ENERGY 2015

The Fifth International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies

ISBN: 978-1-61208-406-0

May 24 - 29, 2015

Rome, Italy

ENERGY 2015 Editors

Carla Westphall, Federal University of Santa Catarina, Brazil

Mario Freire, University of Beira Interior, Portugal

ENERGY 2015

Forward

The Fifth International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies (ENERGY 2015), held between May 24-29, 2015 in Rome, Italy, continued a series of events considering green approaches for smart grids and IT-aware technologies. It addressed fundamentals, technologies, hardware and software needed support, and applications and challenges.

There is a perceived need for a fundamental transformation in IP communications, energy-aware technologies and the way all energy sources are integrated. This is accelerated by the complexity of smart devices, the need for special interfaces for an easy and remote access, and the new achievements in energy production. Smart Grid technologies promote ways to enhance efficiency and reliability of the electric grid, while addressing increasing demand and incorporating more renewable and distributed electricity generation. The adoption of data centers, penetration of new energy resources, large dissemination of smart sensing and control devices, including smart home, and new vehicular energy approaches demand a new position for distributed communications, energy storage, and integration of various sources of energy.

The conference had the following tracks:

- Smart grid applications
- Energy-aware technologies
- Green communications and Smart grid technologies

Similar to the previous edition, 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 ENERGY 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 ENERGY 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 ENERGY 2015 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that ENERGY 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 green approaches for smart grids and IT-aware technologies. We also hope that Rome, Italy provided a pleasant environment during the conference and everyone saved some time to enjoy the historic beauty of the city.

ENERGY 2015 Chairs

ENERGY Advisory Chairs

Stefan Mozar, CCM Consulting / CQ University - Sydney International Centre, Australia
Mardavij Roozbehani, Massachusetts Institute of Technology, USA
Mark Apperley, University of Waikato, New Zealand
Steffen Fries, Siemens, Germany

ENERGY Industry Liaison Chairs

Dave Cavalcanti, Philips Research North America, USA
Marco Di Girolamo, Hewlett-Packard Company - Cernusco sul Naviglio, Italy
Dragan Obradovic, Siemens AG, Germany

ENERGY Special Area Chairs on Nano-Grids

Peter Müller, IBM-Zurich, Switzeland

ENERGY Special Area Chairs on Smart Grids

Ritwik Majumder, ABB AB / Corporate Research Center - Vasteras, Sweden

ENERGY Special Area Chairs on IT-energy- aware, Planning

Brian P. Gaucher, IBM Research Division - Yorktown Heights, USA

ENERGY Special Area Chairs on Grid, Green Communication

Gargi Bag, ABB Corporate Research, Sweden
Ken Christensen, University of South Florida, USA

ENERGY Special Area Chairs on Vehicular

Grzegorz Swirszcz, IBM Watson Laboratory, USA

ENERGY 2015

Committee

ENERGY 2015 Advisory Chairs

Stefan Mozar, CCM Consulting / CQ University - Sydney International Centre, Australia
Mardavij Roozbehani, Massachusetts Institute of Technology, USA
Mark Apperley, University of Waikato, New Zealand
Steffen Fries, Siemens, Germany

ENERGY 2015 Industry Liaison Chairs

Dave Cavalcanti, Philips Research North America, USA
Marco Di Girolamo, Hewlett-Packard Company - Cernusco sul Naviglio, Italy
Dragan Obradovic, Siemens AG, Germany

ENERGY 2015 Special Area Chairs on Nano-Grids

Peter Müller, IBM-Zurich, Switzeland

ENERGY 2015 Special Area Chairs on Smart Grids

Ritwik Majumder, ABB AB / Corporate Research Center - Vasteras, Sweden

ENERGY 2015 Special Area Chairs on IT-energy- aware, Planning

Brian P. Gaucher, IBM Research Division - Yorktown Heights, USA

ENERGY 2015 Special Area Chairs on Grid, Green Communication

Gargi Bag, ABB Corporate Research, Sweden
Ken Christensen, University of South Florida, USA

ENERGY 2015 Area Chairs on Vehicular

Grzegorz Swirszcz, IBM Watson Laboratory, USA

ENERGY 2015 Technical Program Committee

Amir Abtahi, Florida Atlantic University - Boca Raton, US
Nizar Al-Holou, University of Detroit Mercy, USA
Ahmed Al-Salaymeh, University of Jordan, Jordan
Lachlan Andrew, Swinburne University of Technology – Melbourne, Australia
Luca Ardito, Politecnico di Torino, Italy
Mehdi Bahrami, University of California, Merced, USA
Chakib Bekara, Fraunhofer Fokus Institute – Berlin, Germany
Fabio Luigi Bellifemine, Telecomitalia, Italy
Rachid Benchrifa, Mohammed V University, Morocco
Siegfried Benkner, University of Vienna, Austria
Riad Benelmir, University of Lorraine, France
Frede Blaabjerg, Aalborg University, Denmark
Robert Brewer, Aarhus University, Denmark
Antonio Caló, NorTech Oulu -Thule Institute / University of Oulu, Finland
Davide Careglio, Universitat Politècnica de Catalunya - Barcelona, Spain
Mari Carmen Domingo, Barcelona Tech University, Spain
Dave Cavalcanti, Philips Research North America - Briarcliff Manor, USA
Konstantinos Chasapis, University of Hamburg, Germany
Antonin Chazalet, IT&Labs, France
Mohamed Cheriet, ETS/GreenStar - Montreal, Canada
Abdeljabbar Cherkaoui, ENSA Tanger, Morocco
Howard Choe, Raytheon - McKinney, TX, USA
Sangho Choe, The Catholic University of Korea, Korea
William Cheng-Chung Chu, Tunghai University, Taiwan
Delia Ciullo, EURECOM Sophia Antipolis, France
Peter Corcoran, College of Engineering & Informatics, NUI Galway, Ireland
Margot Deruyck, Ghent University/IBBT, Belgium
Marco Di Girolamo, Hewlett-Packard Company, Italy
Yong Ding, Karlsruhe Institute of Technology (KIT), Germany
Ron Doyle, IBM | Software Group Strategy and Technology - RTP, USA
Venizelos Efthymiou, University of Cyprus, Cyprus
Soumia El Hani, Université Mohammed V, Morocco
Larbi El Farh, Université Mohammed 1er, Morocco
Tullio Facchinetti, University of Pavia, Italy
Farid Farahmand, Sonoma State University, USA
Eugene A. Feinberg, Stony Brook University - New York, USA
Alexandre Peixoto Ferreira, IBM Austin Research Laboratory, USA
Steffen Fries, Siemens Corporate Technology - Munich, Germany
Brian P. Gaucher, IBM Research Division - Yorktown Heights, USA
Xiaohu Ge, Huazhong University of Science and Technology, China
Erol Gelenbe, Imperial College London, UK
Hamid Gharavi, National Institute of Standards and Technology, USA

Georgios B. Giannakis, University of Minnesota - Minneapolis, USA
Manimaran Govindarasu, Iowa State University, USA
Mesut Günes, Freie Universität Berlin, Germany
Rune Gustavsson, KTH, Sweden
Daniela Hossu, University 'Politehnica' of Bucharest, Romania
Chun-Hsi Huang, University of Connecticut - Storrs, USA
Zhiyi Huang, University of Otago, New Zealand
Ahmed Ihlal, Ibn Zohr University, Morocco
Muhammad Ali Imran, University of Surrey, UK
Canturk Isci, IBM TJ Watson Research Center, - New York, USA
Philip Johnson, University of Hawaii - Honolulu, USA
Aravind Kailas, University of North Carolina - Charlotte, USA
Essam E. Khalil, Cairo University, Egypt
Young Sun Kim, Korea Electrotechnology Research Institute (KERI), Korea
Thierry E. Klein, Bell Labs / Alcatel-Lucent, USA
Janine Kniess, Santa Catarina State University, Brazil
Sileshi Kore, Dilla University, Ethiopia
Dejan Kostic, Institute IMDEA Networks, Spain
Paul J. Kuehn, Institute of Communication Networks and Computer Engineering - University of Stuttgart, Germany
Dimosthenis Kyriazis, University of Piraeus, Greece
Salah Laghrouche, Université de technologie Belfort-Montbéliard, France
DongJin Lee, University of Auckland, New Zealand
Marco Listanti, University Sapienza of Roma, Italy
Eugene Litvinov, ISO New England, USA
Shanshan Liu, Electric Power Research Institute, USA
Marco Lützenberger, Technische Universität Berlin, Germany
Priya Mahadevan, PARC, USA
Thair Shakir Mahmoud, Edith Cowan University - Western Australia, Australia
Mikko Majanen, VTT Technical Research Centre of Finland, Finland
Michael Massoth, University of Applied Sciences - Darmstadt, Germany
Satoshi Matsuoka, Tokyo Institute of Technology, Japan
Rafael Mayo-García, CIEMAT, Spain
Jean-Marc Menaud, Ecole des Mines de Nantes, France
George Michailidis, University of Michigan, USA
Marilena Minou, Athens University of Economics & Business, Greece
Nicolas Montavont, Telecom Bretagne, France
Daniel Mossé, University of Pittsburgh, USA
Gero Mühl, Universitaet Rostock, Germany
Masayuki Murata, Osaka University, Japan
Simin Nadjm-Tehrani, Linköping University, Sweden
Ahmed Nait-Sidi-Moh, University of Picardie Jules Verne (UPJV), France
Bruce Nordman, Lawrence Berkeley National Laboratory, USA
Dragan Obradovic, Siemens AG - München, Deutschland

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

Demand Response for Increasing Renewable Energy Penetration in Isolated Power Systems

Michael Negnevitsky
University of Tasmania
Hobart, Australia
e-mail:
Michael.Negnevitsky@utas.edu.au

Dusan Nikolic
Hydro Tasmania
Hobart, Australia
e-mail: Dusan.Nikolic@entura.com.au

Martin de Groot
CSIRO
Sydney, Australia
e-mail: Martin.deGroot@csiro.au

Abstract— **Electric power industry is undergoing a profound change. The change is driven by technical, economic and environmental factors. The emerging challenges are particularly significant for distribution grids, where the level of automation or "smartness" is relatively low. With the push for energy conservation, demand response is becoming a vital tool under the smart grid paradigm. This paper outlines some experience obtained at the University of Tasmania, Hydro Tasmania and the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia in developing fast DR for isolated power systems. The paper presents results of the implementation of DR for enabling higher wind energy penetration in isolated power systems. The presented approach is based on the centralized control of residential and commercial loads with the DR execution time of 1 second. The technology discussed in the paper has been implemented in the King Island power system in Australia.**

Keywords-smart grid; distribution system; demand response.

I. INTRODUCTION

Electric power systems are undergoing a profound change. This change is driven by several factors that include technical, economic and environmental factors. We need to deal with an aging infrastructure of power systems and maintain the required level of grid reliability. We need to integrate renewable energy sources, particularly wind and solar, and provide secure power supply to our customers, and at the same time improve operational efficiency. The emerging changes and challenges are particularly significant for distribution grids, where the level of automation or "smartness" is relatively low. Manual and "blind" operations along with old electromechanical relays are to be transformed into a "smart grid". This transformation is necessary to meet environmental targets, accommodate distributed generation, and support plug-in electric vehicles. In fact, these needs present the power industry with the biggest challenge it has ever faced. On one hand, the transition to the "grid of the future" has to be evolutionary – we still need to supply electricity to our customers to keep the lights on. On the other hand, the challenges associated with the smart grid are significant enough to expect revolutionary changes in power system design and operation.

With the push for energy conservation, demand-side management and demand response are becoming vital tools under the broad smart grid paradigm.

The term "demand-side management" (DSM) was first introduced by Electric Power Research Institute (EPRI) in the 1980s, and since then has been widely used around the world. In fact, DSM is a term that implies many activities such as direct load control, peak shaving, peak shifting, and various load management strategies. Effective load management programs are often referred to as demand response (DR). According to the US Federal Energy Regulatory Commission, DR is defined as:

"Changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized."

Customers living in remote areas often cannot be supplied from conventional interconnected power systems. These customers are usually serviced by a local electricity generation and distribution system with electricity generated using diesel fuel. Due to remoteness and consequent high cost of diesel fuel supply, the cost of electric energy in isolated power systems is high compared to conventional interconnected systems. In some locations, the price exceeds US $1/kWh, which is an obvious incentive for introducing renewable energy (RE) generation. Unfortunately, RE from the two most abundant energy sources – wind and solar – incurs significant stability and reliability issues due to the intermittency of those sources.

This paper outlines and builds on recent research in utilizing DR in power systems. It presents fast (i.e., sub-second) DR as an enabling technology for increasing renewable energy penetration in isolated power systems (IPS). The contribution of this paper lies in presenting the concept of implementing fast DR in a real IPS case study, presenting preliminary results proving the concept of fast DR and outlining its potential future applications.

The paper has the following structure. In Section II, we discuss potential benefits of the DR application in isolated power systems. In Section III, we consider DR as a virtual power plant, or a DR generator, for supporting higher RE

 ISBN: 978-1-61208-406-0

penetration by providing required spinning reserves. We present the architecture of the DR generator, communication protocols and the structure of the control system. In Section IV, we present a case study based on the King Island Renewable Energy Integration Project. Finally, in Section V, we report our conclusions.

II. DEMAND RESPONSE IN POWER SYSTEMS

A. *Standard approach to DR*

Recent research has shown that DR can provide benefits to power systems and their customers by:

- Supporting frequency and/or voltage regulation [1, 2];
- Reducing operational costs and emissions by increasing utilization of RE sources [3]. Note that reducing operating costs in turn leads to a greater return on investment which would incentivize expansion of the RE industry;
- Reducing operational costs and emissions caused by traditional generators installed to provide spinning reserve for RE [4];
- Relieving stress from transmission and distribution infrastructure by coordinating loads close to RE sources [4];
- Reducing utility operating costs through advanced metering infrastructure installed to enable DR [5].

Most of the research is focused on large systems, price-driven signals to DR [6], day-ahead scheduling [7], or demand response which responds in period of 10-minutes or longer [8].

In recent years, researchers have started recognizing the need for fast direct load control; one which does not depend on price responsiveness [9]. Benefits of this fast DR was taken further by showing that sub-second response of DR can support primary frequency regulation [10].

This paper focuses on benefits of fast DR in IPS.

B. *Special case of IPS*

Most DR applications focus on large power systems, which have a steady and predictable load profile defined by morning and evening peaks. The time and size of these peak periods can be accurately estimated using historical load data and weather forecasts. In contrast, isolated power systems not only supply less power (MWs, rather than GWs) they are also geographically much smaller. Being smaller in capacity means that demand is less predictable. Being smaller in area means that supply from RE sources is more variable, as a larger percentage of the RE generators are likely to be affected by the same weather events (e.g., a lull in the wind or passing clouds). Due to their reduced demand predictability and increased variability of RE supply, conventional generation scheduling in isolated systems with RE is more challenging. From a generation scheduling perspective - where RE generation is usually treated as a load offset – the daily load curve becomes extremely volatile. An example of an IPS with RE load curve is giving in Figure 1. Notice that it does not even have predetermined daily peaks.

High load variation makes scheduling of diesel generation more difficult and less efficient, as diesel engines will rarely operate at their peak efficiency, and more generator start-ups are required. This is where fast DR can help. It can smooth the variability of required diesel generation by quickly adjusting system load. To be able to do that, fast DR cannot rely on typical load patterns – it must be executed in real-time. If DR is to complement RE in an isolated system, it has to be as fast as the speed at which RE generators change their power output.

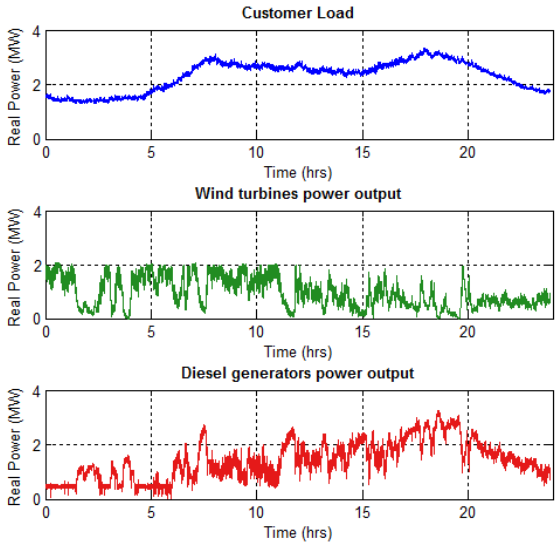


Figure 1. Daily load diagram in an isolated power system.

III. DEMAND RESPONSE AS A VIRTUAL POWER PLANT

The idea of aggregating and controlling small loads to create a large block of variable demand has been discussed in the power engineering literature where it is often referred to as a 'virtual power plant' [11]. Since isolated power systems generally have only one power station, the idea of adding another (virtual) power plant to this system might be a bit misleading. However, isolated systems with RE might have several generating sources (e.g., diesel, wind, solar, etc.). Therefore, aggregated DR could be treated as a 'virtual generator'. Although such a 'DR Generator' (DRG) is not generating real power, the power system controller perceives it as one due to the DRG's ability to decrease the amount of energy needed from other generating sources.

Isolated systems are usually controlled by a single controller. The controller is typically implemented with a programmable logic controller (PLC). The role of the controller is to schedule available generation in accordance with the current power system constraints, and to maintain system stability. In addition, the controller can be programmed to maximize the amount of RE generation and, consequently, minimize running costs. The controller

 ISBN: 978-1-61208-406-0

effectively controls the entire system by collecting data on the current system status and by issuing commands to various generation sources, as shown in Figure 2.

When the controller has a goal to maximize the use of RE, it will dispatch as much renewable generation as the power system can handle while simultaneously maintaining an appropriate level of spinning reserve to ensure system stability. If the amount of RE drops and the system suddenly does not have enough spinning reserve, the controller starts a diesel generator. The role of the DRG is to support higher RE penetration by providing additional spinning reserve. If sufficient spinning reserve is provided by the DRG, diesel generator start-up is prevented.

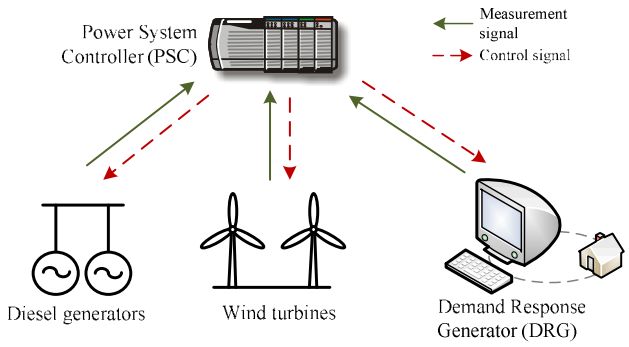


Figure 2. Smart Grid generator in IPS control system.

A DRG consists of three main components, as shown in Figure 3:

- DRG Master controller,
- communications network, and
- Slave controllers.

