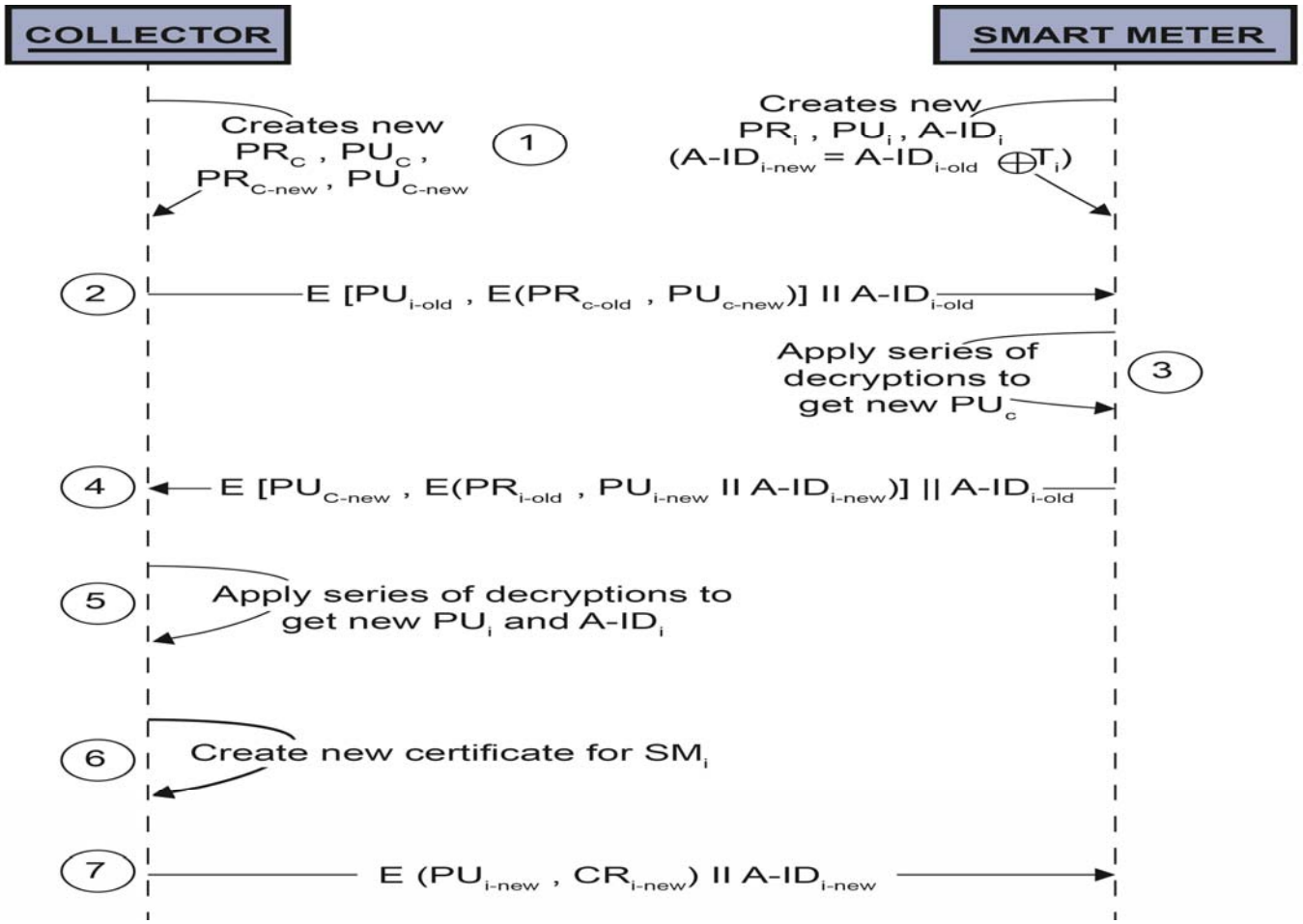


Figure 6. Exchanging new keys, IDs, and certificates

A. *Confidentiality*

Confidentiality ensures that the message sent can only be disclosed to the authorized parties. This implies that authorization restrictions are in place to ensure personal and information confidentiality.