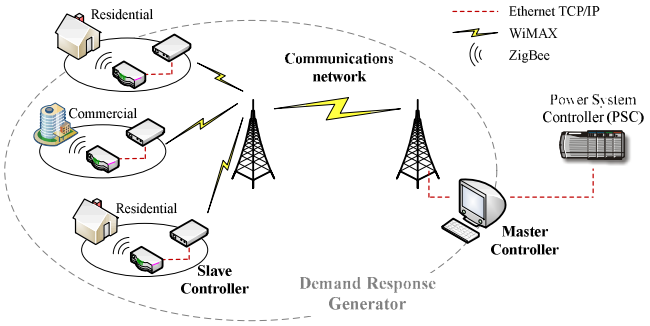


Figure 3. The DRG architecture.

A. DRG Master Controller

The master controller collects metering data (voltage, current, real and reactive power) from each available load and aggregates it into predefined virtual loads (e.g., feeders, geographical regions). It only aggregates the loads, which are currently consuming energy.

The Master controller then communicates available DR capacity to the power system controller (PSC) while checking for DR dispatch requests from the PSC. DR requests identify a target virtual load and the amount of demand to curtail. The master controller selects the individual loads to curtail from the virtual load, and immediately sends a switch-off signal to each. It also ensures that customers are not greatly affected by load curtailment, and that all customers are treated evenly. This is achieved by imposing constraints on the maximum length of control for each individual load.

Finally, Master controller controls only individual consumer loads and not entire feeders or their parts.

B. Communications Network

A multi-protocol bidirectional communications network delivers information between all elements of the DRG. Ethernet is used within the control system. A dedicated WiMAX network provides the backhaul capability. Within individual customer sites, a WiMAX gateway is connected by Ethernet to a ZigBee gateway for the final link to the load metering and switching devices. This communications configuration is configured to ensure a sub-second round trip for DR requests from the PSC out to the load control devices and back again.

C. Slave Controllers

Slave controllers are located in each DR capacity providing site. They consist of a pair of gateways to provide WiMAX-Ethernet-ZigBee signal translation between the backhaul network and the individual load control devices. The load control devices perform both metering and load switching. They provide a range of power metrics and also support set-points.

A DRG providing spinning reserve must be extremely responsive and reliable. The master controller must be able to monitor and dispatch slave controllers at all times. This critical requirement becomes obvious in the two most common scenarios:

- If the PSC requests DR for extended periods of time, some slave controllers may override the dispatch as they exceed their maximum dispatch duration. In this situation, the master controller must quickly identify and dispatch another device (or devices) with an equivalent load.
- If a slave controller or communication link is unreliable, the DRG may be forced to always dispatch more DR than requested to ensure a suitable margin of error in either load switching or reporting. This is not an efficient use of capacity and may reduce the overall effectiveness of the DRG.

IV. Case Study

The fast DR technology discussed above was implemented in an IPS as part of the King Island Renewable

 ISBN: 978-1-61208-406-0

Energy Integration Project [12]. King Island lies in the Bass Strait between Tasmania and the Australian mainland. It has a population of approximately 2000 people, and an economy based on agriculture and food processing.

A. The King Island Power System

Customer load on King Island ranges between 1 MW and 3 MW, with an average of around 1.5 MW. The King Island power system is shown in Figure 4. There is one power station on the island with four distribution feeders delivering electricity to customers. The power station houses four diesel generators with a total generation capacity of 5.8MW. Three fixed speed Nordex N29 (250kW each) wind turbines are installed on a nearby hill, together with two Vestas V52 turbines (850kW each) with doubly fed induction generators. Two 800kW diesel engines with flywheels are also connected to the system. In these generators the flywheels are separated by a clutch from a diesel engine and provide system with additional inertia.

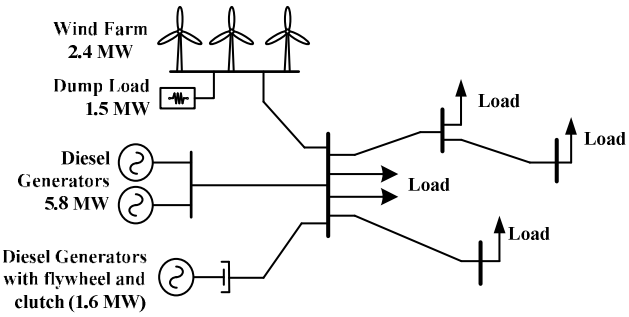


Figure 4. The simplified schematic of the King Island power system.

Normally the spinning reserve in the King Island power system is provided by diesel generators. During the wind-only operation of the system, the spinning reserve is equal to the dump load, as it represents the surplus of renewable energy in the system. If wind power suddenly drops, quick start-up of diesel generation is needed.

By adding to the spinning reserve (represented by dump load) during the wind-only operation of the system, the available DR has the opportunity to delay or prevent diesel generation start-ups, and by doing so, increase RE penetration in the system.

B. The King Island Smart Grid Project

The ongoing King Island Smart Grid project has the goal of supporting higher levels of wind energy integration in the King Island power system by providing:

1. Spinning reserve by implementing the DRG concept, and
2. Fast fine-grained under frequency load shedding (based on the slave controller level).

The master controller constantly monitors each available load, aggregates it and passes this information to the PLC-based power system controller (PSC). At the same time, the PSC monitors the current level of power system spinning reserve.

If the spinning reserve falls below a predefined threshold, the controller instructs the DRG to curtail some load and effectively raise the spinning reserve. This function can be observed in Figure 5, where due to a sudden drop in wind generation, the spinning reserve falls. The reduction in spinning reserve causes the power system controller to initiate a request for all available DR, as shown in the lower graph.

The results shown in Figure 5 demonstrate that the implemented DRG was able to respond accurately to given set-points. It also shows that DR capacity can be dispatched reliably in 1 second.

C. Selection of Smart Grid technology

As presented in Section III, there are three components of the Smart Grid project: master controller, communications network and slave controllers. The King Island Smart Grid project is, in fact, a research and development project. However, all the equipment installed during the project duration is meant to be industrial grade and capable of operating in real power systems for extended periods of time. Having this in mind, ruggedized and proven equipment has been used during in the Smart Grid project:

- Master controller is an industrial in-rack server computer with operating systems for embedded devices. It requires very little or no maintenance.
- Communications network is built using widely available and proven telecommunications technology. Since the requirements of the King Island Smart Grid are much lower than usual operating requirements to communication networks, the network maintenance cost is also very low. To ensure the minimum response time, a proprietary long-polling protocol has been developed by CSIRO and used for communications between master and slave controllers. This protocol maintains a permanent open communication link between devices. There is no need for opening and closing the communication link, which provided an obvious advantage in the overall response time compared with industrial standards.
- Slave devices are modem, gateway and miniature circuit breaker (MCB) switch capable of at least 50,000 operations. If any of these devices malfunctions, their replacement is relatively inexpensive and fast.

An essential part of the King Island Smart Grid project is to provide training to local electricians in maintaining the system including its repair when equipment fails. The Smart Grid has been in operation for over a year now, with no serious equipment failures. The outcome of selecting widely available devices is a relatively low capital cost of the installed equipment. Selecting proven and ruggedized equipment resulted in very low maintenance costs.

 ISBN: 978-1-61208-406-0

D. *Preliminary Implementation Results*

Currently the King Island DRG has 50 sites under management and has been fully integrated into the power system. When complete, the DRG will be extended to include 150 households and several commercial loads.

Prior to roll-out on King Island, the DRG was tested with 10,000 simulated customer loads with minimal performance degradation [13]. This implies that the full DRG will supply more than 100 kW of sub-second DR capacity.

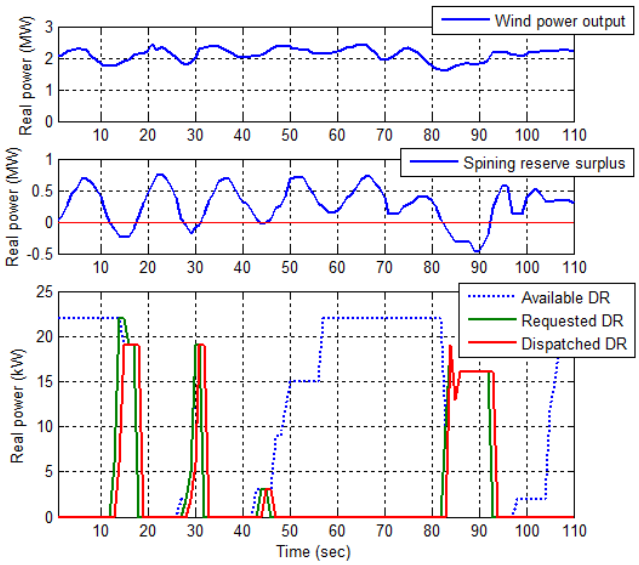


Figure 5. Results of the King Island DRG operation.

The effectiveness of the King Island DRG depends largely on its integration with the power system controller. Figure 6 demonstrate how the controller requests for the DRG and uses it as a tool for regulating demand accurately.

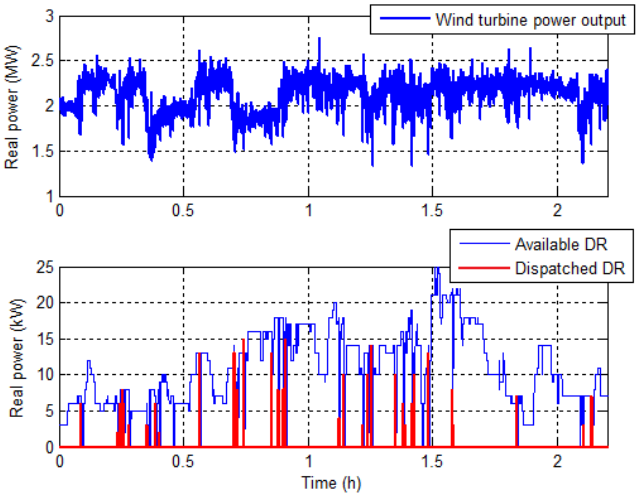


Figure 6. Operation of the King Island DRG.

During the period of over 2 hours, the King Island power system was running in zero-diesel operation and the DRG was used under small and short dips in wind power generation. In this mode of operation, the power system controller prioritizes DRG dispatch, and thus postpones unnecessary start-ups of a diesel generator.

V. CONCLUSION

With a strong drive for energy conservation, demand-side management and demand response are becoming vital for the implementation of the smart grid concept. This paper outlines some experience obtained at the University of Tasmania, Hydro Tasmania and the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia in developing fast DR for isolated power systems.

The paper presented DR as an enabling technology for higher penetration of renewable energy in isolated power systems. These systems are often based on diesel generators. However, due to high costs of diesel fuel supply, the cost of electricity in isolated power systems is much high compared to conventional interconnected systems. This presents an incentive for introducing renewable energy generation in isolated power systems.

Unfortunately, the integration of renewable generation presents significant stability and reliability challenges due to their intermittency. The solution proposed in this paper is based on centralized two-way communication and control of residential and commercial loads. DR can be dispatched and confirmed within 1 s. The technology has been installed and successfully tested in an isolated power system on King Island in Australia.

ACKNOWLEDGMENT

The authors would like to thank power operators of the King Island system, Australia for sharing insights their extensive knowledge and experience in operating isolated power systems.

REFERENCES

[1] J. A. Short, D. G. Infield, and L. L. Freris, “Stabilization of Grid Frequency Through Dynamic Demand Control,” IEEE Transactions on Power Systems, vol. 22, no. 3, pp. 1284-1293, 2007.

[2] T. L. Vandoorn, B. Renders, L. Degroote, B. Meersman, and L. Vandevelde, “Active Load Control in Islanded Microgrids Based on the Grid Voltage,” IEEE Transactions on Smart Grid, vol. 2, no.1, pp. 139-151, 2011.

[3] M. de Groot, J. Forbes and D. Nikolic, “Demand Response in Isolated Power Systems”, Australasian Universities Power Engineering Conference, Proc. AUPEC 2013, Hobart, Tasmania, Australia, 29 Sep. – 3 Oct. 2013, pp. 1-6.

[4] D. Westermann and A. John, "Demand Matching Wind Power Generation With Wide-Area Measurement and Demand-Side Management," IEEE Transactions on Energy Conversion, vol. 22, no. 1, pp. 145-149, 2007.

 ISBN: 978-1-61208-406-0

[5] N. Rajakovic, D. Nikolic, and J. Vujasinovic, “Cost benefit analysis for implementation of a system for remote control and automatic meter reading,” Proc. PowerTech, 2009 IEEE Bucharest, 2009, pp. 1-6.

[6] A. J. Conejo, J. M. Morales, and L. Baringo, "Real-Time Demand Response Model," IEEE Transactions on Smart Grid, vol. 1, no. 3, pp. 236-242, 2010.

[7] A. S. Kowli and S. P. Meyn, "Supporting wind generation deployment with demand response," Pros. the IEEE/PES General Meeting, Detroit, MI, USA, 24-28 July, 2011, pp. 1-6, IEEE Catalog Number CFP11POW-USB, ISBN 978-1-4577-1001-8.

[8] B. J. Kirby, "Spinning Reserve From Responsive Loads," Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, March 2003.

[9] D. S. Callaway and I. A. Hiskens, "Achieving Controllability of Electric Loads," *Proceedings of the IEEE,* vol. 99, pp. 184-199, 2011.

[10] S. A. Pourmousavi and M. H. Nehrir, "Real-Time Central Demand Response for Primary Frequency Regulation in Microgrids," *Smart Grid, IEEE Transactions on,* vol. 3, pp. 1988-1996, 2012.

[11] J. Kumagai, “Virtual power plants, real power,” IEEE Spectrum, vol. 49, pp. 13-14, 2012.

[12] Hydro Tasmania. (2013). King Island Renewable Energy Integration Project (KIREIP). [Online]. Available from: www.kireip.com.au 2015.04.10

[13] D. Nikolic, M. Negnevitsky, M. de Groot, S. Gamble, J. Forbes and M. Ross, “Fast Demand Response as an Enabling Technology for High Renewable Energy Penetration in Isolated Power Systems”, Pros. the IEEE/PES General Meeting, Washington DC, USA, 27-31 July, 2014, pp. 1-5, IEEE Catalog Number CFP14POW-USB, ISBN 978-1-4799-6414-7.

 ISBN: 978-1-61208-406-0

Towards a Quality of Service Based Complex Event Processing in Smart Grids

Epal Njamen Orleant
University of Grenoble, LIG
Saint Martin d'Hères, France
Email: Orleant.Epal-Njamen@imag.fr

Christine Collet
Grenoble INP, LIG
Saint Martin d'Hères, France
Email: christine.collet@grenoble-inp.fr

Genoveva Vargas Solar
LIG-LAFMIA
Saint Martin d'Hères, France
Email: genoveva.vargas@imag.fr

***Abstract*—This paper investigates distributed event processing in smart grids, considering the following quality of service (QoS) dimensions: event priority, network occupation, memory occupation, and notification latency. It proposes an architecture for a QoS based distributed event processing. Our approach considers that the event processing is implemented as a network of operators that are executed by distributed event processing units. We also investigate on strategies used by event processing units in order to address QoS requirements.**

Keywords–Complex event processing; Quality of service; Smart Grids.

I. INTRODUCTION

Computer systems are more and more distributed (smart grids, sensor networks, cloud based applications, etc). In general, the complexity to manage or supervise distributed systems increases with the number and type of participanting systems (potentially geographically separated), which continuously generate data. Those data can be considered as events that refer to happenings of interest produced within the system environment.

In most cases, the capacity of monitoring and supervising a distributed system relies on the capacity to process low level events, for inferring higher level events, semantically rich for end user applications. This process includes events filtering, aggregations, correlations, windowing, and other computations on events. Infrastructures able to achieve this are referred to as complex event processing systems[1–5].

For example, in a smart grid, smart meters and sensors generate different types of event streams. Let us consider for example CoverOpenAlert and BadVoltage event types, the former being generated each time the cover of a smart meter is opened, and the latter being generated each time a bad voltage is detected by a sensor over an electrical line. An application may be interested in the sequence of CoverOpenAlert and a BadVoltage occurring at the same place, within a two minutes time window. This pattern detects suspicious activities (MeterSuspected event type) on smart meters. The detection of such a high level event includes event filtering (type and attribute based filtering), windowing and temporal correlation.

The production of event streams in distributed contexts, associated with the need to quickly process them to have an aggregated view of a system's state, requires the definition of complex event processing systems like [1–5]. Moreover, those systems should be able to be deployed in distributed architectures. They must efficiently achieve event filtering, correlation, aggregation and composition while adapting to their environment in terms of the multiplicity of data sources (sensors, smart meters, existing databases, etc.) and applications quality of service (QoS) [6].

a) Multiplicity of data sources: Distributed systems like smart grids consist of different types of components that can act as event producers or consumers, with different interaction modes (synchronous or asynchronous, push or pull based style), as illustrated by sensors, smart meters, existing databases. The diversity of interaction modes, coupled with the difference in data formats make it difficult to integrate data from different producers for event processing purposes.

b) Applications quality of service (QoS): The need to detect and notify complex events from basic events is sometimes correlated with some quality of service requirements like latency, memory consumption, network occupancy, event priority, notification latency, etc. Those QoS requirements generally constrain the way the event processing must be achieved. In addition, they are not independent of each other. For example, the reduction of network occupancy generally decreases the notification latency. Therefore, the trade-offs among these QoS metrics has to be done according to application requirements. Existing systems are limited in the sense that they fail to associate QoS preoccupations to event processing.

The problem we address in this paper can be summarized as follows: given smart grid needs in terms of event composition and QoS, how to provide the complex event processing system that best fulfills expected QoS requirements?

Our approach considers an event based abstraction of smart grids functions and services. This abstraction allows to reason on the smart grid in terms of event streams that are generated by smart grid components. In order to identify relevant or critical situations (complex events) among those event streams, we propose a distributed complex event processing architecture. The event processing logic is implemented as a network of operators that have to be executed by distributed event processing units. We also investigate on strategies applicable to event processing units in order to address the following QoS dimensions: event priority, network occupation, memory occupation and notification latency.

The remainder of the paper is organized as follows: Section II presents some related works. Section III presents the overview of our approach for QoS based complex event processing in smart grids, Section IV presents the model and architecture of our approach. Section V discusses the QoS adoption in event processing and finally, Section VI concludes the paper.

 ISBN: 978-1-61208-406-0

II. RELATED WORKS

Many works have been achieved on event streams analysis and composition, and many event processing systems have been proposed so far [1–5], either for centralized or distributed architectures.

In centralized architectures, the generated events are processed by a single node acting as an event processing server [2][4][5][7][8]. This requires event streams to be routed to that server node, which increases the latency of the event processing, overloads the network and the server, which can become a point of failure. Therefore, this approach is not suited for distributed contexts.

In distributed architectures, the event processing logic is performed by a set of distributed communicating nodes, each one achieving a part of the work. This offers a better scalability and availability than centralized approaches. Some distributed event processing systems are [1][3][9][10].

References [11] and [6] identify some QoS dimensions (latency, priority, etc.) relevant for distributed event processing, but they do not propose mechanisms for their adoption. However, some other systems provide QoS support. They optimize the query processing according to a particular objective, and differs from each others by the adopted QoS dimension. For example, [1] has focused on reducing the network traffic whereas [9] studied energy consumption. In wide networking environments, it is not reasonable to expect that all applications share the same objective. In our approach, we identify a set of QoS properties relevant for event processing in smart grids, and we study their adoption by the event processing system.