Consumers definitely do not like others to intrude on their confidentiality in terms of energy quantity used or how it was used. They need assurance that no unauthorized disclosure of the transferred information will take place.

The proposed protocols ensure that confidentiality is met through four levels. First and most important, the message that is forwarded to the next smart meter or directly to the collector in the case of SM_0 is encrypted with the public key of the collector ($Y_i = E(PU_c, X_i \| A\text{-}ID_i)$). Only the party that has the private key (collector), PR_C, can decrypt this message. In fact, because the contents of X_i are encrypted with PR_i, and then X_i is encrypted with PU_c, authentication and digital signature are also taken care of.

In addition, the replacement of real IDs with anonymous ones will make it hard to relate a reading to a particular smart meter. Furthermore, the use of pseudorandom number generator (PRNG) introduced further hardship in judging the link between the reading and smart meter. Finally, readings are XORed with a random value that modifies the actual reading. This will make it very hard for attackers to extract the actual reading.

B. *Integrity*

Customers and utilities need an assurance that the data received is exactly as sent. This assurance guarantees that the data received has not been subject to any modification, insertion, deletions, or replay on its way to its destination. This is referred to as data integrity.

Message authentication is a technique used to ensure the integrity of the message. With regards to smart meter-to-collector communication, meter readings messages and commands should arrive exactly as they left the source that issued them.

The reading, R_i, in the proposed schemes has its integrity fulfilled through the use of cryptographic hash function, $H(R_i)$. Upon receiving the message, the collector extracts R_i and find its $H(R_i)$. It then compares the computed $H(R_i)$ with the received one. Any mismatch indicates the message has been modified. Further guard to ensure the integrity of the message was carried out by using digital signature. The hash value, $H(R_i)$, is encrypted with the private key of each smart meter ($X_i = E [PR_i, M_i \| H(R_i) \| T_i]$). Only the receiver with

 ISBN: 978-1-61208-422-0

the public key of the smart meter, which is the collector, can decrypt the hash value.

C. Nonrepudiation

Non-repudiation guarantees that the sender cannot deny it sends the information, and the receiver cannot deny it receives it. No smart meter can deny its reading because the reading and its hash value are encrypted with the private key of smart meter ($X_i = E [PR_i, M_i \| H(R_i) \| T_i]$). Provided the key was not compromised, no party but the smart meter knows its own private key.

In a similar analysis, the collector cannot deny it sent each smart meter its certificate [$(CR_i = E(PR_c, PU_i\|\|A\text{-}ID_i\|\|PRV)$] because it is encrypted with its private key and no other party knows its private key. Furthermore, the use of hash functions, $H(R_i)$, is used for nonrepudiation of the origin (dispute resolution).

V. CONCLUSION

Smart meter-to-collector communication plays a critical role within the Advanced Metering Infrastructure. Protecting them against possible cyber-attacks is a vital requirement. To contribute to this effort, two cryptographic protocols based on PKI were introduced. One of these protocols involved using certificates issued by the collector. Only the indirect communication of smart meters with the collector was investigated. Securing such a connection is harder than the direct one because readings have to travel through other smart meters before reaching the collector. The introduced schemes satisfied the security requirements; confidentiality, integrity, and nonrepudiation.

REFERENCES

[1] S. M. Amin, "Smart Grid Security, Confidentiality, and Resilient Architectures: Opportunities and Challenges," IEEE Power and Energy Society General Meeting, San Diego, CA, 2012, pp. 1-2.

[2] V. Aravinthan, V. Namboodiri, S. Sunku, and W. Jewell, "Wireless AMI Application and Security for Controlled Home Area Networks," IEEE Power and Energy General Meeting, San Diego, CA, 2011, pp. 1-8.

[3] M. Chebbo, "EU Smart Grids Framework: Electricity Networks of Future 2020 and Beyond", IEEE Power Engineering Society General Meeting, Tampa, FL, 2007, pp. 1-8.

[4] G. Chen, Z. Y. Dong, J. H. David, G. H. Zhang, and K. Q. Hua, "Attack Structural Vulnerability of Power Grids: A Hybrid Approach Based on Complex Networks," Physica A: Statistical Mechanics and its Applications, vol. 389, 2010, pp. 595-603.