Reference [12] presents a survey on the QoS requirements of the smart grid communications system. It focuses on the functionalities that have to be provided by the smart grid communication infrastructure in order to address application requirements. References [13] and [14] study the QoS adoption in the smart grid communication network. The former proposes to add QoS by providing differentiated service for data traffic with different priority at the MAC (Media Access Control) layer, while the latter propose GridStat, a publish-subscribe middleware framework that has been designed to meet the QoS requirements for the electric power grid. In our work, we assume the existence of QoS support at the networking layer (e.g, message priority) in order to propose a complex event processing system in smart grid that deals with event priority, memory occupation, network occupation and notification latency.

III. APPROACH OVERVIEW

Our approach to integrate complex event processing technologies into smart grids is presented in Figure 1. It consists in three layers of abstraction, namely the smart grid, event streams, and event processing network layers.

- The smart grid layer consists in the real physical smart grid architecture, which includes telecommunication based devices such as smart meters, sensors, data concentrators, etc. Those devices are connected by communication networks technologies including power line communications, wireline communications or wireless communications [15]. The smart grid is described in terms of information being used and exchange between functions, services and components. This layer of abstraction is referred to as the *Information layer* in the smart grid reference architecture model [16]. In our approach, information is seen as events that happen within the smart grid.
- The event stream layer, which considers that data generated by smart grids components are event streams. In that level, smart grid components act as sources which can generate different types of events in a continuous manner. The event type and event stream models considered in this work are presented in Section IV.
- The event processing network layer consists in a set of distributed event processing units which are connected by event channels. It is created according to complex event subscriptions, and its deployment may be distributed across multiple physical networks, computers and software artifacts. The complex event subscriptions can be tagged with applications QoS requirements such as event priority and notification latency. Those QoS requirements have to be translated into contraints applicable to event processing units at execution time. In addition to those constraints derived from QoS requirements, the smart grid infrastructure itself (processing devices and network technologies) dictates other constraints such as ressources (memory, CPU) limitations and network occupation limitations.

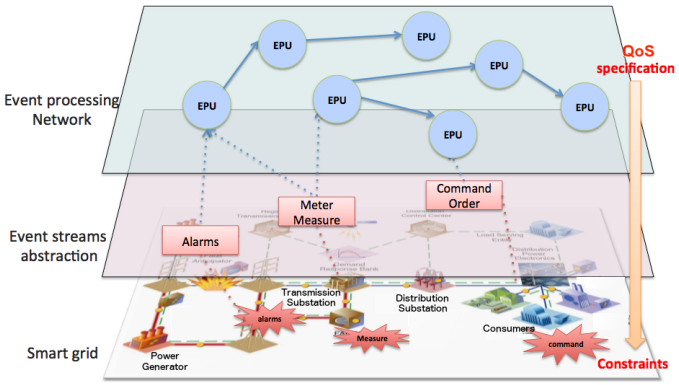


Figure 1. Approach overview

IV. MODEL AND ARCHITECTURE

This section presents the event model (event type and event stream), and the runtime architecture of our approach (event processing network).

A. Event Model

1) Event Type: An event type represents a class of significant facts (events) and the context under which they occur. The defintion of an event type includes the attributes presented in Table I.

The *typeName* attribute refers to the name of the event type. The *producerID* attribute refers to the id of the entity who produced the event occurence. The *detectionTime* attribute refers to the time at which the event occurence has been detected by a source. The *productionTime* attribute refers to the time at which the event has been produced (as a result of a processing on others events) by an event processing unit.

 ISBN: 978-1-61208-406-0

TABLE I. EVENT TYPE ATTRIBUTES

Name	Type
typeName	String
producerID	String
detectionTime	Number
productionTime	Number
notificationTime	Number
receptionTime	Number
priority	Number
context	Set<Attribute >

The *notificationTime* attribute refers to the time at which the event is notified to interested consumers. The *receptionTime* refers to the time at which the event is received by an interested consumer. The *priority* attribute represents the priority value associate to the event occurrence. The context (*context* attribute) of an event type defines all the attributes that are particular to this event type. They represent the others data manipulated by the producer which are relevant to this event type. For example, the context of a *MeterMeasure* event type generated by a smart meter includes the *voltage* and *current* attribute.

An event instance (or simply event) is an occurrence of an event type. An event type can be simple or composite.

Simple event types are event types for which instances are generated by producers (sensors, smart meters, etc.). They are not generated as results of processings on others events. In the example considered in Section I, *BadVoltage* and *CoverOpenAlert* are simple event types. More generally in a smart grid, the event types include *Alarms*, *MeterMeasure* and *SensorMeasure* generated by electric devices and such as smart meters and sensors, and *Command*, *ControlOrder*, *ControlAction* generated by utility applications.

Complex (or composite) event types are event types for which instances are generated as results of processings on others events. Reference [17] includes a set of operators applicable to events. They capture particular situations (relevant or critical) that can be inferred from occurrences of others events. Those situations have to be notified to utility applications so that the system can be automatically or manually controlled. In the same example, *MeterSuspected* is a complex event type. Complex event types can also capture aggregated values, like the daily electricity consumption of a household. This can be generated as result of an aggregation on a one-day window of *MeterMeasure* event instances.

2) Event Stream: An event stream is a continuous, append-only sequence of events. We note $Stream(s,T)$ the stream of events of type T generated by the source s. If S is a set of sources, then $\{\bigcup stream(s,T), s \in S\}$ defnes a stream of events of type T, denoted $Stream(T)$.

B. Event Processing Network

As introduced in Section III, the event processing logic is implemented by the event processing units. The runtime deployment of event processing units with associated event channels is called the event processing network [18][19]. This is illustrated in Figure. 2.

The general vision of our QoS based complex event processing system can be briefly described as follows: applications subscribe to composite events by issuing complex event patterns to the system, with associated QoS requirements. The system then deploys a set of distributed event processing units, which apply different strategies to meet QoS requirements during event processing. The complex events generated by the event processing units are notified to consumers. In a smart grid, such an infrastructure can act as a middleware on which utility applications can rely for detecting interesting or critical situations (sensors errors, alarms, etc.) over the electrical grid, with some QoS guarantees (e.g., priority, notification latency, etc.).

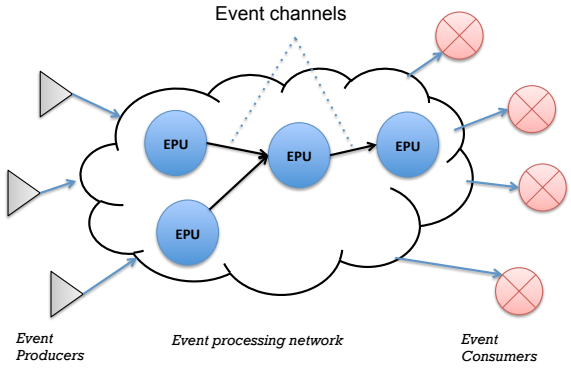


Figure 2. Event processing network

1) Event Processing Unit: An event processing unit can be defined by three types of components (see Figure 3):

- a set of input queues, on which parts of input event streams are maintained.
- an operator, which implements a three step event processing logic: *fetch-produce-notify*. In the first step (fetch), some events are selected from the input queues and marked as ready to be used to produce new composite events. In the second step (produce), the events selected at step 1 are used to produce new composite events according to the operator semantic. The composite events produced are stored in the output queue. In the third step (notify), events in the output queue are notified either to other event processing units or to interested consumers.
- an output queue, which contains events to be notified.

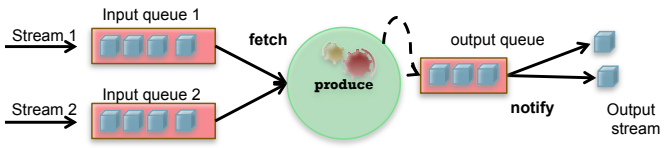


FIGURE 3. EVENT PROCESSING UNIT

2) Event Channel: Event processing units communicate through event channels. Event channels are means of conveying events [20]. This can be done via standard tcp or udp connections, or a higher level communication mechanism like publish/subscribe [21] or group communication [22] provided by a middleware layer.

V. QOS SUPPORT IN EVENT PROCESSING

A. QoS Dimensions

The need to detect and notify complex events from basic events is sometimes correlated with some QoS requirements. The QoS dimensions we proposed to address in this paper are event priority, memory occupation, network occupation and notification latency.

1) Event Priority: Event priority defines a priority order between events. In some contexts, there may exist priorities between events that have to be captured by the event processing runtime. For example in a smart grid, a BadVoltage event can be higher priority than a CoverOpenAlert event. Events that have a higher priority have to be processed and notified earlier than less priority events.

2) Memory Occupation: Differents devices may have different available memory capacity. To adapt the event processing to the memory capacity of each device, it must be a way to specify the maximum memory occupation incurred by an event processing unit at the execution time. The memory occupation constraint gives an upperbound on the number of events that an event processing unit can maintain in its main memory at execution time.

3) Network Occupation: Event processing units can be distributed accross different locations. The communication between those event processing units is achieved by messages sending at the notification step. The underlying network may be overloaded by a high event notification rate among event processing units. The network occupation constraint gives an upperbound on the number of networked event notifications per time unit on a given event processing unit.

4) Notification latency: In the common practice for power device protection, the circuit breaker must be opened immediately if the voltage or current on a power device exceeds the normal values. The notification latency of an event is the time elapsed between its production and its notification to interested consumers (end users or event processing units). The notification latency constraint imposed on an event processing unit defines an upper bound on the notification latency of events produced by that event processing unit.

B. QoS Adoption in Event Processing

Event processing networks are created from application event subscriptions. Those event subscriptions are specifications of event types (simples or complex) that applications are interested in, with their associated QoS requirements. The QoS requirements defined by smart grid applications constrain the way event processing and notification must be achieved. To address those application level QoS requirements, we have to provide:

- an operator placement algorithm that ensures load balance between the available processing devices while minimizing the end to end latency. A first step will be to have a look at works presented in [23][24], and adapt them in our context.
- strategies applicable to event processing units allowing to ensure that high priority events will be processed and notified earlier than less priority events. An idea consists in adopting priority queues for implementing event processing units input/output queues, and provide a mapping function between event priotity and message prioritization, if available in the underlying networking technology.
- strategies to adapt the event processing unit ressources (memory and network) utilisation in a consistent way wich respect to existing notification latency constraints.

VI. CONCLUSION

This paper shows that monitoring of smart grids can be done using an event based approach where event streams generated by distributed sources are processed by distributed event processing units. Such units may produce complex events indicating interesting or critical situations that are notified to interested consumers. Since the invocation of business critical processes is now triggered by events, the QoS of the event processing infrastructure becomes a key issue. We have identified key QoS dimensions relevant to smart grids, and proposed the first step towards a QoS based event processing system.

ACKNOWLEDGMENT

This work was carried out as part of the SOGRID project (www.so-grid.com), co-funded by the French agency for Environment and Energy Management (ADEME) and developed in collaboration between participating academic and industrial partners.

REFERENCES

[1] G. Cugola and A. Margara, "RACED : an Adaptive Middleware for Complex Event Detection," Proceedings of the 8th International Workshop on Adaptive and Reflective MIddleware, 2009, pp. 1–6.

[2] "Homepage of Esper," 2015, URL: http://esper.codehaus.org/ [Accessed: 2015-03-27].

[3] "Homepage of TIBCO StreamBase," 2015, URL: http://www.streambase.com/ [accessed: 2015-03-27].

[4] D. Gyllstrom, E. Wu, H.-J. Chae, Y. Diao, P. Stahlberg, and G. Anderson, "SASE: Complex Event Processing over Streams," 3rd Biennial Conference on Innovative Data Systems Research (CIDR), 2006, pp. 1–5.

[5] "Homepage of Oracle CEP," 2015, URL: http://www.oracle.com/technetwork/middleware/complex-event-processing/overview/index.html [accessed: 2015-03-26].

[6] S. Appel, K. Sachs, and A. Buchmann, "Quality of service in event-based systems," in CEUR Workshop Proceedings, vol. 581, 2010, pp. 1–5.

[7] A. Demers, J. Gehrke, B. Panda, M. Riedewald, V. Sharma, and W. White, "Cayuga : A General Purpose Event Monitoring System," CIDR 2007, Third Biennial Conference on Innovative Data Systems Research, 2007, pp. 412–422.

[8] D. Luckham, "Rapide: A Language and Toolset for Simulation of Distributed Systems by Partial Orderings of Events," Stanford University, Tech. Rep., 1996.

[9] O. Saleh and K.-U. Sattler, "Distributed Complex Event Processing in Sensor Networks," 2013 IEEE 14th International Conference on Mobile Data Management, Jun. 2013, pp. 23–26.

[10] P. Pietzuch, B. Shand, and J. Bacon, "A framework for event composition in distributed systems," Proceedings of the ACM/IFIP/USENIX, vol. 2672, 2003, pp. 62–82.

 ISBN: 978-1-61208-406-0

[11] S. Behnel, L. Fiege, and G. Mühl, "On quality-of-service and publish-subscribe," in Proceedings - International Conference on Distributed Computing Systems, 2006, pp. 1–6.

[12] Y.-h. Jeon, "QoS Requirements for the Smart Grid Communications System," Journal of Computer Science and Network Security, vol. 11, no. 3, 2011, pp. 86–94. [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6038941

[13] W. S. W. Sun, X. Y. X. Yuan, J. W. J. Wang, D. H. D. Han, and C. Z. C. Zhang, "Quality of Service Networking for Smart Grid Distribution Monitoring," Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on, 2010, pp. 373–378.

[14] H. Gjermundrø d, D. E. Bakken, C. H. Hauser, and A. Bose, "GridStat: A flexible QoS-managed data dissemination framework for the power grid," IEEE Transactions on Power Delivery, vol. 24, no. 1, 2009, pp. 136–143.

[15] W. Wang, Y. Xu, and M. Khanna, "A survey on the communication architectures in smart grid," Computer Networks, vol. 55, no. 15, Oct. 2011, pp. 3604–3629.

[16] CEN-CENELEC-ETSI Smart Grid Coordination Group, "Smart grid reference architecture," 2012, URL: http://ec.europa.eu/energy/gas_electricity/smartgrids/doc/xpert_group1_reference_architecture.pdf [accessed: 2015-03-27].

[17] G. Cugola and A. Margara, "Processing flows of information: From data stream to complex event processing," ACM Computing Surveys (CSUR), vol. V, no. i, 2012, pp. 1–70.

[18] L. Perrochon, W. Mann, S. Kasriel, and D. C. Luckham, "Event Mining with Event Processing Networks," in Methodologies for Knowledge Discovery and Data Mining. Third Pacific-Asia Conference, PAKDD-99 Beijing, China, April 2628, 1999 Proceedings, 1999, pp. 474–478.

[19] G. Sharon and O. Etzion, "Event-processing network model and implementation," IBM Systems Journal, vol. 47, no. 2, 2008, pp. 321–334.

[20] "Event processing glossary version 2.0," 2011, URL: http://www.complexevents.com/2011/08/23/event-processing-glossary-version-2/ [accessed: 2015-03-27].

[21] P. T. Eugster, P. A. Felber, R. Guerraoui, and A.-M. Kermarrec, "The many faces of publish/subscribe," ACM Computing Surveys, vol. 35, no. 2, 2003, pp. 114–131.

[22] G. V. Chockler and R. Vitenberg, "Group Communication Specifications : A Comprehensive Study," ACM Computing Surveys, vol. 33, no. 4, 2001, pp. 427–469.

[23] G. T. Lakshmanan, Y. Li, and R. Strom, "Placement Strategies for Internet-Scale Data Stream Systems," IEEE Internet Computing, vol. 12, no. 6, Nov. 2008, pp. 50–60.

[24] C. Thoma, A. Labrinidis, and A. J. Lee, "Automated operator placement in distributed Data Stream Management Systems subject to user constraints," 2014 IEEE 30th International Conference on Data Engineering Workshops, Mar. 2014, pp. 310–316.

 ISBN: 978-1-61208-406-0

Sensor Placement for Real-Time Power Flow Calculations in Transmission Networks

E. Feinberg, M. Li, R. Samulyak

Stony Brook University
Stony Brook, NY
Email: Eugene.Feinberg@stonybrook.edu
Email: Muqi.Li@stonybrook.edu
Email: Roman.Samulyak@stonybrook.edu

B. Fardanesh, G. Stefopoulos

New York Power Authority
White Plains, NY
Email: Bruce.Fardanesh@nypa.gov
Email: George.Stefopoulos@nypa.gov

***Abstract*—This paper describes a sensor placement algorithm for real-time parallel power flow computations for transmission networks. In particular, Phasor Measurement Units (PMUs) can be such sensors. Graph partitioning is used to decompose the system into several subsystems and to locate sensors in an efficient way. Power flow calculations are then run in parallel for each area. Test results on the IEEE 118- and 300-bus systems show that the proposed approach is suitable for real-time applications.**

Keywords–Parallel computing; power system; power flow calculation; PMU placement.

I. INTRODUCTION

Fast load flow analysis is essential for the successful implementations of advanced real-time control of transmission systems.The natural tool for this is parallel computing. To this end, various parallel methods have been proposed. Many traditional approaches [1][2][3] use factorization and the forward-backward solution of linear equations to achieve parallelism. As many serial computations are needed for such approaches, their parallel efficiency is not high.

Another approach to alleviating the computational burden is to decompose a large problem into a number of small problems and perform computations for each sub-problem in parallel. These smaller sub-problems are usually coordinated by a master process. Rafian et al. [4] presented a method for load-flow analysis based on tearing the network into 2-3 subsystems. In every iteration, the subsystems are solved in parallel and will communicate with a coordinating program. After the communication, the coordinating program is then conducted to determine the global solution for the original system. Chan [5] et al. proposed a parallel solution based on piecewise method. The Jacobian matrix is converted into a bordered block diagonal form. The block diagonal form leads to subproblems that can be solved independently, after which a problem corresponding to the border is solved to coordinate the subproblems and obtain a solution to the original problem. Amano et al. [6] employed a block-parallel method for load-flow analysis. In particular, the Jacobian matrix is constructed by applying the epsilon decomposition algorithm, which eliminates weak coupling elements from the matrix. One drawback of this approach is that the transformation of the matrix into a balanced diagonal matrix takes time; further, the speed of convergence is affected by the choice of partition method.

Traditional power flow calculations use only power injection measurements as the input. However, at the present time it is also possible to have synchronized voltage measurements at buses. For example, it is possible to use Phasor Measurement Units (PMUs) for this. These devices provide accurate real-time measurements at multiple points on the grid. A number of significant improvements in control and analytical capabilities have been made possible by this technology. However, due to their advanced features and the need for communications infrastructure, PMUs are relatively costly to implement and maintain. Hence it is typically not economical to install a PMU at each bus [7]. This has motivated a growing literature on the optimal placement of PMUs, and various PMU placement algorithms have been developed for different situations. Xu and Abur [8] have modeled the PMU placement problem as an integer linear programming problem with the constraint that the entire power system should be observable. A mixed integer linear programming formulation was introduced by Aminifar et al. [9] for PMU placement that accounts for line and PMU contingencies. Gou [10] presented a model that allows for redundant PMU placement and incomplete observability, which can be solved by integer linear programming. Chakrabarti and Kyriakides [11] proposed an exhaustive search algorithm for PMU placement that takes so-called zero injection buses, i.e., buses with neither generation nor load, into account. In particular, systems with such buses may need fewer PMUs.