[5] S. Choi, S. Kang, N. Jung, and I. Yang, "The Design of Outage Management System Utilizing Meter Information Based on AMI (Advanced Metering Infrastructure) System," in Proc. the 8th International Conference on Power Electronics, Shilla Jeju, Korea, 2011, pp. 2955-2961.

[6] S. Clements and H. Kirkham, "Cyber-security Considerations for the Smart Grid," IEEE Power and Energy Society General Meeting, Minneapolis, MN, 2010, pp. 1-5.

[7] F. M. Cleveland, "Cyber Security Issues for Advanced Metering Infrastructure (AMI)," IEEE Power and Energy General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, PA, 2008, pp. 1-5.

[8] G. N. Ericsson, "Cyber-security and Power System Communication: Essential Parts of a Smart Grid Infrastructure," IEEE Transactions on Power Delivery, vol. 25, no. 3, 2010, pp. 1501-1507.

[9] D. G. Hart, "Using AMI to Realize the Smart Grid," IEEE Power Engineering Society General Meeting, Pittsburgh, PA, 2008, pp. 1-2.

[10] A. Ipakchi and F. Albuyeh, "Grid of the Future," IEEE Power and Energy Magazine, vol. 7, no. 2, 2009, pp. 52-62.

[11] X. Jin, Y. Zhang, and X. Wang, "Strategy and Coordinated Development of Strong and Smart Grid, in Proc. the 2012 IEEE Conference on Innovative Smart Grid Technologies – Asia (ISGT Asia), Tianjin, China, 2012, pp. 1-4.

[12] I. Joe, J. Y. Jeong, and F. Zhang, "Design and Implementation of AMI System using Binary CDMA for Smart Grid," in Proc. the Third International Conference on Intelligent System Design and Engineering Applications, Hong Kong, 2013, pp. 544-549.

[13] J. H. Khan and J. Y. Khan, "A Heterogeneous WiMAX-WLAN Network for AMI Communications in the Smart grid," in Proc. the IEEE third International Conference on Smart Grid Communication (SmartGridComm), Tainan, Taiwan, 2012, pp. 710-715.

[14] A. R. Metke and R. L. Ekl, "Smart Grid Security Technology," in Proc. IEEE Conference on Innovative Smart Grid Technologies (ISGT), Gaithersburg, MD, 2010, pp. 1-7.

[15] X. Miao, X. Chen, X. Ma, G. Liu, H. Feng, and X. Song, "Comparing Smart Grid Technology Standards Roadmap of the IEC, NIST, and SGCC," in Proc. 2012 China International Conference on Electricity Distribution (CICED 2012), Shanghai, China, 2012, pp. 5-6.

[16] R. O'neill, "Smart grid sound transmission investments," IEEE Power and Energy Magazine, vol. 5, no. 5, 2007, pp. 104-102.

[17] Y. Simmhan, A. G. Kumbhare, B. Cao, and V. Prasanna, "An Analysis of Security and Confidentiality Issues in Smart Grid Software Architectures on Clouds," in Proc. IEEE 4th International Conference on Cloud Computing (CLOUD 2011), Washington, DC, USA, 2011, pp. 582-589.

[18] H. Tai and E. Hogain, "Behind the Buzz: Eight Smart-Grid Trends Shaping the Industry," IEEE Power and Energy, vol. 7, no. 2, 2009, pp. 96-97.

[19] U.S. Department of Energy (DOE), Available: www.smartgrid.gov/the_smart_grid, [retrieved: April, 2015].

[20] U.S. Department of Energy (DOE), "The Smart Grid: an Introduction," Available: http://energy.gov/oe, [retrieved: April, 2015].

[21] B. Vaidya, D. Makrakis, and H. Mouftah, "Secure Multipath Routing for AMI Network in Smart Grid," in Proc. IEEE 31st International Conference on Performance Computing and Communications (IPCCC), Austin, TX, 2012, pp. 408-415.

[22] M. Wagner, M. Kuba, and A. Oeder, "Smart Grid Cyber Security: A German Perspective," in Proc. International Conference on Smart Grid Technology, Economics and Policies (SG-TEP), Nuremberg, 2012, pp. 1-4.

[23] J. Wang and V. C. M. Leung, "A Survey of Technical Requirements and Consumer Application Standards for IP-based Smart Grid AMI Network," in Proc. the International Conference on Information Networking (ICOIN), Barcelona, 2011, pp. 114-119.

[24] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A Survey on Smart Grid Communication Infrastructures: Motivations, Requirements, and Challenges," IEEE Communications Surveys and Tutorials, vol. 15, no. 1, 2013, pp. 5-20.

[25] Y. Yan, Y. Qian, and H. Sharif, "A Secure and Reliable In-network Collaborative Communication Scheme for Advanced Metering Infrastructure in Smart Grid," in Proc. IEEE Wireless Communications and Networking Conference (WCNC), Cancun, Quintana Roo, 2011, pp. 909-914.

[26] P. Schaumont, "True Random Number Generation," Circuit Cellar, no. 268, 2012, pp. 52-58.

 ISBN: 978-1-61208-422-0

Detection of Advanced Persistent Threats Using System and Attack Intelligence

Alberto Redondo-Hernández, Aitor Couce-Vieira, Siv Hilde Houmb

Secure-NOK AS
Hamar, Norway
Email: {albertoredondo,aitorcouce,sivhoumb}@securenok.com

***Abstract*—Cyber attacks have evolved from being mostly harmless to sophisticated and devastating Advanced Persistent Threats (APT), such as the Stuxnet or Aurora attacks. APTs have the capabilities to stop business operations and cause physical damage to plants and equipment. This is a serious threat to Industrial Control Systems common in critical infrastructures such as pipelines, refineries, electrical grids or nuclear plants. This paper discusses why existing cyber attack detection technologies and solutions are not able to detect APTs, and makes use of a flawed detection paradigm based on prior knowledge of attacks. This paper also introduces a novel approach to detect APTs that is based on deep monitoring over large time intervals combined with correlation and analysis of monitored events over these time periods to detect indications of a cyber attack. The paper also provides an example of using the proposed approach to detect Stuxnet.**

Keywords–Malware; APT; Monitoring System; Intrusion Detection Systems; Intrusion Prevention Systems; Cybersecurity.

I. INTRODUCTION

The computerization of industrial environments has introduced new cybersecurity problems [1]. Cybersecurity breaches, espionage, insiders, and threats to privacy continue to increase in frequency, impact and sophistication [2]. Indeed, their impact on the global economy has been estimated at more than $400 billion in annual cost, or around 0.8% of the global Gross Domestic Product (GDP) (in comparison, drug trade represents 0.9%, and international crime, 1.2%) [3]. World leaders are raising their concerns on cyber attacks and the serious menace they pose to critical infrastructure and intellectual property [4]. Governments, in coordination with the industry, are developing strategies and guidelines to improve critical infrastructure cybersecurity and prevent the increasing social and economic impact of attacks.

The challenge is that today's Industrial Control Systems (ICS) and critical infrastructure rely on outdated security models and invalid assumptions. At the same time, the frequency and sophistication of cyber attacks against ICS are increasing and these critical assets are becoming prime targets both by criminal and terrorist organizations. These sophisticated attacks are difficult to detect and they operate covertly; they typically start with seemingly benign activities that do not trigger any warning, as was the case with the Stuxnet [5] and Aurora [6].

These attacks are called Advanced Persistent Threats (APT) [7], characterized as attacks that remain unnoticed until the consequences become visible in the system or its environment. APTs cannot be detected using conventional security tools and represent a significant challenge and risk to industrial environments and critical infrastructure.

The remainder of the paper provides an evaluation of current detection paradigms and proposes a new paradigm based on event analysis. Section II provides an overview of APTs and their phases. Section III discusses why existing detection solutions are not designed to detect APTs and use a flawed paradigm based on prior knowledge of attacks. Section IV introduces a novel detection approach tailored to the nature of APTs and based on (1) deep monitoring over large time intervals combined with (2) the analysis of monitored events over such periods to detect indications of a cyber attack. Section IV provides an example of the proposed approach applied to detect the early phases of Stuxnet.