This paper investigates the possibility of using additional sensor measurements, e.g., from PMUs, to decompose the power system into several parts and conduct power flow calculation in parallel. The idea of using PMU is motivated by decomposition method, e.g., [4][5], for solving power flow problem. Instead of using a master process to coordinate the subsystems in each iteration, we achieve the coordination directly with PMU measurements. Using this approach, no data needs to be transferred between subsystems, which results in a simple and efficient solution. The placement of PMUs is based on solving a graph partition problem. By placing a relatively small number of additional PMUs, a large power system can be divided into several non-overlapping smaller subsystems that can be solved in parallel. Numerical results on 118 and 300 bus power networks demonstrate the efficacy of the proposed approach in significantly reducing the computation time.

The rest of the paper is organized as follows. Section II

 ISBN: 978-1-61208-406-0

briefly describes the power flow problem, and Section III introduces the proposed method. Sections IV describes the details of PMU placement. Section V illustrates the implementation of the algorithm via numerical examples on two power systems.

II. Power flow Analysis

Power flow analysis, commonly known as load flow analysis, is an important part of power system analysis. The goal of the power flow analysis is to obtain complete voltage angle and magnitude information for each bus in a power system under balanced three-phase steady state conditions. The power balance equations can be written as:

$$P_i = \sum_{j=1}^{n} |V_i||V_j|(G_{ij}cos\theta_{ij} + B_{ij}sin\theta_{ij}), \quad (1)$$

$$Q_i = \sum_{j=1}^{n} |V_i||V_j|(G_{ij}sin\theta_{ij} - B_{ij}cos\theta_{ij}). \quad (2)$$

Here, P_i and Q_i are the real and reactive power injections at bus i; $|V_i|$ is the bus voltage magnitude at bus i; θ_{ij} is the voltage angle difference between buses i and j; G_{ij} and B_{ij} are the real and imaginary parts of elements in the bus admittance matrix corresponding to buses i and j. Equations (1) and (2) constitute a set of nonlinear equations, and the number of equations is approximately twice the number of network buses. Expanding the above equations using Taylor series and ignoring the higher order terms results in the following equations:

$$\begin{bmatrix} \Delta P_2 \\ \vdots \\ \Delta P_n \\ \Delta Q_2 \\ \vdots \\ \Delta Q_n \end{bmatrix} = \begin{bmatrix} \frac{\partial P_2}{\partial \theta_2} & \cdots & \frac{\partial P_2}{\partial \theta_n} & \frac{\partial P_2}{\partial |V_2|} & \cdots & \frac{\partial P_2}{\partial |V_n|} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial P_n}{\partial \theta_2} & \cdots & \frac{\partial P_n}{\partial \theta_n} & \frac{\partial P_n}{\partial |V_2|} & \cdots & \frac{\partial P_n}{\partial |V_n|} \\ \frac{\partial Q_2}{\partial \theta_2} & \cdots & \frac{\partial Q_2}{\partial \theta n} & \frac{\partial Q_2}{\partial |V_2|} & \cdots & \frac{\partial Q_2}{\partial |V_n|} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial Q_n}{\partial \theta_2} & \cdots & \frac{\partial Q_n}{\partial \theta_n} & \frac{\partial Q_n}{\partial |V_2|} & \cdots & \frac{\partial Q_n}{\partial |V_n|} \end{bmatrix} \begin{bmatrix} \Delta \theta_2 \\ \vdots \\ \Delta \theta_n \\ \Delta |V_2| \\ \vdots \\ \Delta |V_n| \end{bmatrix} \quad (3)$$

Equation (3) can be written in the form:

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = J \begin{bmatrix} \Delta \theta \\ \Delta |V| \end{bmatrix}, \quad (4)$$

where J is the Jacobian (or the matrix of partial derivatives) displayed in (3). The linearized system of equations is solved to determine the next approximation of voltage magnitude $|V|$ and angle θ, and the process continues until a stopping condition is met.

III. Proposed Real-Time Power Flow Calculation Method

In this paper, we propose an approach for power flow calculation based on power system decomposition and PMU placement. The proper placement of PMUs will decrease the dimensions of the sub-problems and improve the efficiency of computational procedures. Our method contains following steps:

A. Step 1: Partition Power System and PMU Placement

A power system is decomposed into k non-overlapping subsystems of approximately the same size using a graph partition algorithm. PMUs will be installed at selected boundary buses to make sure the voltage phasors at all boundary buses are known. In other words, PMUs are placed to make all boundary buses observable [12]. Given a particular bus, installing a PMU obviously makes that bus observable. In addition, all adjacent buses to that bus become observable since their voltage phasors can be calculated from the transmission line current measured by the PMU and transmission line parameters. The procedure of power system decomposition and PMU placement will be described in Section IV.

Let S denote the set of buses in a power system. Suppose this power system is decomposed into k subsystems. We denote by S_i the set of buses in subsystem $i = 1, \ldots, k$, where $S_i \cap S_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{k} S_i = S$. Thus, each bus in the subsystem i belongs to one of the following types:

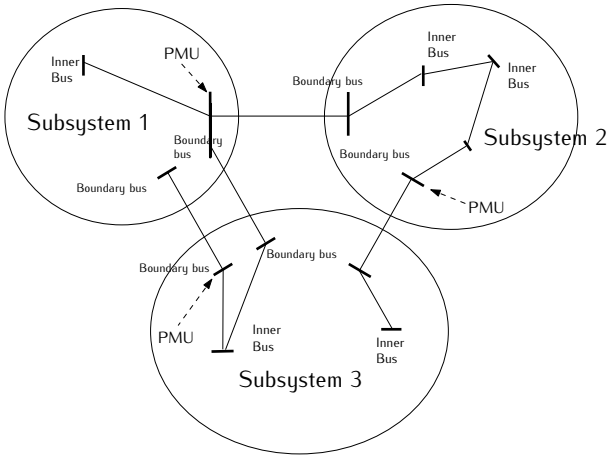


Figure 1. Power system decomposition and bus classification for k =3

- Inner Bus: All its neighboring buses also belong to subsystem i.
- Boundary Bus: At least one of its neighboring buses belongs to a different subsystem.

This is illustrated in Figure 1.

B. Step 2: Select Reference Buses

In order to perform power flow analysis for each subsystem, we need to choose a reference bus for each subsystem. Since the PMUs are placed in such a way that all boundary buses are observable, the voltage magnitude and phase angle for each boundary bus is known. This means that any boundary bus can serve as a reference bus for the subsystem it belongs to. We experimented with performing power flow calculation using different boundary buses as reference buses, and found that the computation time was not affected by the different selection of reference buses.

C. Step 3: Update Power at Boundary Buses

Before subsystem i can be solved independently of the others, the real and reactive power at each boundary bus needs to be updated to account for power flows from neighboring subsystems.

 ISBN: 978-1-61208-406-0

In particular, consider a boundary bus b that belongs to subsystem i. Let P_b and Q_b be the real and reactive power, respectively, at the boundary bus b. For any bus c connected to b, let $P_{b,c}$ and $Q_{b,c}$ be the real and reactive line power flow, respectively, on the line from bus b to bus c. Also, let $A(b)$ denote the set of boundary buses connected to bus b and not belonging to subsystem i. For any boundary bus b_1 connected to b the line power flows P_{b,b_1} and Q_{b,b_1} can be calculated from PMU measurements. The power flowing between bus b and its neighbors in subsystem i can be adjusted as follows:

$$P_b^{new} = P_b - \sum_{c \in A(b)} P_{b,c},$$
$$Q_b^{new} = Q_b - \sum_{c \in A(b)} Q_{b,c}.$$

D. Step 4: Calculation for Subsystems and Aggregate the Results

Power flow calculations are performed for each subsystem with respect to its reference bus. The voltage phasors of each reference bus are determined by PMU measurement. After each subsystem is solved in parallel, the solution to the entire system is then obtained by aggregating the solutions for each subsystem.

It is possible that some of the subsystems consist of multiple connected components. In this case the number of connected component m is greater than k. In our computational experiments, this increase of components does not significantly increase the computational time.

IV. PMU Placement Algorithm

The proposed real-time power flow calculation method decomposes a large power system into several subsystems, which are then solved in parallel. The implementation cost associated with such a decomposition depends on the number of PMUs installed at boundary buses. Having roughly the same number of buses per subsystem is desirable from the standpoint of balancing computational load. Thus our objective is to minimize the number of PMUs installed at boundary buses subject to the condition that each subsystem has approximately the same number of buses.

The basic structure of the PMU placement algorithm contains two steps. The first step is to divide the power system into k parts using a graph partitioning algorithm. This algorithm attempts to minimize the number of lines whose incident buses belong to different subsystems, while keeping the number of buses per subsystem approximately equal. The second step is to place PMUs based on power system decomposition. Given the partition k, the PMU locations are obtained by solving an integer linear programming problem.

A. Step 1: Power System Decomposition

Viewing the system as a graph $G = (V, E)$, where the vertex set V is the set of buses and the edge set E is the set of transmission lines, we obtain a k-way partitioning problem. The k-way partitioning problem divides a graph into k subgraphs with roughly the same number of vertices such that the edge cut, i.e., the number of edges connecting different subgraphs, is minimized. This problem is NP-hard, and several heuristics for its solution have been developed; see [13]. In this paper, we implement two heuristics for graph partitioning. One is a spectral partitioning algorithm, another one is a multilevel k-way partitioning algorithm.

Spectral partitioning: The spectral partitioning algorithm uses the eigenvectors of the adjacency matrix of a graph to find partitions. In this paper, we are using a spectral factorization based algorithm [14]. This spectral partitioning algorithm consists of four steps.

1) Form the Adjacency Matrix: For a power system with n buses, let $A = \{a_{i,j}\}$ be the corresponding $n \times n$ graph adjacency matrix, where $a_{ij} = 1$, when bus i and bus j are connected by a transmission line, and $a_{ij} = 0$ otherwise.
2) Adjacency Matrix Normalization: Normalize the nonnegative symmetric matrix A to obtain a doubly stochastic matrix A', i.e., $A' = \{a'_{ij}\}$ satisfies $\sum_i a'_{ij} = \sum_j a'_{ij} = 1$ for each j and i.
3) Compute Eigenvectors: Compute the k largest eigenvectors u_i, $i = 1, 2, ..., k$, of matrix A'. It is convenient to define the $n \times k$ matrix $U := [u_1, u_2, ..., u_k]$.
4) Clustering: Obtain a partition of the network into k subsystems by clustering the n rows of U into k clusters using the k-means algorithm [15].

Multilevel k-way partitioning: We used an implementation of the multilevel k-way partitioning algorithm from METIS [16]. The algorithm consists of three major steps:

1) Graph coarsening: Given the original graph $G_0 = (V_0, E_0)$, if weight information is not provided, it assumed that the vertex weights and edge weights are all equal to 1. Let $|V|$ be the number of vertices in V. A series of successively smaller graphs $G_i = (V_i, E_i)$, $i = 1...m$, is derived from the input graph such that $|V_{i-1}| > |V_i|$. Each successive graph G_i is constructed from the previous graph G_{i-1} by collapsing together a set of pairs of adjacent vertices; these sets can be obtained from finding a maximal matching. If two vertices are merged, the weights need to be updated in order to preserve the structure of the previous graph. The weight of the new vertex is set equal to the sum of the weights of its constituent vertices. If two merged vertices are adjacent to the same neighbor, then the two edges will be replaced by a new edge whose weight is the sum of weights of the edges it replaced. The graph coarsening step ends when the coarsest graph G_m is smaller than a given threshold.
2) Initial partitioning: A k-way partition of the coarsest graph G_m is computed such that each sub-graph contains roughly same vertex weight. This is done using a relatively simple approach such as the multilevel bisection algorithm.
3) Uncoarsening and refinement: The partition of the smallest graph G_m is projected back to G_0 through a successively larger graphs $G_{m-1}, G_{m-2}, ..., G_1$. This projection step reverses the process in step 1 to obtain the uncoarsen graph. After each projection step, the partition is refined using an algorithm based on the Kernighan-Lin method [17] that iteratively moves vertices between sub-graphs as long as such moves

improve the quality of the partition.

B. Step 2: PMU placement based on decomposition

Once a partition of the system is obtained, PMUs need to be installed to provide measurements for boundary buses. In most cases, the number of PMUs needed is not necessarily equal to the number of branches connecting different subsystems. Based on the approach in [8], the optimal PMU placement problem can be formulated as an Integer Linear Programming (ILP) problem.

Given a partition of the system with n boundary buses, we can define an n by n constraint matrix M for the boundary buses. The entries of M are defined as follows:

$$M_{i,j} = \begin{cases} 1, & \text{if } i = j \\ 1, & \text{if boundary buses } i \text{ and } j \\ & \text{are connected} \\ 0, & \text{otherwise} \end{cases}$$

The PMU placement problem can be formulated as follows:

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} c(i)x_i \\ \text{subject to} \quad & MX \geq \hat{1} \\ & x_i \in \{0,1\},\ i = 1, \dots n, \end{aligned} \tag{5}$$

where n is the number of boundary buses, $c(i)$ is the cost of placing a PMU at boundary bus i, $\hat{1}$ is a vector of ones, and X is a binary decision vector whose entries are:

$$x_i = \begin{cases} 1, & \text{if a PMU should be installed at boundary bus } i \\ 0, & \text{otherwise.} \end{cases}$$

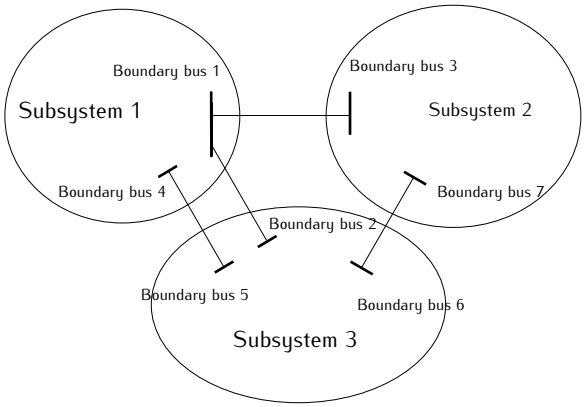


Figure 2. Example for optimal PMU placement

For example, consider the power system shown in Figure 2 that has been decomposed into three subsystems. There are 4 transmission lines between 7 boundary buses. The optimal PMU placement problem can be solved as follows: First, initialize the constraint matrix M for the boundary buses. Building the M matrix for Figure 2 yields:

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Then the inequalities in (5) takes the following form:

$$\begin{cases} x_1 + x_2 + x_3 \geq 1 \\ x_1 + x_2 \geq 1 \\ x_1 + x_3 \geq 1 \\ x_4 + x_5 \geq 1 \\ x_4 + x_5 \geq 1 \\ x_6 + x_7 \geq 1 \\ x_6 + x_7 \geq 1 \end{cases}$$

The first constraint means that a PMU should be placed either on bus 1, bus 2 or bus 3 to make bus 1 observable. Similarly, the second constraint implies that a PMU should be installed at either bus 1 or 2 to make bus 2 observable. After solving the ILP problem for the boundary buses, the placement of PMUs for the entire system is obtained.

V. Numerical Results

We use the IEEE 118 and 300 bus systems from [18] to illustrate the performance of our parallel algorithms. The 118 bus system was decomposed into 2, 4, and 8 subsystems, while the 300 bus system was decomposed into 2, 4, 8, and 16 subsystems. The power system was partitioned by both the spectral algorithm [14] and the multilevel k-way method [16] described in Section IV. The PMU measurements at the boundary buses were emulated using solutions obtained by traditional serial power flow methods. After decomposition, each of the resulting sub-networks was solved using Newton's method [19]. The convergence criterion was set to 10^{-5}, and the maximum iteration number for Newton's method was set to 10. The system information is shown in Table I. The partition results and computation times for different numbers of sub-networks are summarized in Tables II and III.

TABLE I. Network Information for two power system

System	118 bus	300 bus
number of nodes	118	300
number of branches	186	411
number of generator buses	53	68

TABLE II. Test Results on IEEE 118 bus system

Method	Partition number	Max subsystem size	Edge-Cut size [a]	PMU number	Calculation time (sec.)
Spectral Method	1	118	0	0	0.183
	2	79	5	3	0.090
	4	38	15	10	0.029
	8	22	26	13	0.016
Multilevel k-way	1	118	0	0	0.183
	2	78	5	3	0.089
	4	33	14	10	0.026
	8	19	29	14	0.014

[a] Edge-cut size: the number of branches connecting different subsystems

The results obtained using our methods were compared with the corresponding results obtained using serial Newton's method. The maximum deviation of node voltage in our method compared to Newton's method was less than 10^{-4} p.u., which illustrates that our method is accurate and feasible.

 ISBN: 978-1-61208-406-0

TABLE III. TEST RESULTS ON IEEE 300 BUS SYSTEM

Method	Partition number	Max subsystem size	Edge-Cut size [a]	PMU number	Calculation time (sec.)
Spectral Method	1	300	0	0	1.440
	2	184	6	5	0.570
	4	98	11	7	0.163
	8	56	19	13	0.075
	16	30	39	26	0.029
Multilevel k-way	1	300	0	0	1.440
	2	173	6	6	0.530
	4	97	11	10	0.162
	8	51	24	20	0.058
	16	25	49	32	0.025

[a] Edge-cut size: the number of branches connecting different subsubsystems

A. Speedups compared to serial method

The computation times for the power flow calculation are presented in Tables II and III. In particular, the computation time T_s corresponding to partition number 1, the case where no partitioning is done, was obtained by the serial Newton's method. The rate of speedup can be obtained by the formula $S = T_s/T_p$, where T_p is the computation time of parallel method. Figures 3 and 4 show these speedups. Comparing the the details of these timing studies, we present the following conclusions:

1. High speedups and parallel efficiency are achieved in both the 118 and 300 bus systems by spectral and multilevel k-way method. Using spectral method as example, when the system is divided into two subsystems, a speedup by about a factor of 2 was obtained for both systems. Installing more PMUs usually leads to faster power flow calculations. In the 300-bus system, the traditional serial method takes 1.440 seconds, while with 5 PMUs the calculation time is reduced to 0.570 seconds; here the speedup rate is 2.53. The speedup rate increases to 49.66 by using 26 PMUs. These speedups suggest that the parallel algorithm proposed in this paper is efficient.

2. The speedup associated with the parallel method increased as the size of the power system increased. For the case of splitting system into two subsystems, by using spectral method, the calculation time for the 118 bus system was sped up by a factor of 2.03, while for the 300 bus system it was sped up by a factor of 2.53. For the case of splitting system into four subsystems, the calculation time for the 118 bus system was 6.31 times faster, and for the 300 bus system it was 8.83 times faster. This suggests that the algorithms presented in this paper may perform well on larger power systems.

B. Performance of Spectral Partitioning vs. Multilevel k-way Partitioning

Different partition methods may result in different PMU placements. In particular, the practicality of the partitioning algorithm becomes especially important when real systems are considered. For large power systems, the optimal PMU placement problem is hard to solve exactly. A common practice is to compare the performance of different methods. Here we compared the spectral partitioning algorithm with the k-way partitioning method. As Figures 3 and 4 show, the multilevel k-way method outperforms the spectral partitioning algorithm in terms of calculation time. However, the multilevel k-way method usually requires more PMUs.

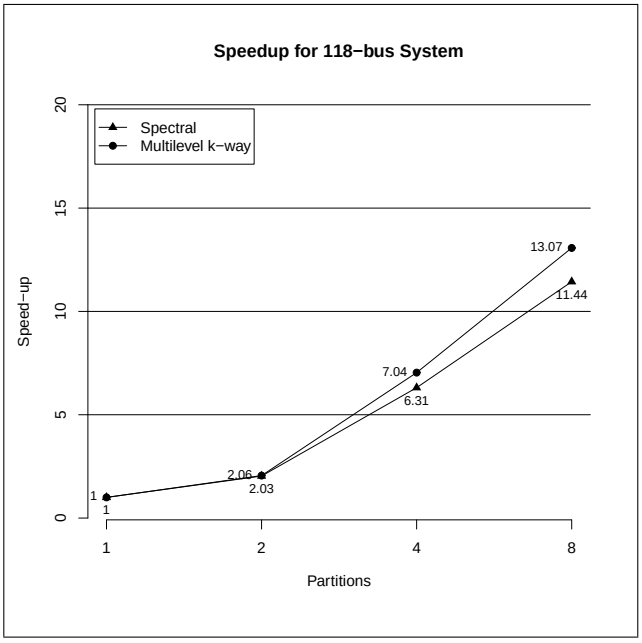


Figure 3. Speedup for 118 bus system

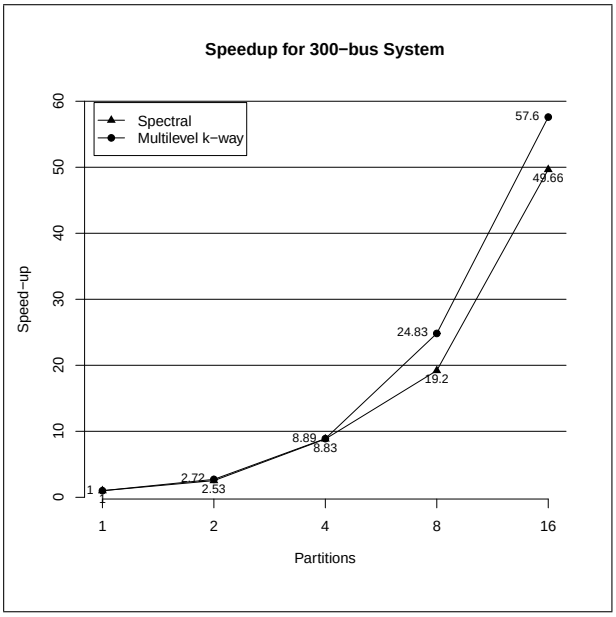


Figure 4. Speedup for 300 bus system

TABLE IV. COMPARISON OF THE SPECTRAL METHOD AND MULTILEVEL k-WAY METHOD FOR IEEE 300 BUS SYSTEM

Partition Number	Spectral method		Multilevel k-way method	
	Max Subsystem Size	PMU Number	Max Subsystem Size	PMU Number
2	184	5	173	6
4	98	7	97	10
8	56	13	51	20
16	30	26	25	32

 ISBN: 978-1-61208-406-0

The partition results obtained via the spectral method and the multilevel k-way method for the 300 bus system are listed in Table IV. As shown in Table IV, with the same partition number, the max subsystem size obtained using the multilevel k-way method is smaller than the one obtained using the spectral algorithm. It indicates that the multilevel k-way method allocates the buses to subsystems in a more balanced way than the spectral method. More balanced subsystems may result in faster calculation time; see Figures 3 and 4. On the other hand, these allocations require more PMUs. There is a tradeoff between having balanced subsystems and using a small number of PMUs, considering the high cost of PMU implementation and maintenance in the long run, the spectral method may be more appropriate for our decomposition scheme.

VI. Conclusion

In this paper, we have described a new approach to power flow analysis. Our problem formulation explicitly takes into account the placement of PMUs. A graph partition approach was proposed to partition the power system and determine the placement of PMUs, after which the power flow problem can be solved in parallel. The effectiveness of the approach was illustrated on a 118 and 300 bus system using two different graph partitioning methods. In each of these cases, significant speedups compared to the serial Newton's method were obtained. The spectral algorithm requires fewer PMUs while keeping the approximately the same number of buses for each subsystem, thus it is more suitable for our decomposition approach.

Acknowledgment

This study was supported by NYSERDA (New York State Energy Research and Development Authority) contract 28814.

References

[1] S. Lin and J. Van Ness, "Parallel solution of sparse algerbraic equations," IEEE Transactions on Power Systems, vol. 9, no. 2, 1994, pp. 743–749.

[2] Y. Wang and H. Gooi, "New ordering methods for sparse matrix inversion via diagonalization," IEEE Transactions on Power Systems, vol. 12, no. 3, 1997, pp. 1298–1305.

[3] J. Wu and A. Bose, "Parallel solution of large sparse matrix equations and parallel power flow," IEEE Transactions on Power Systems, vol. 10, 1995, pp. 1343–1349.

[4] M. Rafian, M. Sterling, and M. Irving, "Decomposed load-flow algorithm suitable for parallel processor implementation," IEE Proceedings, vol. 132, no. 6, 1985, pp. 281–284.

[5] K. Chan, R. Dai, and C. Cheung, "A Coarse Grain Parallel Solution Method for Solving Large Set of Power Systems Network Equations ," Proc. Int. Conf. Power System Technology, vol. 4, 2002, pp. 2640–2644.

[6] M.Amano, A. Zecevic, and D. Siljak, "An improved block-parallel newton method via epsilon decompositions for load-flow calculations," IEEE Transactions on Power Systems, vol. 11, no. 03, 1996, pp. 1519–1527.

[7] M.Esmaili, K.Gharani, and H. Shayanfar, "Redundant observability PMU placement in the presence of flow measurements considering contingencies," IEEE Transactions on Power Systems, vol. 28, no. 4, 2013, pp. 3765–3773.

[8] B.Xu and A.Abur, "Observability Analysis and Measurement Placement for Systems with PMUs," in Power Systems Conference and Exposition, 2004. IEEE PES, vol. 2, 2004, pp. 943–946.

[9] F.Aminifar, A.Khodaei, M.Fotuhi-Firuzabad, and M.Shahidehpour, "Contingency-constrained PMU placement in power networks," IEEE Transactions on Power Systems, vol. 25, no. 1, 2010, pp. 516–523.

[10] B.Gou, "Generalized Integer Linear Programming Formulation for Optimal PMU Placement," IEEE Transactions on Power Systems, vol. 23, no. 03, 2008, pp. 1099–1104.

[11] S.Chakrabarti and E.Kyriakides, "Optimal placement of phasor measurement units for power system observability ," IEEE Transactions on Power Systems, vol. 23, no. 3, 2008, pp. 1433–1440.

[12] R. Kavasseri and S. K. Srinivasan, "Joint Placement of Phasor and Power Flow Measurements for Observability of Power Systems," IEEE Transactions on Power Systems, vol. 26, no. 4, 2011, pp. 1929–1936.

[13] P.-O. Fjllstrm, "Algorithms for Graph Partitioning: A Survey," Linkping Electronic Articles in Computer and Information Science, 1998.

[14] J. Hespanha, "An efficient MATLAB Algorithm for graph partitioning," Technical Report, University of California, Santa Barbara, CA, 2004, [retrieved: 03, 2015]. [Online]. Available: http://www.ece.ucsb.edu/~hespanha/techreps.html

[15] J. MacQueen, "Some methods for classification and analysis of multivariate observations," Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., vol. 1, 1967, pp. 281–297.

[16] G. Karypis and V. Kumar, "A fast and highly quality multilevel scheme for partitioning irregular graphs," SIAM Journal on Scientific Computing, vol. 20, no. 1, 1999, pp. 359–392.

[17] B. W. Kernighan and S. Lin, "An Efficient Heuristic Procedure for Partitioning Graphs," Bell System Technical Journal, vol. 49, 1970, pp. 291–307.

[18] Power Systems Test Case Archive, [retrieved: 03, 2015]. [Online]. Available: https://www.ee.washington.edu/research/pstca

[19] H. Saadat, Power System Analysis. New York: McGraw-Hill, 1999.

 ISBN: 978-1-61208-406-0

A Matrix Model For An Energy Management System Based On Multi-Carrier Energy Hub Approach

Diego Arnone, Massimo Bertoncini,
Giuseppe Paternò, Alessandro Rossi
Research and Development Laboratory
Engineering Ingegneria Informatica S.p.A.
Palermo, Italy
e-mail: diego.arnone@eng.it, massimo.bertoncini@eng.it,
giuseppe.paterno@eng.it, alessandro.rossi@eng.it

Mariano Giuseppe Ippolito,
Eleonora Riva Sanseverino, Gaetano Zizzo
Dipartimento di Energia, Ingegneria dell'Informazione e
Modelli Matematici (DEIM)
Università degli Studi di Palermo
Palermo, Italy
e-mail: marianogiuseppe.ippolito@unipa.it,
eleonora.rivasanseverino@unipa.it, gaetano.zizzo@unipa.it

Sandra Elizabeth Jenkins
Engineering Systems Division, Technology and Policy Program '14
Massachusetts Institute of Technology
Cambridge, MA, USA
e-mail: sandjenkins@gmail.com

Abstract—**The INGRID FP7 European co-funded project studies several methodologies concerning hydrogen production and storage, aiming to provide services to electricity system operators for suitably balancing electrical supply and demand. In such a context, the problem of integrating different carriers into a single multi-hub optimiser represents a challenging topic for the research. This paper depicts the Energy Management System (EMS) of the plant which will be developed and built as a prototype of the INGRID system. The approach followed for the EMS design and development takes the cue from the matrix model presented in the rest of the paper, as well as the general optimisation problem formulation and the algorithm selected for its solution.**

Keywords-hydrogen solid-state storage; multi-carrier energy hub; matrix modelling; multi-objective optimisation.

I. INTRODUCTION

Last two decades have been characterized by the restructuring of vertically integrated energy systems towards liberalized markets, and by the increase of small distributed generation technologies and non-traditional energy carriers. At the same time, information and communication technology (ICT) equipment, like SCADA devices, data acquisition centres and control systems, have been the focus of development in order to provide better services for energy systems and devices. Furthermore, the growing awareness of the impacts of pollution has led to increased interest in sustainability, and environmentalism within governmental organizations, following the concept of sustainable development in order to increase "green" energy sources and reduce the emission of CO_2 in the atmosphere.

These changes in the energy systems, especially in electric power grids, are forcing the system operators, i.e., Distribution Service Operators (DSOs) and Transmission Service Operators (TSOs), to manage and operate their networks in new ways. Thus, it seems reasonable for future development on innovation within the existing infrastructure, to reuse it as much and as long as possible, by means of studying and designing new energy technologies and facilities. The INGRID FP7 European co-funded project [1]-[6] is studying several solutions concerning hydrogen storage, aiming to suitably balance electrical supply and demand. The proposed integrated approach for the INGRID system involves combining a solid-state high-density hydrogen storage system with advanced solutions for smart distribution grids, electrical power grids and methane pipeline networks. A concrete demonstrator will be implemented in Troia (Puglia, Italy), which will operate as an implementation and proof of concept of the overall INGRID system.

The possibility of planning and operating this kind of hybrid systems bridges both technical and economic demands, such as DSO requests, ancillary services, and, above all, price signals from energy markets.

In Section II, the structure of the INGRID system and its components are defined; in Section III, the concept of multi-carrier energy hub is introduced and adapted to the INGRID system structure; Section IV, depicts the proposed matrix model for the INGRID system and the related equations; Section V, is focused on the control algorithm chosen for the EMS implementation; in Section VI, the conclusions are given.

II. THE INGRID PROJECT SYSTEM

A. System structure and working conditions

The structure of the system proposed inside the INGRID Project is composed by a Renewable Energy Source (RES) power generation plant (PV and/or windmill), an Electric

 ISBN: 978-1-61208-406-0

Vehicle Recharge Station (EVRS), a Water Electrolyser (WE) with its AC/DC conversion device, a Fuel Cell (FC) with its DC/AC conversion device and an innovative Hydrogen Solid-state Storage System (HSS). This grid is connected to the Medium Voltage grid (MV, 20 kV) and the Low Voltage grid (LV, 0.4 kV). Figure 1 shows the block diagram of the system [5].

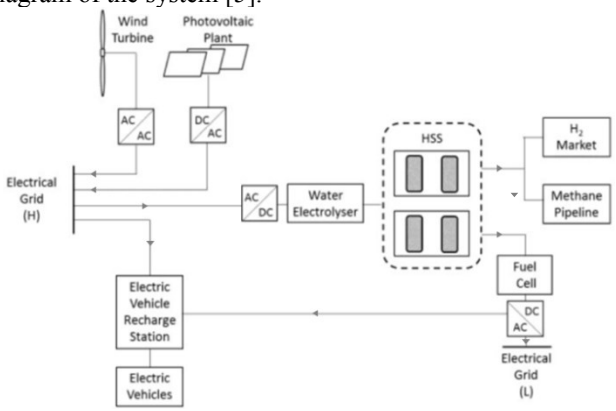


Figure 1. Block diagram of the INGRID system.

One of the most particular features of this system is the brand new hydrogen storage technology based on magnesium hydride, which offers several practical advantages: it adsorbs and desorbs hydrogen at a constant pressure, the high-temperature metal hydride and the low temperature metal hydride have similar dissociation pressures (a feature that allows the process to be self-regulating), and the hydrogen solid-state storage has the capability to store more hydrogen per mass unit than gas and liquid storage forms [17][18].

The WE can absorb electric power both from the MV grid and RES plant. The hydrogen produced by the WE may be sent either to the hydrogen market or to the methane distribution network (as considered in [19]), forming a so called "Open Loop" (OL). Otherwise, the hydrogen may be used to re-obtain electricity by means of the FC, in order to create a "Closed Loop" (CL). The EVRS can be supplied either from the internal RES plant, or directly from the MV grid or the FC generator; the EVRS is managed by an Intelligent Dispenser (ID). These power flows are outlined in Figure 1. The possibility of providing ancillary services directly to the LV grid, by means of the internal RES and/or the FC is being experimented.

A key aspect of the INGRID system is the close collaboration with the DSO for the general balance of the grid. For example, the DSO may request a power injection in the MV or LV feeder to facilitate their system operations. This could include delivering ancillary services, as well as balancing the consumption and generation of power, especially when the risk of power reverse flow occurs to the TSO network (very common in regions characterized by huge levels of RES production). In all these cases, an INGRID based plant would respond according to the availability of power and storage capacity, to fulfil the incoming requests, while keeping in mind its own strategic objectives. This goes beyond the already existing storage systems mainly composed of devices directly controlled by the network operators, and promotes a new form of collaboration and negotiation between the plant and the DSO, as well as the energy market.

B. *INGRID demonstrator*

The INGRID Project comprises the design and development of a demonstrator prototype, which is planned to be deployed in Troia (Puglia, Italy). The WE maximum rated power is equal to 1152 kW, the FC nominal power is 90 kW, the hydrogen storage capacity of the CL tank is equal to 25 kg of hydrogen, while the OL tanks has a total capability of 900 kg of hydrogen (450 kg in use, 450 kg as reserve). The CL storage system has better performances than the OL, because of its technological features, e.g., specifically the Phase Change Material (PCM) and the heat management unit, that allow higher efficiency if the stored energy is effectively used within few hours. Otherwise, the more capacious yet less efficient OL tanks can be used instead.

The prototype will be connected to a MV primary substation (150 kV/20 kV) and will absorb electricity from the electric grid as a flexible load from 0.365 to 1.2 MW. As a variable generator (0-90 kW), the plant will be connected to a low voltage power line. The main difference between traditional auxiliary storage systems, entirely controllable by the DSO, and the proposed one is that the INGRID system is intended to collaborate with DSO, rather than being directly commanded, to optimize its power consumption and generation according to several possible custom strategies, (e.g., to maximise the economical incoming of the plant, to minimise the carbon emissions, to follow the DSO profile, etc., or a combination of several strategies). For this reason, the multi objective optimisation can represent a promising approach to address the problem.

III. MULTI-CARRIER ENERGY HUB

As reported in the introduction, the development of modern energy systems is focused on defining hybrid structures involving various energy carriers. Therefore, this paper presents a hybrid and integrated multi-carrier energy system. The multi-carrier energy hub is defined as a "hybrid energy system which interfaces between energy producers, consumers and the transportation infrastructure" [7]. In Figure 2, an example of a possible multi-carrier energy hub is shown.

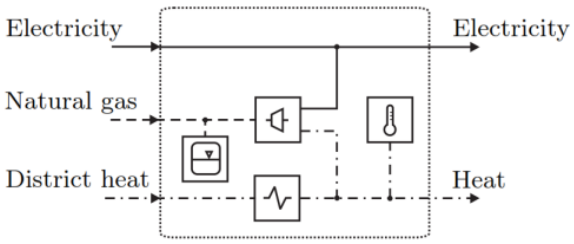


Figure 2. Example hub with gas turbine, heat exchanger, gas tank and heat storage [7].

 ISBN: 978-1-61208-406-0

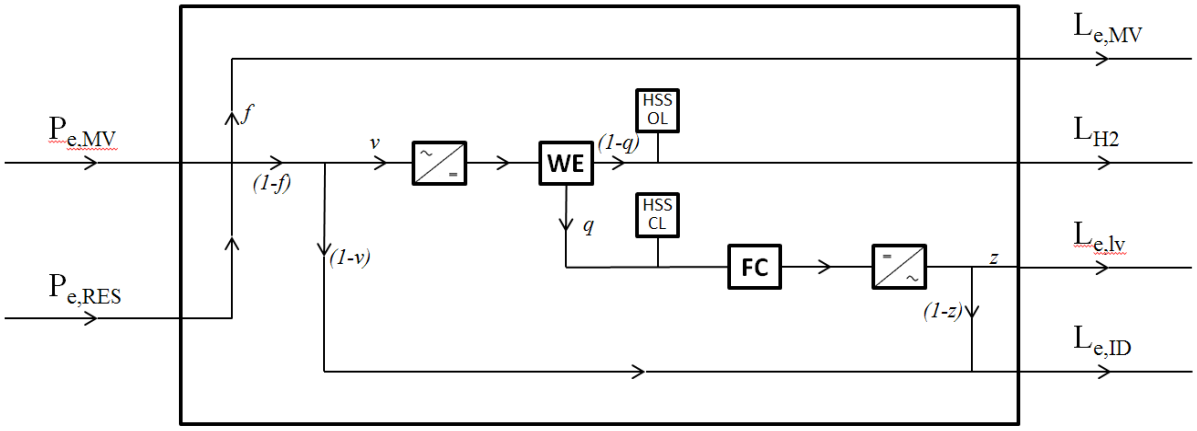


Figure 3. Flow chart of the proposed INGRD system.

$$\begin{bmatrix} L_{H_2} \\ L_{e,LV} \\ L_{e,ID} \\ L_{e,MV} \end{bmatrix} = \begin{bmatrix} (1\text{-}f)v\eta_R\eta_{WE}(1\text{-}q)c_F & (1\text{-}f)v\eta_R\eta_{WE}(1\text{-}q) & e_{OL} & 0 \\ (1\text{-}f)v\eta_R\eta_{WE}\eta_{FC}\eta_I zqc_F & (1\text{-}f)v\eta_R\eta_{WE}\eta_{FC}\eta_I zq & 0 & e_{CL}z \\ (1\text{-}f)\left[v\eta_R\eta_{WE}\eta_{FC}\eta_I(1\text{-}z)q+(1\text{-}v)\right]c_F & (1\text{-}f)\left[v\eta_R\eta_{WE}\eta_{FC}\eta_I(1\text{-}z)q+(1\text{-}v)\right] & 0 & e_{CL}(1\text{-}z) \\ fc_F & f & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} P_{e,MV} \\ P_{e,RES} \\ \dot{E}_{OL} \\ \dot{E}_{CL} \end{bmatrix} . \quad (1)$$

G. Andersson et al. [7]-[11] have proposed a matrix modelling approach for the entire hybrid system, even with storage devices inside. Through this model, the conversion and storage stages of multiple energy carriers may be represented generally and comprehensively. Furthermore, the matrix-based approach is appropriate to use in the formulation of optimisation problems and allows easy implementation of complex optimisation algorithms. Accordingly, a model of the INGRID hybrid system using the approach briefly explained above (whose complete description is available in the referenced works) is proposed in the next section.