II. ADVANCED PERSISTENT THREATS

APTs [7] works in the background conducting espionage or sabotage actions that could result in considerable monetary, environmental or safety loses. The steps taken by these threats usually go unnoticed until they have reached their goal or have penetrated large parts of the infected systems, making their removal costly and difficult. Table I outlines a selection of APTs attacks happened in the last decade [8][9]:

TABLE I. IMPORTANT ATTACKS WITHIN THE LAST SIX YEARS.

Attack	*Entry Method*	*Date*	*Classification*
Aurora operation	Malware	2007	Espionage
Stuxnet	Malware	2009	Sabotage
Energetic Bear	Malware	2011	Espionage
Flame	Malware	2012	Espionage
Shamoon	Malware	2012	Sabotage
Heartbleed	Malware	2014	Espionage

A. APT Characteristics

There has been significant research and analysis of APTs and experts have defined their main features [10][7][11] as follows:

Targeted: APTs target organizations with the purpose of stealing specific data or causing damage.

Persistent: APTs play out varied phases over a long period of time. To steal data, the attacker must identify vulnerabilities, evaluate existing security controls, gain access to privileged hosts within the target network, find the target data and, finally, ex-filtrate or manipulate them.

Evasive: APTs are designed to evade traditional security products gaining, for instance, privilege access in hosts within the target network while avoiding firewalls, antivirus and

 ISBN: 978-1-61208-422-0

other security protective mechanisms.
Complex: APTs apply a complex mix of attack methods adapted to the multiple vulnerabilities that the attacker identifies in the targeted system.

B. The APT Process

The APT process includes three dominant phases which may take place over a period of several months [12].

1) Phase 1: The attacker performs reconnaissance, identifies vulnerabilities, launches the attack and infects target hosts.

- **Recognize:** Attackers look for entry points, vulnerabilities, key individuals, and key assets.
- **Launch:** Common methods to gain access to a privileged host may include email traps with hidden malware, malicious websites aimed at extracting passwords, or social engineering to get access to specific accounts.
- **Infect:** Code is installed into a targeted host and the malware reports back to a Command and Control (C&C) fueling the attack.

2) Phase 2: The attacker controls infected hosts, updates the malware, spread it to other machines, and collects data.

- **Control:** The attacker remotely controls infected hosts with a C&C service on the Internet, often on a dynamic Domain Name System host. C&C provides the attacker with remote control.
- **Discover:** The infected host downloads additional components to identify target data on the infected hosts, on mapped network drivers, and at other network locations.
- **Persist:** An important difference between traditional malware and an APT is the ability to persist. Traditional malware will often remove itself or be removed by an antivirus program. However APT operations are designed to go on in silence and persist by downloading new code to avoid being detected.

3) Phase 3: Once the attackers have taken control of one or more hosts within the target network, they may establish access credentials to expand their reach. In case of exfiltration, the attacker send the data out of the network through the C&C server or a previously unused server. At this point, the consequences can result in the public disclosure or selling of sensitive information, blackmail, or the share of the attack methods to other attackers that may repeat the attack. In addition, if the purpose is sabotage, the attackers may manipulate the data of specific targets in order to alter the normal operations and processes that the attacked system supports.

C. The Challenge of Detecting APTs

APTs are methodical, adaptive and efficiently covering tracks while carefully penetrating the network, ceasing the attack or staying ”under the radar” for days to avoid raising suspicion, gaining knowledge of the system, or taking advantage of zero-day vulnerabilities [13] in the underlying operating system that cannot be defined through patterns or signatures.

This adaptive behaviour is hard to detect. Standard antivirus systems are not able to detect these attacks and perimeter defences; even the most sophisticated ones, are frequently breached [14]. Therefore, monitoring and detection mechanism need to implement a meticulous surveillance strategy focused on tracking the footprints of cyber attacks to be able to detect data-thefts and other loses using both system and attack intelligence.

III. Related work and Existing Cybersecurity Solutions for Industrial Control Systems

There exist multiple solutions and techniques aimed at detecting sophisticated cyber attacks, such as Tofino and Industrial Defender.