IV. MATRIX-BASED MODEL FOR INGRID SYSTEM

The system proposed by the INGRID Project can be modelled through the matrix-based approach introduced in the previous section: the final aggregate instantiation is a multi-carrier energy system that involves both electricity and hydrogen; this model also permits dealing with storage devices, and so it is perfectly suitable for the system purposes.

A graphical representation is depicted, in terms of a flow chart reporting all the devices working in the INGRID system, in Figure 3.

In this flow chart, the different "P_i" denote the input carriers of the system, which are the electric powers coming from the MV grid and RES internal to the plant; the different "L_j", in their turn, denote the output carriers considered as loads. These are: (i) the possible electric power injection in the MV feeder; (ii) the hydrogen flow toward the specific market, (iii) the electric power sent to the LV grid and (iv) the electric power used for supplying the EVRS through the ID. The WE and the FC are considered, respectively, with their AC/DC and DC/AC conversion devices. Eventually, the HSSs, for both OL and CL, are modelled by their own blocks.

The flow chart presents dispatching factors whenever a flow line splits in two different paths (cf. "*f*", "*v*", "*q*" and "*z*" in Figure 3). These factors allow the definition of the amount of energy sent to one process or another. All of them can assume a value between 0 and 1.

For instance, the factor "*q*" fixes how many kilograms of hydrogen per second flow in the OL or in the CL tanks.

The relationships between different energy carriers can be directly deduced by the flow chart and can be expressed through a matrix equation (1).

In this matrix model, each device is represented by its instantaneous efficiency "η_i"; the storage system contribution is represented by the parameters "e_{OL}" and "e_{CL}", that can be seen as the forward or reverse energy efficiencies of the whole storage system, depending on their working state:

$$e_{OL} = \begin{cases} -e_{OL}^{+} & \text{charging} \\ 1/e_{OL}^{-} & \text{discharging} \end{cases} \quad (2)$$

$$e_{CL} = \begin{cases} -e_{CL}^{+} & \text{charging} \\ 1/e_{CL}^{-} \cdot \eta_{FC}\eta_I & \text{discharging} \end{cases} . \quad (3)$$

These parameters will be multiplied by $\dot{E}_\alpha$ (where α = OL, CL) that corresponds to the internal flow of the device, assumed to be the time derivative of the stored energy.

The "c_F" coefficient allows the model to provide the possible plant configurations, depending on the DSO requests: (i) for the plant as a power consumer, and (ii) for the plant as a power producer when a possible power injection into the MV feeder is requested by the DSO. When c_F=1 and f=0, the entire system is supplied both by the grid and the RES plant; if c_F=0 and f≠0, the power produced by the internal RES plant is enough to supply the INGRID system and also to send power to the MV grid (e.g., if f=0.4, the 40% of the RES power is sent to MV grid and 60% is sent to the storage systems). Finally, the system equations for both the described conditions can be derived;

- if c_F=1 and f=0:

$$L_{H_2}=\left(P_{e,MV}+P_{e,RES}\right)v\eta_R\eta_{WE}\left(1-q\right)+\dot{E}_{OL}e_{OL} \quad (4.1)$$

$$L_{e,LV}=\left(P_{e,MV}+P_{e,RES}\right)v\eta_R\eta_{WE}\eta_{FC}\eta_I zq+\dot{E}_{CL}e_{CL}z \quad (4.2)$$

$$L_{e,ID}=\left(P_{e,MV}+P_{e,RES}\right)\left[v\eta_R\eta_{WE}\eta_{FC}\eta_I(1-z)q+(1-v)\right]+$$
$$+\dot{E}_{CL}e_{CL}\left(1-z\right) \quad (4.3)$$

$$L_{e,MT}=0; \quad (4.4)$$

- if c_F=0 and f≠0:

$$L_{H_2}=P_{e,RES}\left(1-f\right)v\eta_R\eta_{WE}\left(1-q\right)+\dot{E}_{OL}e_{OL} \quad (5.1)$$

$$L_{e,LV}=P_{e,RES}\left(1-f\right)v\eta_R\eta_{WE}\eta_{FC}\eta_I zq+\dot{E}_{CL}e_{CL}z \quad (5.2)$$

$$L_{e,ID}=P_{e,RES}\left(1-f\right)\left[v\eta_R\eta_{WE}\eta_{FC}\eta_I\left(1-z\right)q+\left(1-v\right)\right]+$$
$$+\dot{E}_{CL}e_{CL}\left(1-z\right) \quad (5.3)$$

$$L_{e,MV}=P_{e,RES}f. \quad (5.4)$$

V. Multi-Objective Optimisation Algorithm For EMS Structure

By means of the proposed matrix-based model, the operating conditions of the INGRID system can be simulated and, thus, all the energy flows of each carrier of the system can be calculated. The INGRID EMS has to indicate the most suitable configuration of all the devices, sending an optimal working set-point to the local control units of each device.

The logic control criteria are defined by the EMS in order to meet one or more objectives, which are expressed by one or more functions depending on the operating conditions of the system. The most common objective functions [15], which the optimisation problems are referred to, are:

- the total generation cost of the whole supplied power;
- the energy losses in the grid;
- the total polluting emissions of the system.

This kind of problem can be solved with multi-objective optimisation, which optimises several objective functions simultaneously. In such a kind of problem, there is not a unique optimal solution, but a set of suitable solutions fitting the user's goals; these are approximated solutions that can be very near to the ideal solution of an optimisation problem.

Assuming that $f_i(\mathbf{X})$ are *k* different *objective functions*, the multi-objective problem is defined [16] as minimizing the vector of functions:

$$\mathbf{F(X)}=[f_1(\mathbf{X}), \ldots, f_k(\mathbf{X})]^T \quad (6)$$

which is subject to a set of *inequality constraints*, defined as:

$$g_i(\mathbf{X})\leq 0 \qquad i=1, \ldots, m \quad (7)$$

and a set of *equality constraints*, defined as:

$$h_j(\mathbf{X})=0 \qquad j=1, \ldots, p. \quad (8)$$

The *decision variables* are the numerical values which are expected to be chosen, as solutions, by the multi-objective optimisation. They are the independent variables of the objective functions and are used to minimise them; they are represented as the vector:

$$\mathbf{X}=[x_1, \ldots, x_n]^T \qquad \mathbf{X}\in\boldsymbol{\Omega}. \quad (9)$$

where $\boldsymbol{\Omega}$ is a n-dimensional vector space and represent the set of the feasible region of $\boldsymbol{X}$.

The INGRID system EMS relies upon the optimisation of a costs-benefit objective function [5] which refers to the purchase of electrical energy from the MV grid and to the sale of the produced hydrogen, the provision of ancillary services to the LV grid and the supply of the ID. A second fundamental objective function refers to the collaboration with the DSO: one of the goals of the system is following the power consumption profile requested by the DSO (which, in turn, is calculated by the DSO for overall grid system balancing) in order to achieve an economic discount on the energy acquisition price. Indirectly, this goal connects the contribution of the plant to the stability of the whole smart grid. A third possible objective function can involve the reduction of Greenhouse Gas (GHG) emissions, which is still under investigation.

Some of the problem's inequality constraints are fixed by the electrical power and hydrogen minimum/maximum rated values. Other constraints are related to the state of the storage devices, in order to avoid the production of hydrogen when the HSSs are not able to store it. The only equality constraint on the system is defined by the total produced hydrogen volume during the optimisation time horizon.

Such multi-objective optimisation can be resolved through heuristic algorithms [12]-[14], because of the non-linear nature of the objective functions. Previous works of the INGRID Project [5][6] have tested some heuristic

 ISBN: 978-1-61208-406-0

algorithms, such as Generalized Reduced Gradient (GRG), Simulated Annealing (SA) and Tabu Search (TS) towards a simplified formulation of the optimisation problem, which have been used to deal with a mono-objective function related to the economic benefits for the system owner.

The appointed algorithm in order to manage the control issue of EMS is the Non-dominated Sorting Genetic Algorithm-II (NSGA-II) [13][14]. The algorithm is based on Pareto-dominance and Pareto-optimality, which are fundamental concepts in multi-objective optimisation. A solution $\mathbf{X} \in \Omega$ is said to be *Pareto-optimal* with respect to Ω if and only if there is no $\mathbf{X'} \in \Omega$ for which

$$\mathbf{v}=\mathbf{F}(\mathbf{X'})=[f_1(\mathbf{X'}), \ldots, f_k(\mathbf{X'})]^T \quad (10)$$

dominates

$$\mathbf{u}=\mathbf{F}(\mathbf{X'})=[f_1(\mathbf{X'}), \ldots, f_k(\mathbf{X'})]^T. \quad (11)$$

A vector

$$\mathbf{u}=[f_1(\mathbf{X'}), \ldots, f_k(\mathbf{X'})]^T \quad (12)$$

is said to *dominate* another vector

$$\mathbf{v}=[f_1(\mathbf{X'}), \ldots, f_k(\mathbf{X'})]^T \quad (13)$$

(denoted by $\mathbf{u} \preceq \mathbf{v}$) if and only if $\mathbf{u}$ is partially less than $\mathbf{v}$, i.e.,

$$\forall i \in \{1,\ldots, k\}, u_i \leq v_i \wedge \exists i \in \{1,\ldots, k\}: u_i < v_i. \quad (14)$$

Therefore, the Pareto-optimal Set, $\mathcal{P}^*$, is defined as:

$$\mathcal{P}^* := \{\mathbf{X} \in \Omega \mid \neg \exists\, \mathbf{X'} \in \Omega\ \mathbf{F}(\mathbf{X'}) \preceq \mathbf{F}(\mathbf{X})\}. \quad (15)$$

Some recent papers [19]-[26] have proved that the NSGA-II is quite efficient, especially in the field of power distribution operation and planning.

This genetic algorithm separates and orders the population (the set of whole individuals) following a criterion based on non-dominance: individuals which are not dominated by any other individual are labelled with a higher rank; individuals that do not dominate any other individual have a lower rank. By means of a Binary Tournament Selection, individuals with the best rank are chosen to be the parents of the next generation. Following the analogy with the biological phenomena, crossover and mutation actions are performed on the selected elements, through a Simulated Binary Crossover and Polinomial Mutation [14]. Now, the parent and the child population are mixed and only the best N individuals are chosen for the next generation, where N is the population size. This is one of the most important features of the NSGA-II, called "elitism", which avoids possible losses of valuable solutions.

The algorithm results are an entire population of Pareto-optimal solutions. An important component of the EMS in INGRID will be its Decision Support System (DSS). This subsystem allows the set-point configuration of all devices of INGRID system to be chosen. Such a choice can be done either manually by a human operator who supervises the plant, or automatically by a decision maker algorithm which autonomously sets the operating configuration, choosing it from the feasible solutions provided by optimisation algorithm. The DSS will be designed once the demonstrator infrastructure is physically realised.

VI. CONCLUSIONS

In this paper, the problem of managing a multi-carrier energy hub has been discussed. The problem arises when other products, such as hydrogen, are used to store and later re-use the electrical energy flowing from the network, virtually decoupling the electrical side from the other. The concept analysed in depth, which will be implemented on a test site in the INGRID FP7 project, is a very promising approach for extending the concept of a multi-carrier energy hub to include other industrial intermediate products in process industry. This paper, in particular, has introduced the possibility to use multi-objective optimisation approaches to solve the problem in a more complete way, finding solutions that provide the fulfilment of other objectives than just the costs minimization. In the future, a comprehensive study will be performed on the economical sustainability of the entire system, exploiting the results of the implementation of algorithms over the multi-objective function optimisation.

ACKNOWLEDGMENT

This work is part of the INGRID project, co-funded by the European Commission within the FP7 Framework Programme. Authors thank the members of the INGRID Consortium as well as the European Commission for supporting any project dissemination activities. This work reflects only the authors' views. The Commission is not liable for any use that may be made of the information contained therein.

REFERENCES

[1] INGRID (High-Capacity Hydrogen-Based Green-Energy Storage Solutions for Grid Balancing) EU FP7 Project: http://www.ingridproject.eu/ [retrieved: January, 2015].

[2] D. Arnone et al., "An ICT-based Energy Management System to Integrate Renewable Energy and Storage for Grid Balancing," ACM e-Energy '13, May 2013, pp. 259-260, doi:10.1145/2487166.2487196.

[3] A. Rossi, D. Arnone, M. Bertoncini, and D. Moneta, "Sistema di gestione energetico innovativo con stoccaggio a idrogeno per l'integrazione di fonti rinnovabili a supporto del bilanciamento della rete," AEIT 2013, October 2013.

[4] D. Arnone et al., "Energy Management in a Smart Grid-integrated Hydrogen-based Storage - Electric Grid balancing when integrating large-capacity Renewable Energy sources," IARIA ENERGY 2014,April 2014, pp. 86-89.

[5] R. Proietto et al., "Mixed Heuristic-Non Linear Optimisation of Energy Management for Hydrogen Storage-based Multi Carrier Hub," IEEE International Energy Conference (ENERGYCON), May 2014, pp. 1019-1026, doi:10.1109/ENERGYCON.2014.6850550.

[6] R. Proietto et al., "A Novel Heuristics-Based Energy Management System For A Multi-Carrier Hub Enriched With

 ISBN: 978-1-61208-406-0

Solid Hydrogen Storage," ACM e-Energy '14, June 2014, pp. 231-232, doi:10.1145/2602044.2602081.

[7] M. Geidl, "Integrated Modeling and Optimisation of Multi-Carrier Energy System," Swiss Federal Institute of Technology (ETH), Power Systems Laboratory, Zurich, Switzerland, 2007.

[8] G. Andersson and M. Geidl, "A modelling and optimisation approach for multiple energy carrier power flow," Power Tech, 2005 IEEE Russia, June 2005, pp. 1-7, doi:10.1109/PTC.2005.4524640.

[9] M. Geidl, G. Koeppel, P. Favre-Perrod, B. Klöckl, G. Andersson, and K. Fröhlich, "The energy hub - A powerful concept for future energy systems," Third Annual Carnegie Mellon Conference on the Electricity Industry, March 2007, pp. 1-10.

[10] M. Arnold and G. Andersson, "Decomposed electricity and natural gas optimal power flow," Power Systems Computation Conference (PSCC), July 2008, pp. 1-7.

[11] M. Arnold, R.R. Negenborn, G. Andersson, and B. De Schutter, "Multi-area predictive control for combined electricity and natural gas systems," Proceedings of the European Control Conference 2009, August 2009, pp. 1408-1413.

[12] N. Kokash, "An introduction to heuristic algorithms," University of Trento, Department of Informatics and Telecommunications, Italy, 2005.

[13] K. Deb, S. Agrawal, A Pratap, and T.A. Meyarivan, "Fast elitist non-dominated sorting genetic algorithm for multi-objective optimisation: NSGA-II," Lecture Notes in Computer Science, vol. 1917, pp. 849-858, September 2000.

[14] A. Seshadri, "A Fast Elitist Multiobjective Genetic Algorithm: NSGA-II," 2008, http://www.mathworks.com [retrieved: January, 2015].

[15] M.G. Ippolito, M.L. Di Silvestre, E. Riva Sanseverino, G. Zizzo, and G. Graditi, "Multi-objective optimized management of electrical energy storage systems in an islanded network with renewable energy sources under different design scenarios," Energy, vol. 64, no. 1, January 2014, pp. 648-662, doi:10.1016/j.energy.2013.11.065.

[16] C.A. Coello, D.A.V. Veldhuizen, and G.B. Lamont, "Evolutionary algorithms for solving multi-objective problems," New York: Kluwer Academic/Plenum Publishers, 2002.

[17] D. N. Harries, M. Paskevicius, D. A. Sheppard, T. E. C. Price, and C. E. Buckley, "Concentrating solar thermal heat storage using metal hydrides," Proceedings of the IEEE, vol. 100, February 2012, pp. 539-549, doi:0.1109/JPROC.2011.2158509.

[18] M. Fruchart and D. Jehan, "McPhy-Energy's Proposal for Solid State Hydrogen Storage Materials and Systems," Journal of Alloys and Compounds, vol. 580, 15 December 2013, April 2013, pp. 343–348, doi:doi:10.1016/j.jallcom.2013.03.266.

[19] D. Haeseldonckx and W. D'haeseleer, "The use of the natural-gas pipeline infrastructure for hydrogen transport in a changing market structure," International Journal of Hydrogen Energy, vol. 32, no. 10-11, August 2007, pp. 1381-1386, doi:6:5738e59.

[20] B.F. Hobbs and F.A.M. Rijkers, "Modeling strategic generator behavior with conjectured transmission price responses in a mixed transmission pricing system: formulation," IEEE Transaction Power Systems, vol. 19, June 2004, pp. 707-717, doi:10.1109/PES.2004.1373030.

[21] R.P.J. Benvindo, A.M. Cossi, and J.R.S. Mantovani "Multiobjective short-term planning of electric power distribution systems using NSGA-II," Journal of Control, Automation and Electrical Systems, vol. 24, issue 3, June 2013, pp 286-299, doi: 10.1007/s40313-013-0022-5.

[22] B. Tomoiaga, M. Chindris, A. Sumper, A. Sudria-Andreu, and R. Villafafila-Robles, "Pareto optimal reconfiguration of power distribution systems using a Genetic algorithm based on NSGA-II," Energies 6, no. 3, March 2013, pp. 1439-1455, doi:10.3390/en6031439.

[23] H. Xianchao and G.A. Taylor, "Service restoration of distribution systems based on NSGA-II," Universities power Engineering conference (UPEC), September 2010, pp.1-6.

[24] D. Buoro, M. Casisi, A. De Nardi, P. Pinamonti, and P. Reini, "Multicriteria optimisation of a distributed energy supply system for an industrial area," Energy, vol. 53, September 2013, pp. 128-137, doi:10.1016/j.energy.2012.12.003.

[25] S. Özyöna, H. Temurtas, B. Durmus, and G. Kuvat, "Charged system search algorithm for emission constrained economic power dispatch problem," Energy, vol. 46, issue 1, October 2012, pp. 420–430, doi:10.1016/j.energy.2012.08.008.

[26] S. Wu, Z. Yu, X. Feng, G. Liu, C. Deng, and K.H. Chu, "Optimisation of refinery hydrogen distribution systems considering the number of compressors," Energy, vol. 62, December 2013, pp. 185–195, doi:10.1016/j.energy.2013.09.041.

 ISBN: 978-1-61208-406-0

Ener-SCAPE: A Novel Persuasive Game to Improve the Energy Consumption Awareness

Diego Arnone, Alessandro Rossi, Enrico Melodia
Research and Development Laboratory
Engineering I.I. S.p.A.
Palermo, Italy
email: diego.arnone@eng.it, alessandro.rossi@eng.it,
enricogabriele.melodia@eng.it

Marzia Mammina
DemetriX S.r.l.
Palermo, Italy
email: marzia.mammina@demetrix.it

Sandra Elizabeth Jenkins
Engineering Systems Division, Technology and Policy Program '14
Massachusetts Institute of Technology
Cambridge, MA, USA
e-mail: sandjenkins@gmail.com

Abstract— **Ener-SCAPE is an educational game, relying on a novel software framework, whose main aim is trying to improve the awareness of the energy consumption of its users, at home where they live, as well as in the office where they usually work. Players participate by trying to exit from their home or office, by improving some predefined energy efficiency indexes which show the environmental performances of the place. The software application uses a common escape room game framework, tailoring the game archetype to focus on energy efficiency and consumption awareness. Furthermore, Ener-SCAPE implements a novel feedback mechanism based on real energy consumption that induces consumers to apply what they have learned from the virtual reality of the game into their daily real lives.**

Keywords-energy-awareness; ***serious game***; ***educational game; escape game; virtual currency; energy consumption; energy efficiency; social network interaction.***

I. INTRODUCTION

World energy consumption is steadily increasing, despite mitigation efforts, particularly in the countries that are having a great economic growth. This is a global issue for the environmental protection as well as an economic and political issue for countries dependent on foreign energy supplies.

In order to address increasing consumption, many solutions are being proposed and implemented: incentives for renewable energy sources, new technologies for highly efficient buildings, new technologies for more efficient appliances, more sustainable transportation, etc. However, the aggregate energy consumption behaviour of the general population is responsible for a significant percentage of the overall energy consumption worldwide. For this reason, the energy policy adopted by many countries is focused on Energy Consumption Awareness (ECA). The main factors that may influence energy consumption behaviours of citizens are environmental education, real time control of energy consumption, involvement of young people and greater energy consumption awareness. Thus, ECA policies can be strategic for the reduction of energy consumption.

With the growing number of products and devices in homes and offices, learning and keeping up to date with energy efficiency measures can be difficult. However, as more devices are able to communicate with each other, consumers are going to increasingly look for more innovative and accessible ways to manage and learn about energy efficiency. It is essential that all citizens are "energy aware", because improving energy efficiency does not mean that people should give up activities to save energy. Rather, new technologies and more effective behaviour will actually allow citizens to do more, improving their living conditions without compromising their standard of living. Furthermore, while the improvement of energy efficiency means lower costs and greater sustainability, it is also a great opportunity to stimulate economic growth.

This work aims at following and going beyond the series of current initiatives. Through a novel application, our goal is to improve the energy consumption awareness of people both at home and in the office by employing three features: (i) a well-known and successful game model, (ii) a real-time intuitive feedback on consumption and (iii) a strict collaboration with other players/consumers.

In Section II, the state of the art of serious and pervasive games is briefly reported. In Section III, the Ener-SCAPE applications are fully described. In Section IV, the conclusions are given.

II. LITERATURE REVIEW

In the last few years several institutions, including the European Union (EU), have become increasingly committed to promoting a higher awareness of energy consumption. With this aim, the EU has launched several initiatives and published several documents, especially directed at children and young adults. In addition, information campaigns for

 ISBN: 978-1-61208-406-0

greater awareness in energy consumption, not only directly related to public institutions, have spread. Nevertheless, these are actions that will produce long-term results; particularly, those focused on children which can get significant results across generations [1].

In order to obtain immediate results, feedback on consumption is frequently adopted to improve the energy awareness. Research projects, like BeAware [2], or commercial products, like Energy Orb [3], aim at motivating users to be more responsible in their use of electric appliances by showing real-time energy consumption feedback through an intuitive user interface. Recent studies demonstrate that this kind of feedback alone is enough to reduce the consumption by 23% to 27% [4]. Moreover, persuasive technologies have proven useful in modifying behaviours related to energy usage. A relatively unexplored potential way to drive awareness is through the development of the so called "serious games", i.e., interactive virtual simulations whose fundamental purpose is the development of the user's abilities and skills in a simulated environment, so they can be then applied in the real world. Ecoville [5] is a web-browser simulation game that develops a city and its population, handling the energy balance, CO_2 emissions, garbage disposal, etc. EnerCities [6] takes this concept further, based on the Theory of Planned Behaviour [7], where players have to balance three variables: population, planet, profit. Power Agent [8] and Power Explorer [9] are two "pervasive games" that put family members in competition to reduce domestic consumption, as measured by sensors, and points out the difference between players' and non-players' behaviours utilizing an avatar to help convey best practices in energy efficiency.

Even though there have been relatively good results, serious games still remain a "niche" field. The main limits to their popularity are their target audience (scholars and families, often very different from the average users and surely far from the player community), the perception that educational games are boring, and the lack of entertainment and achievement typical in traditional games. One of the most downloaded mobile games, Angry Birds [10], is a very simple physics based game. As many other F2P ("free-to-play") games, it is an example of the translation of various objectives (commercial, in this case) into gaming ones (credits, in-app purchases). F2P games strength lies in their simplicity, intuitive interfaces, and their repay value. Moreover, applications like Farmville [11], a farming simulation social network game, demonstrate that people like cooperative games in large online communities, consisting of traditional gamers and non-gamers alike.

Ener-SCAPE is meant to exploit the strengths of serious games, while overcoming their weaknesses, by posing them in a new light to make them more entertaining and, thus, more likely to be successful.

III. ENER-SCAPE

Ener-SCAPE is a persuasive game aiming at improving the awareness of energy consumers both in their homes and in buildings where they usually work. Players are tasked with escaping from their apartment or office by solving rebus puzzles, through the use of specific information and the implementation of strategies. Improving energy consumption behaviour in the game helps the player to earn virtual currency redeemable for an easy escape. By simulating real situations, the information and sensations experienced remain strongly impressed, and, thus, allow the player to sharpen perception, attention and memory by promoting behavioural changes through "learning by doing".

Two versions of the game Ener-SCAPE are being developed: one related to a home environment and the other to a building/office environment.

Both versions include and integrate three different applications: a dashboard for monitoring real consumption, which shows performance indicators and hints to improve performance, an escape game (Point & Click), and a tool for integration with social networks (Facebook/Twitter) to allow collaboration between players/consumers. The innovative aspect of the proposed approach consists of a simple but effective mechanism to integrate the three applications with the aim of converting existing successful models to an "edutainment" (education and entertainment) function.

The cornerstone of the Ener-SCAPE game, as previously introduced, is essentially to escape from a house (or a building in the second version of the game) as in any escape-game. Escaping can only be achieved after the resolution of a sequence of rebus puzzles (find items, combine them, use them, look for clues, piece together clues, gather information, find combinations, compose puzzle, solve riddles, etc.). The main feature of the game lies in the themes of each action: eco-sustainability, savings, efficiency, and energy-awareness. The most innovative aspect of the proposed application is the feedback from the real world. The game is integrated with a very simple and intuitive interface (monitoring dashboard), which processes the data collected through a sensing infrastructure and allows the consumers to see their real energy performance. The application allows the users to compare their current energy consumption with the historical average values.

While playing, users acquire information (gained from the game or the social networks where they get useful tips for solving puzzles) which can also be used in the real world. The ultimate aim of the game is impacting the consumers' behaviour in their real life. The monitoring dashboard provides a means to measure performance improvement, which is then translated into virtual currency which can be spent in the game to ease an escape. In order to facilitate the solution of a rebus puzzle, the users utilize the virtual currency that they can obtain from both their real performance improvements, and their social network interactions. The user can cooperate with other consumers in specific social network groups : each consumer builds his or her own reputation by cooperating with the community in solving the escape room rebuses as well as in achieving better energy performance in the real world. Reputation is also translated into virtual currency.

In this way users at home will strive to improve their behaviour by trying to reach the optimal values of real performance indexes. Therefore, as can easily be seen from Figure 1, Ener-SCAPE implements a beneficial cycle that

 ISBN: 978-1-61208-406-0

allows the consumer to learn about how to efficiently use energy simply playing the game [12]. At home, the user can save or waste energy. Then, the dashboard graphically represents the real performances coming from a sensing infrastructure installed in the home. The transition from the physical world to the virtual one corresponds to the translation of the real results into virtual currency that affects the consumer's ability to solve the rebuses in the virtual escape room. Finally, the beneficial cycle results in simulated solutions, information the user collects, applied hints, cooperation with other players, monitoring virtual performance, and sharing ideas inside a community. In this way, players may become consumers with increased energy consumption awareness that can help them to detect inefficiencies and improve their energy behaviour.

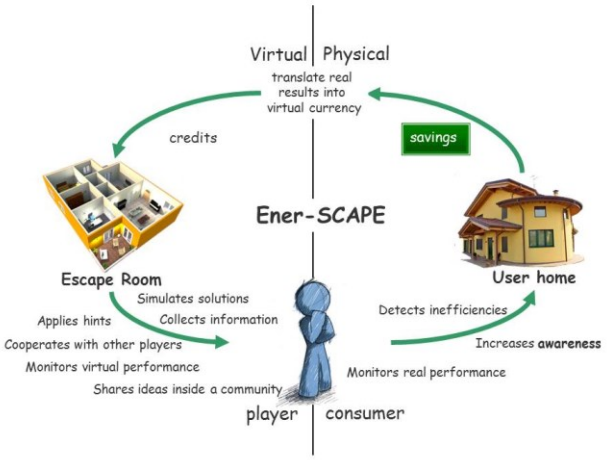


Figure 1. Ener-SCAPE approach to energy consumption awareness (home environment).

The game environment has been designed to be as physically attractive as possible, through a good quality architectural rendering.

Finally, the application was designed to be used from both Personal Computers (PC) and mobile devices, such as smartphones and tablets, in order to meet the requirements of accessibility and to avoid limiting the game to a single kind of device.

Figure 2 shows a mock-up of the game in the tablet version in the home environment. On the right, there is a basket where discovered items are collected; at the bottom, there are some widgets where players can view their virtual and real performances, as well as their scores and hints. Figure 3 shows a mock-up of the dashboard in the tablet version in the home environment. The monitoring interface includes diagrams which give the users a representation of their consumption and performance. In particular, through these diagrams, they can check their current consumption and the ideal average consumption for similar classes of consumers. Improvements in energy performance are translated into money (virtual currency) that can be spent in order to virtually buy needed items in the game.

Figure 2. Mock-up of the game (tablet version) in the home environment.

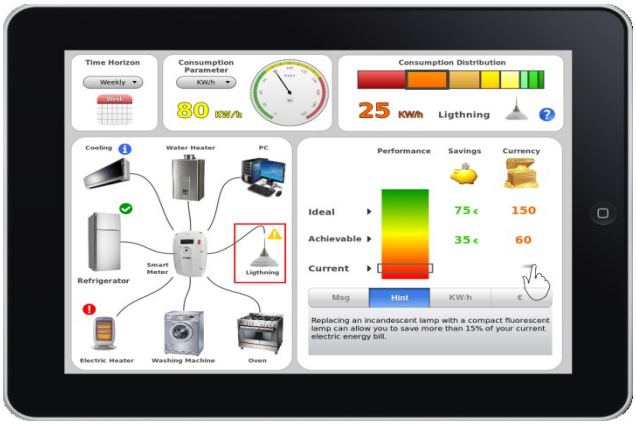


Figure 3. Mock-up of the dashboard (tablet version) in the home environment.

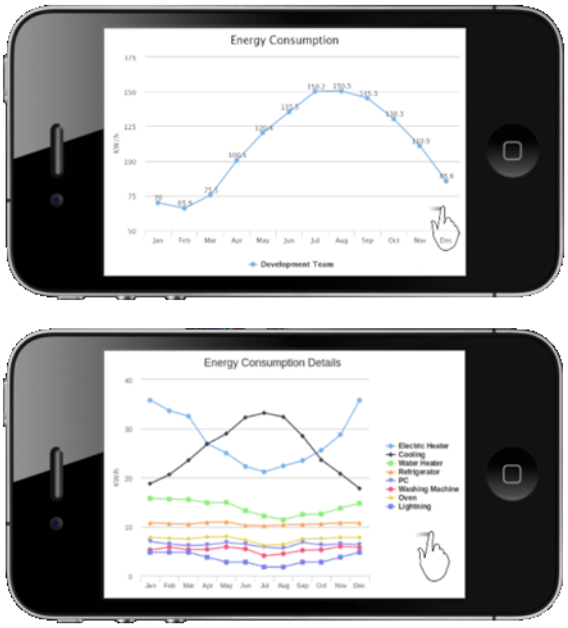


Figure 4. Energy consumption diagrams (smartphone version).

The access to energy consumption and performance values is enabled in multiple ways: by selecting household appliances in the star structure that represents the set of monitored devices (see Figure 3), or by selecting a consumption category (e.g., heating, lighting, appliances, sockets etc.). The user may also access focused hints and check any possible abnormalities in their consumption caused by inefficiencies.

As already mentioned, users can also compare their current consumption with their historical consumption, as showed in Figure 4 which depicts the dashboard for smartphones.

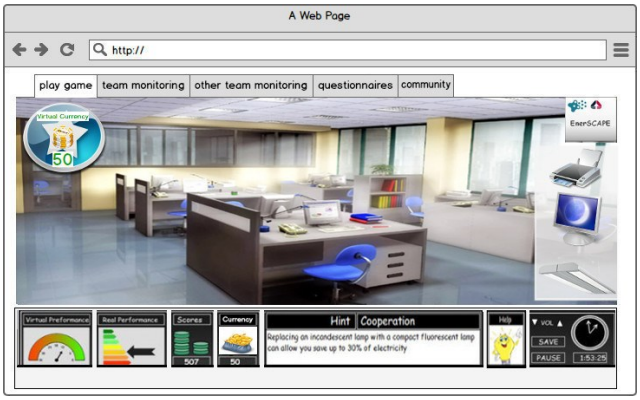


Figure 5. Mock-up of the game (PC version) in the office environment.

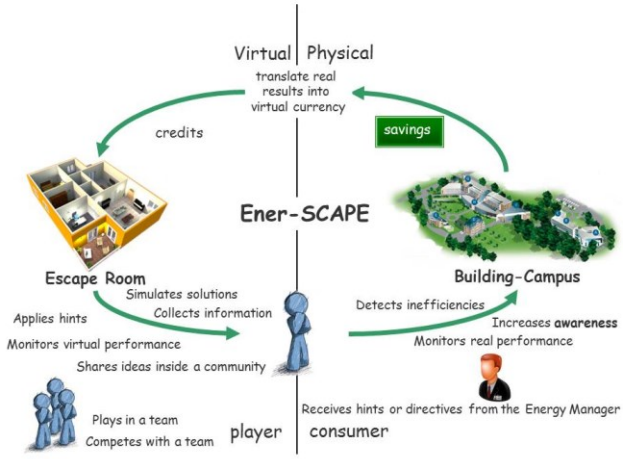


Figure 6. Ener-SCAPE approach to energy consumption awareness (building environment).

The game version for an office environment (see Figure 5 for a mock-up of the PC version) has only one significant difference from the one for the home environment: it is an escape game played in teams. The basic idea of the game is identical to the previous one, but with additional consideration of other actors besides the user/player: the energy manager (responsible for the coordination, management and efficient use of energy resources in the building/field) and other users/players, who, being in the same environment, are energy consumers as well. In this case, a mechanism for collaboration among users belonging to the same team and competition with other groups/teams of users is implemented. Members of the same team can exchange tips, information, and objects in the virtual world where they move and every player can constantly see not only the results of his/her team but also the results of other teams. In order to improve the energy performance the building(s), the energy manager will oversee the progress of the game by having the opportunity to suggest changes in behaviour, precautions, or to impose guidelines. Furthermore, as with the home environment case, the user may take advantage of social networks as an important resource to acquire information and to earn virtual currency.

The beneficial cycle previously described for the office environment is represented in Figure 6.

IV. EXPERIMENTATION PLAN

Ener-SCAPE is under development and no experimentation phase has been started yet. Nevertheless a plan to validate the proposed approach has been defined. Since the application integrates components already known to users neither acceptance problems nor transients for the achievement of a steady operation are expected.

Some metrics have been identified to evaluate the results and have been divided into quantitative metrics and qualitative metrics.

Quantitative metrics measure the effects of greater energy awareness. No preliminary consumption baseline is provided for the evaluation of the performance: users are clustered on the basis of a set of parameters (age, job, education, skills etc.) they provide during a preliminary registration phase and their performance, both in a home and in a building/campus environment, are evaluated on a well-defined set of time slots (hour, day, week, month, etc.) and compared to their own past average values or to the values of users belonging to the same cluster (e.g. the best user, the average one or the worst one). It is assumed that in order to have a greater persistence of the results it is needed to acquire and assimilate slowly a massive amount of information that will have significant impact on the lifestyle of the consumer, rather than only on the performance recorded during the experimental period. The impact of Ener-SCAPE is expected to increase slowly with time, so that the comparison to the average performance, in the first part of the use of the application, can be considered as the comparison to the baseline.

Qualitative metrics are subdivided into indirect metrics, aimed at assessing the user perception of the Ener-SCAPE application, and direct metrics, aimed at evaluating the effectiveness of Ener-SCAPE by evaluating the sensitivity acquired by the user in the field of eco-sustainability. In both cases, we intend to proceed by administering questionnaires to users. The feedback will allow to activate an iterative process in defining new requirements for the final prototype of Ener-SCAPE. The involvement of users will allow reaching a better understanding of their needs for the

 ISBN: 978-1-61208-406-0

achievement of the aims of the application that will allow the creation of a usable product appropriate to the context of use.

V. CONCLUSIONS

This work has proposed a persuasive and universal application to improving energy consumption awareness by designing and developing an energy-aware application based on an escape room game. The application will prove how the adoption of a well-known game model, as well as its integration with real-time consumption feedback and social collaboration, can significantly improve the energy consumption awareness both at home and in office environments.

ACKNOWLEDGMENT

This work is part of “ENERGETIC – Tecnologie per l’Energia e l’Efficienza Energetica” project, co-funded by MIUR (Italian Ministry for Education, University and Research) by means of the national Program PON R&C 2007-2013, project, PON02_00_355_3391233C.