A. Tofino

Tofino is a device which provides attack detection using Deep Package Inspection (DPI) [15], with a simple installation and rugged hardware design. Whilst this method is efficient for many purposes, it is not efficient in preventing against APT because these attacks are able to change their behaviour according to the purpose of the attack. For example, an APT may, on the fly, adapt itself to a form that is not detectable as an attack using DPI.

1) Strengths: The Tofino configuration is a relatively **simple and straightforward** network monitoring solution that can be configured in various manners using the Tofino Configurator software. Tofino also includes an unique **Test Mode** that allows firewall testing with no risk to current operations, as well as being **compatible** with all Distributed Control Systems (DCS), Programmable Logic Controllers (PLC), Supervisory Control And Data Acquisition (SCADA systems), networking and software products.

2) Weaknesses: Tofino is a perimeter defence and, as we discussed in the previous section, their main problem is that they are breachable. Another issue of Tofino is that it is hardware-based and needs to be installed in a separate module. As some ICS are setup in a cabinet or a limited space type of location it might be infeasible to add additional hardware components to the network.

B. Industrial Defender

Industrial Defender [16] is a platform-agnostic monitoring system for ICS. It is employed to monitor security events, check configurations, collect data, and identify and protect the system. However, Industrial Defender is not very effective in detecting APTs, as this solution monitors the Industrial Components behaviour and not the attacker behaviour. APT attacks do not need to change the component behaviour to perpetuate the attack and therefore are not detected.

1) Strengths: Industrial Defender is comprised of a group of applications depending on the specific objective. One of its biggest benefits that it has a **single** and **unified** view of all assets within the automation systems environment, **actionable security intelligence** and that the system is able to **automate** tedious manual change management processes.

2) Weaknesses: An aspect to take into account is that Industrial Defender does not use DPI, as it obtains information from the control system itself. In addition, Industrial Defender does not provide a method to detect smart attacks, as it focuses solely on the behaviour of the controllers and does not take attack intelligence into consideration.

 ISBN: 978-1-61208-422-0

C. Other Cybersecurity Solutions

- **Darktrace Cyber Intelligence Platform (DCIP):** Evans [17] designed a cyberdefence system based on Bayesian methods with self-learning capabilities tracking evolving patterns of operations and behaviour.
- **Wurldtech Technology & Professional Services:** Ferris and Gilthorpe [18] designed a system to discover operational vulnerabilities in distinct products assessing the root cause.
- **Websense Security Labs:** McCormack [19] has designed the **TRITON** architecture, which is a set of shared security analytics, deployment platforms and management services aimed at identifying infected hosts and data extrusion attempts or prevent infection in APT phase 1.

IV. How to Detect APTs Using System and Attack Intelligence

A. Proposed method

Traditional antivirus products have been proven to be ineffective mitigating APT attacks due to the evasive nature of these threats, as discussed in [20]. The same is the case with other types of attack detection technologies and solutions, as discussed in Section III. Therefore, it is necessary to develop a more sophisticated approach tailored to monitor the behaviour of the system and correlate it with system and attack intelligence. For this purpose, we have designed an approach comprised of the following steps:

1) Monitor relevant events: We monitor events in the hosts that may be related with APTs. The monitoring system must be able to record information on various events in the hosts and network. These events represent the possible movements of the attacker and are used to build the attack patterns, including various steps of the attack. These include: insertion or removal of Universal Serial Bus (USB) devices, activation or deactivation of processes and critical processes, activation or deactivation of firewalls and antivirus, and increase or decrease of the usage of the Central Processing Unit (CPU) or the Random Access Memory (RAM).

2) Check behaviour patterns of different attacks: The second step is to check for behaviour patterns of different kind of attacks such as Denial of Service (DoS) or spyware. An example attack pattern for malware infection could be comprised of the steps illustrated in Figure 1: an attacker inserts an USB, then several processes are activated on the host and finally, the attacker removes the USB device. To separate between false positives and actual APTs, the monitoring system needs to continuously analyse against common patterns of different attacks and adopt a pessimistic algorithm for issuing alarms.

Figure 1. Malware infection pattern.

3) Raise an alarm in case of patterns detected: The monitoring system raises an alarm whenever it detects an attack pattern. This could be done by using alarm notification and by changing for example the alarm colour in a visualization module from green to orange for the affected monitored host as shown in Figure 2. The purpose of this is to display the variation of the risk level on the monitored host and separate between what might be an attack and a false positive.

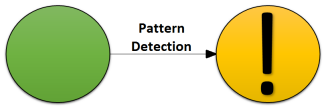


Figure 2. Raising an alarm in monitored host.

B. Example: Stuxnet attack

Stuxnet is a malware that was discovered in June 2010. It was tailored to attack Programmable Logic Controllers (PLCs) in a nuclear facility in Iran. The attack comprised the following six main phases:

1) **Infection** through an USB with a valid certificate.
2) **Search** for targeted machines.
3) **Update** itself with the latest version.
4) **Compromise** with "zero day" vulnerabilities.
5) **Control** of the systems.
6) **Deceive and Destroy.**

We assume that we have a monitor system that detect events such as USB insertion (x_1), USB removal (x_2), process activation (x_3), process deactivation (x_4), firewall deactivation (x_5), firewall activation (x_6) and increased CPU usage (x_7) and that these events are monitored continuously. The following demonstrates how to detect Stuxnet in its two first phases:

C. Phase 1: Infection through USB

Time	1	2	3	4	5	6	7
x1	0	1	1	1	1	0	0
x2	0	0	0	0	0	1	0

USB Insertion Pattern

Figure 3. How a monitoring system might detect Stuxnet in APT phase 1.

As shown in Figure 3, at instant 2, a user inserts an USB device and, after 3 periods of time, and in instant 5, the user removed an device. At this point, the system recognizes the pattern and checks whether it coincides with a common attack pattern behaviour. Finally, the monitoring system may raise an alarm if a pattern is detected or record this behaviour for later analysis.

D. Phase 2: Search targeted machines

As is shown in Figure 4, at instant 9, Stuxnet is trying to search targeted machines generating process activation, firewall deactivation and increased CPU usage. This behaviour will last five periods and will finish in instance 14. At this point, the system will check again whether this behaviour represents an attack action. Finally, the system recognizes the series of events as part of the behaviour pattern of the Stuxnet attack and issues an alarm to inform about the risk.

 ISBN: 978-1-61208-422-0

Time	8	9	10	11	12	13	14
x3	0	1	1	1	1	1	1
x4	0	0	1	0	0	1	0
x7	0	1	1	1	1	1	1
x8	0	0	0	0	0	0	0
x12	0	1	1	1	1	1	1

Search Hosts in Network Pattern

Figure 4. How a monitoring system might detect Stuxnet in APT phase 2.

V. Conclusion and Future Work

APT is an emerging threat that has already caused devastating consequences, such as, with the Stuxnet and Aurora attacks. Antivirus and other type of perimeter defences does not provide sufficient protection against these sophisticated threats for various reasons. The same is the case with existing monitoring and detection technologies and solutions, such as Tofino and Industrial Defender. This paper discussed a sophisticated dynamic attack pattern and behaviour approach to detect APTs. It performs monitoring, detection and analysis of cybersecurity events taking both attack and system intelligence into consideration and is able to detect the behavior of evasive threats such as APTs as they are emerging. In order to mitigate false positives created by the proposed approach the system continuously analyses against common attack patterns. This may be improved by applying stochastic processes to model attack pattern and behaviour.

In the future, we plan to test our patterns with other APT attacks and improving the patterns algorithms adding new events to monitor.

Acknowledgment

This work is supported by the Regionalt Forskingsfond Vestlandet project 245291, “Cybersecurity Incident Response Framework for the O&G Industry to Optimize Oil Production and Prevent Safety and Environmental Disasters”, Norway.

References

[1] A. Couce Vieira, S. H. Houmb, and D. Rios Insua, “A graphical adversarial risk analysis model for oil and gas drilling cybersecurity,” in Proceedings First International Workshop on Graphical Models for Security, ser. EPTCS, vol. 148, 2014, pp. 78–93.

[2] Symantec, “Internet security threat report 2014,” 2014, retrieved: May, 2015. [Online]. Available: http://www.symantec.com/security_response/publications/threatreport.jsp.

[3] Mcafee, “Net losses: Estimating the global cost of cybercrime,” 2014, retrieved: May, 2015. [Online]. Available: http://csis.org/files/attachments/140609_rp_economic_impact_cybercrime_report.pdf.

[4] “Crashing the system,” The Economist, July 10 2014, retrieved: May, 2015. [Online]. Available: http://www.economist.com/news/special-report/21606419-how-protect-critical-infrastructure-cyber-attacks-crashing-system.

[5] M. Kenney, “Cyber-terrorism in a post-stuxnet world,” Orbis, vol. 59, no. 1, 2015, pp. 111–128.

[6] M. Zeller, “Myth or reality—does the aurora vulnerability pose a risk to my generator?” in Protective Relay Engineers, 2011 64th Annual Conference for. IEEE, 2011, pp. 130–136.

[7] C. Tankard, “Advanced persistent threats and how to monitor and deter them,” Network Security, vol. 2011, no. 8, 2011, pp. 16 – 19. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1353485811700861

[8] S. McClure, “Operation cleaver,” Cylance, 2012, retrieved: May, 2015. [Online]. Available: http://www.cylance.com/assets/Cleaver/Cylance_Operation_Cleaver_Report.pdf.

[9] C. Five, “Advanced Persistent Threats: A Decade in Review,” Command Five PTY LTD, 2011, pp. 1–13, .

[10] P. Bondarenko, “Security Indicents Management in Offshore Drilling Rigs,” Master i informasjonssikkerhet IMT4882 Specialization Course, 2013, pp. 0–28, .

[11] Websense, “Advanced persistent threats and other advanced attacks,” A Websense White Paper, 2011, retrieved: May, 2015. [Online]. Available: https://www.websense.com/assets/white-papers/whitepaper-websense-advanced-persistent-threats-and-other-advanced-attacks-en.pdf.

[12] M. Cobb, “Advanced persistent threats - attack and defense,” Hacking, June 13 2013, retrieved: May, 2015. [Online]. Available: http://resources.infosecinstitute.com/advanced-persistent-threats-attack-and-defense/.

[13] L. Bilge and T. Dumitras, “Before we knew it, an empirical study of zero-day attacks in the real world,” Symantec Research Labs, 2012, retrieved: May, 2015. [Online]. Available: http://users.ece.cmu.edu/~tdumitra/public_documents/bilge12_zero_day.pdf.

[14] Norton, “Cibercrime report 2012,” Cibercrime Report 2012, 2012, retrieved: May, 2015. [Online]. Available: http://uk.norton.com/cybercrimereport/promo.

[15] E. B. E. Schweigert and M. Thomas, “Securing ethernet/ip control systems using deep packet inspection firewall technology,” Tofino Security, 2014. [Online]. Available: https://odva.org/Portals/0/Library/Annual20Meeting202014/2014_ODVA_Conference_Byres_Schweigert_Thomas_Securing_EtherNetIP_with_DPI_FINAL.pdf.

[16] M. Brian, “Industrial defender solutions,” Lockheed Martin’s, 1996, retrieved: May, 2015. [Online]. Available: http://www.wurldtech.com/.

[17] J. Evans, “Darktrace cyber intelligence platform (dcip),” GCHQ, retrieved: May, 2015. [Online]. Available: http://www.darktrace.com/.

[18] L. Ferris and H. Gilthorpe, “Wurltech technology & professional services,” Wurltech, 2006, retrieved: May, 2015. [Online]. Available: http://www.wurldtech.com/.

[19] J. McCormack, “Websense security labs,” Websense, 1994, retrieved: May, 2015. [Online]. Available: http://www.websense.com.

[20] D. Goldman, “Hacker hits on u.s. power and nuclear targets spiked in 2012,” The Cybercrime Economy, January 9 2013, retrieved: May, 2015. [Online]. Available: http://money.cnn.com/2013/01/09/technology/security/infrastructure-cyberattacks/.

 ISBN: 978-1-61208-422-0