REFERENCES

[1] EUROPEAN COUNCIL - 25/26 MARCH 2010 - Europe 2020: a New European Strategy for Jobs and Growth. [Online]. Available: http://register.consilium.europa.eu/pdf/en/10/st00/st00007.en10.pdf [retrieved: March, 2015].

[2] BeAware, "Home page," [Online]. Available: http://www.energyawareness.eu/beaware/ [retrieved: March, 2015].

[3] Ambient Devices, "Energy Orb," [Online]. Available: http://www.ambientdevices.com [retrieved: March, 2015].

[4] The Brattle Group, "BGE Smart Energy Pricing Pilot," [Online]. Available: http://www.brattle.com/_documents/uploadlibrary/upload768.pdf [retrieved: March, 2015].

[5] Ecoville, [Online]. Available: http://www.ecovillelejeu.com [retrieved: March, 2015].

[6] EnerCities, "Home page," [Online]. Available: http://www.enercities.eu [retrieved: March, 2015].

[7] I. Ajzen, "The theory of planned behavior," in Organizational Behavior and Human Decision Processes, vol. 50, iss. 2. pp. 179–211, October 1991, doi: 10.1016/0749-5978(91)90020.

[8] A. Gustafsson, M. Bång, and C. Katzeff, “Evaluation of a Pervasive Game for Domestic Energy Engagement Among Teenagers,” in Computers in Entertainment (CIE), vol. 7, iss. 4, art. 54, December 2009, doi: 10.1145/1658866.1658873.

[9] A. Gustafsson, M. Bång, and M. Svahn, “Power Explorer – a casual game style for encouraging long term behavior change among teenagers,” in International Conference on Advances in Computer Enterntainment Technology, pp. 182-189, October 2009, doi: 10.1145/1690388.1690419.

[10] “Angry Birds,” Rovio, [Online]. Available: http://www.angrybirds.com [retrieved: March, 2015].

[11] “FarmVille,” Zynga, [Online]. Available: https://it-it.facebook.com/FarmVille [retrieved: March, 2015].

[12] EscaperRoom24, [Online]. Available: http://www.escapegames24.com/ [retrieved: March, 2015].

 ISBN: 978-1-61208-406-0

Cloud Based Optimal Routing and Powertrain Management for Hybrid and Electric Vehicles

Can Kurtulus
Team Leader, SW & Controls, AVL Turkey
PhD Candidate, Istanbul Technical University
Istanbul, Turkey
email: can.kurtulus@avl.com

Gökhan Inalhan
Assoc. Prof., Aeronautical Engineering Department
Director, Controls and Avionics Laboratory
Istanbul Technical University
Istanbul, Turkey
email: inalhan@itu.edu.tr

***Abstract*— Hybrid and electric vehicles have been gaining significant traction in the vehicle market for the past few years, and this necessitates that further attention should be given to their unique characteristics. This paper illustrates a combined route guidance and power management approach, where the required computation is distributed across an on-board/off-board system architecture. A model of the vehicle and the powertrain is utilized to calculate arc costs of a road network for optimal route guidance purposes, and the calculated power reference is later on used for dynamic programming based power management. This approach has the potential to improve vehicle efficiency significantly, and an application scenario based on recorded real world trip information is used to demonstrate system operation.**

Keywords - hybrid-electric vehicles; powertrain management; vehicle route guidance; cloud computing

I. INTRODUCTION

Hybrid and electric vehicles have been gaining a large importance lately as the primary answer to the demands of ever lower levels of noxious emissions as well as lower fossil fuel consumption [1]. These vehicles either have an electric drive component in addition to the conventional combustion engine, or completely get rid of the combustion engine and have fully electric powertrains [2]. The higher efficiency of the electric drive, as well as the possibility to capture normally wasted kinetic energy (i.e., during braking), bring about a significant level of improvement in both fossil fuel consumption and noxious emissions. The main barrier holding back an even more widespread adoption of these types of powertrain configurations is the added system cost, especially for the battery, and the relatively lower energy density of electrochemical energy storage as opposed to fossil fuels, which in turn limits the range of electric driving. The resulting effect of the current situation is that utilization and flow of energy is an even bigger and critical concern for electrified powertrains as compared to conventional, combustion engine based ones.

This work builds upon previous results [3], that demonstrated an algorithm to determine the optimal route with respect to a composite cost function of distance, time, energy and battery wear, for a given source location and target destination pair in a road network. The approach to determination of the optimal route was Bellman–Ford–Moore (BFM) algorithm [4], which builds on the idea of dynamic programming applied to the shortest path problem [5]-[7]. This algorithm has a bound on worst case performance of O(mn) in graph of n nodes and m arcs. Another important feature of the previously reported approach was that external data sources enabled by the benefits of intelligent transportation systems – such as road topology, traffic and weather could be used as additional information to calculate the optimal route.

To be able to obtain the highest efficiency of hybrid electric vehicles, it is a prerequisite to have a combination of route guidance, and management of powertrain operation, which is a currently unexplored area of work. Past work has mainly focused on powertrain management of hybrid vehicles alone [8][9], or on determining the optimal path between two nodes in a road network with respect to time or distance [4] independent of the actual vehicle operation. The aim of this paper is to propose an architecture for this combination, and initial results, hence build upon previous work which focused on route guidance alone [3].

It is clear that for large road networks, which run into millions of nodes and several millions of arcs [10], it is not a straightforward task to compute optimal routes through the network, especially considering the fact that external data is incorporated into the calculation, and has to be processed in real time. For this purpose, it is proposed that the computation and memory intensive operations are shifted off board from the vehicle, to the cloud, and a simple communication strategy ensures that the optimal routing and powertrain operation is relayed to the vehicle, which has limited computing ability on-board.

The main theme of this work revolves around an integrated system architecture where a hybrid and electric vehicle focused route guidance is in constant interaction with hybrid powertrain management in a multi-layered fashion. The route guidance layer hinges on a model of the vehicle and the powertrain to determine energy efficient routes. This model, used to calculate costs for traversing the road network, is also important for the powertrain management

 ISBN: 978-1-61208-406-0

layer since expected future power profile during traversal of the route should be utilized to enable predictive powertrain management.

In Section 2, the problem is introduced, and the algorithm to find the minimum cost route through the road network, and power management is defined. In Section 3, the vehicle and powertrain model to calculate the energy costs for road network links is presented. In Section 4, the combined route guidance and power management algorithm is presented, and finally Section 5 gives an application example using a real world driving profile recorded from a trip, and the associated powertrain management strategy.

II. PROBLEM DEFINITION

A road network is a directed graph G{N,A}, where N is a set of nodes to represent road intersections, and A is a set of arcs to represent road segments connecting two nodes. Each edge of this graph, i.e., the arcs, represents a pair of connected nodes, (i and j), and is associated with a cost c_{ij} and arc(i,j) $\in$ A. The shortest path is a sequence of arcs (a_1, a_2,..., a_n) $\in$ A, that minimizes the total cost $\sum_{k=1}^{n-1} c_{i,i+1}$ over this sequence. The cost, c_{ij} is a weighted sum of distance, time, energy and battery wear where w_x are respective weights such that;

$$c_{ij} = w_d{*}c_d + w_t{*}c_t + w_e{*}c_e + w_b{*}c_b . \quad (1)$$

Previous work [3] demonstrated that equal weighting of all cost functions provide a well-balanced route in terms of time, distance, energy use and battery ageing that can serve as a basis for everyday driving. The weights can also be adjusted based on preferences of each driver.

Power management is the control of power and energy sources in a hybrid powertrain where the power demand of the vehicle is distributed to power providers according to pre-defined criteria. Fuel consumption is a widely used criterion [8], and the associated cost function can be defined as

$$J = \int_0^t \dot{m}_f\,(t, u(t))dt \quad (2)$$

where fuel mass flow $\dot{m}_f$ is minimized over the trip duration, *t*, via determining the optimal control sequence, *u**.

Dynamic Programming is a preferred approach for optimization based power management for hybrid vehicles, and the problem definition is already well covered in the literature [8][11].

III. VEHICLE AND POWERTRAIN MODEL

The vehicle and powertrain architecture selected for modeling is a range extended electric vehicle. This architecture is a part of the "Series Hybrid" family, where the combustion engine does not have a direct connection to the wheels, but drives a generator to support the electrical power network, and charge the battery whenever necessary. The main difference of this architecture to a battery electric vehicle is addition of a generator, which is comprised of a combustion engine, and an e-motor driven by that engine to support the power network. The architecture can be seen in Figure 1.

The powertrain components were modeled in a "Quasi-Static" fashion based on the concept of [12], which is the preferred approach for purposes of energy management and control. This approach neglects high frequency dynamics irrelevant to energy based modeling, and relies on look-up tables of powertrain components (e.g., engine, e-motor etc.) to relate power conversion and efficiency in a steady-state manner. These look-up tables are validated by test data, and dynamics are incorporated to a level they are necessary, such as battery state of charge (SoC) behavior, where the charge stored in the battery is modeled as a function of charge and discharge power of the battery over time, and battery voltage is related to SoC and power [13].

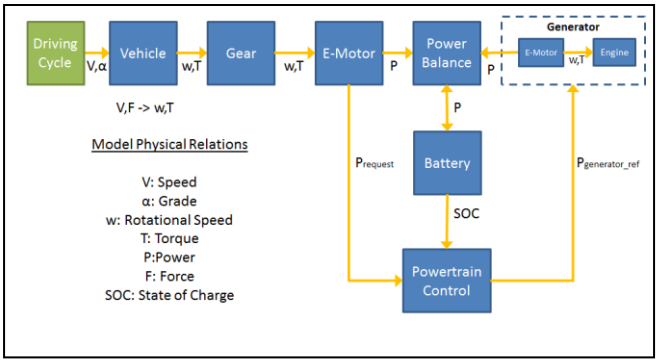


Figure 1. Powertrain and Vehicle Model Structure

The model was parameterized based on publicly available data for the Audi A1 e-tron [14], which is a good example of a range extended electric vehicle for urban use. The parameter set is listed in Table 1.

TABLE I. AUDI A1 E-TRON PARAMETERS

E-Motor		
Type	PMSM	-
Continuous Power	45	kW
Peak Power	75	kW
Range Extender		
Type	Wankel Engine	
Electrical Power	15	kW
Battery		
Capacity	12	kWh
Voltage	270	V
Vehicle		
Mass	1410	kg

 ISBN: 978-1-61208-406-0

IV. ROUTE GUIDANCE AND POWERTRAIN MANAGEMENT ALGORITHM

The overall route guidance and power management algorithm consists of a pre-processing and query steps.

Pre-processing step

- Input: Nodes and arcs of the road network,
- For every arc in the network, periodically carry out the following;
  - Acquire/load any additional data available for the network (e.g., topography, weather, traffic etc.). Data collection period depends on the type of data (i.e., daily for road work, 5 minutes for traffic, 15 minutes for weather)
  - Calculate the time to traverse each arc (might be different due to weather and/or traffic), if there is an update from the previous sample
  - Run the powertrain model to compute energy cost for each link, and post process simulation output for battery ageing calculation, as well as the composite cost.

Query step

- For the source and target nodes, calculate the optimal path with the least composite cost with BFM algorithm
- Run Dynamic Programming for the optimal path utilizing the pre-calculated power profile
- Send the path and range extender on and off command coordinates to the vehicle

The pre-processing step is run periodically for the road network of an area, and only the model based cost calculation of the step is specific to a particular vehicle, whereas the other steps are common for any number of vehicles, therefore benefit all vehicles that make use of the system. For a particular vehicle (or a family of vehicles sharing similar design parameters and powertrains), only the last step of pre-processing needs to be carried out, and it could also be possible to geographically constrain the evaluated arcs for cost calculation, due to the fact that it is usually possible to have knowledge of common and recent locations for a particular vehicle.

The link of route determination algorithm with the power management is through the vehicle and powertrain model and arc cost calculation, where the power demand profile to be negotiated in the calculated path was already found when calculating arc costs. This power profile is later on used to decide when to utilize the generator for range extension as part of a 2nd optimization step that is part of the powertrain management layer. The overall concept is shown in Figure 2.

The interaction between the vehicle and the cloud can be seen in Figure 3, where the vehicle handles user interface functionality to get the source and target locations, and carry out the range extender operation reference command based on location tracking to the range extender operation reference profile. All computations to determine optimal routes and range extender operation strategy are carried out in the cloud. In the case of relatively short trips, communication between the vehicle and the cloud would be necessary only once. Further communication to send route and range extender operation updates from the cloud could be needed for long trips, due to the fact that traffic and weather may change during the trip.

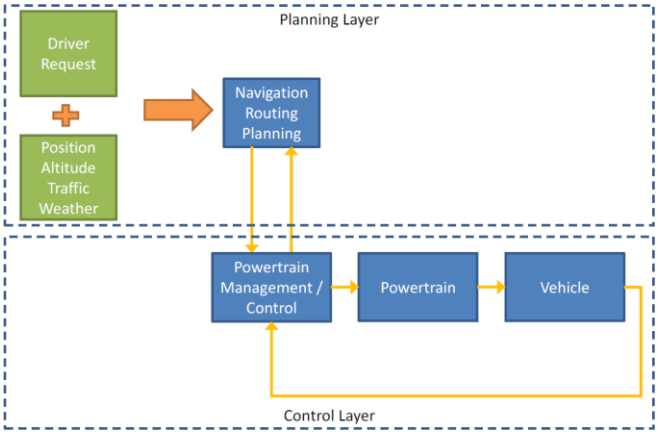


Figure 2. Multi Layered Architecture for Powertrain Management

V. APPLICATION EXAMPLE

The application example is based on a real world recorded trip, where the route has been determined previously, and power management behavior is evaluated to judge feasibility of the approach. The trip route can be seen in Figure 4, and is nearly 64 km long. The route includes a large variation in altitude and speed, which makes it a good candidate to understand behavior of dynamic programming based operation control of the range extender.

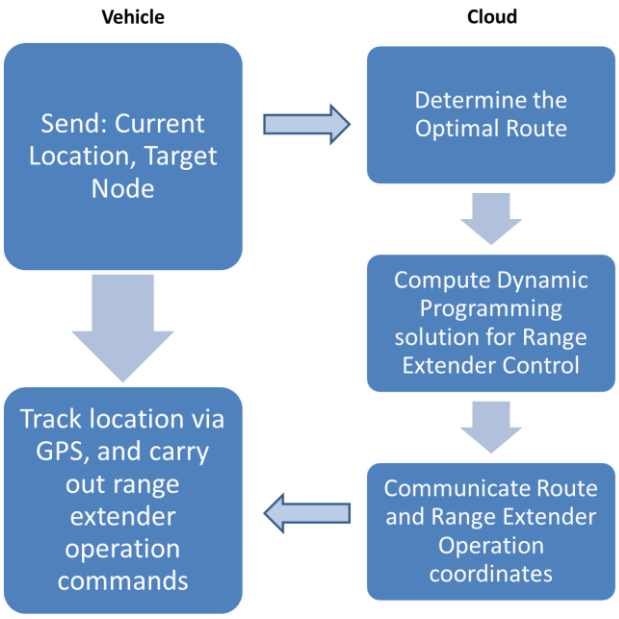


Figure 3. System Architecture

 ISBN: 978-1-61208-406-0

Figures 5 and 6 show the speed profile used during simulations in green (scaled 1/100 to fit in the same graph), battery SOC in blue, and finally the generator operation command in red (scaled by ½ to fit in the same graph). For comparison purposes, the same driving profiles were simulated with Charge-Depleting/Charge-Sustaining (CD-CS) control for generator operation that is used in Plug-in Hybrid Electric Vehicles (PHEV) and range extender electric vehicles (REEV). This strategy is a rule based one, and commonly used in industry as well as serving as the starting point for academic work [13]. The strategy is quite straightforward and aims to initially use most of the electric energy stored in the battery, called the "Charge Depleting" mode, and then switch to conventional hybrid vehicle operation, called the "Charge Sustaining" mode. In this second mode, the generator is operated to keep battery SOC around a pre-determined level.

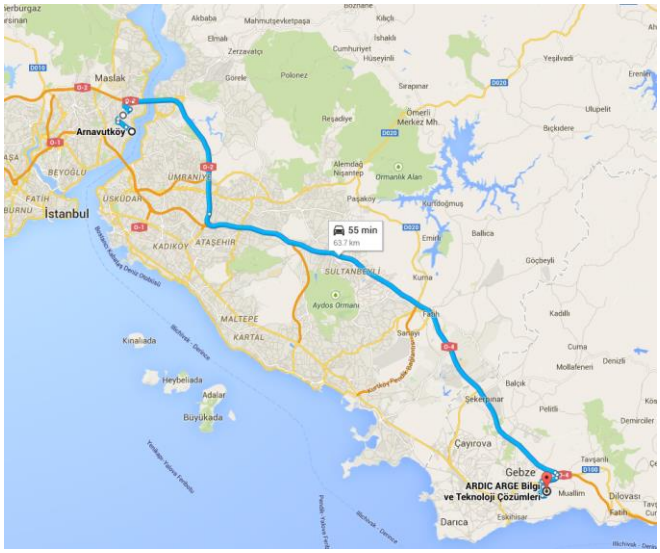


Figure 4. An overview of the route where the real world driving cycle was recorded.

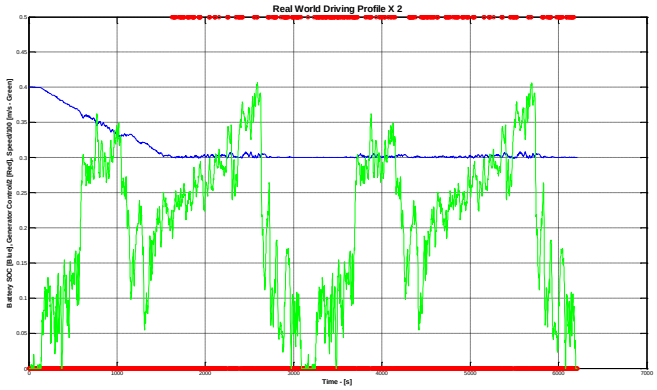


Figure 5. Figure 1 Generator operation and battery SOC behavior during a real world driving profile with CD-CS strategy

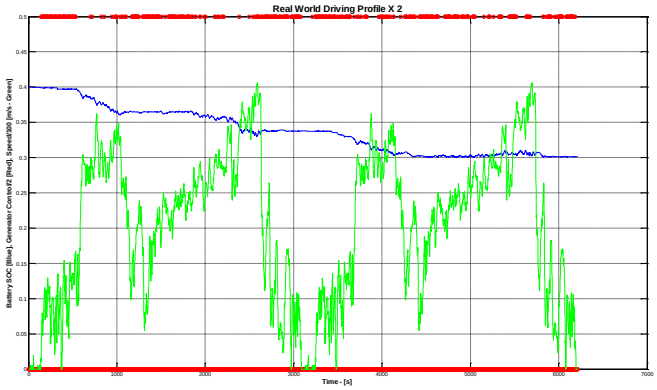


Figure 6. Generator operation and battery SOC behavior during a real world driving profile that minimizes fuel consumption

A quick evaluation of the results demonstrate key differences between a rule based decision on generator operation, and the operation strategy determined via dynamic programming. One of the significant differences is the starting point for generator operation. CD-CS strategy starts to operate the generator whenever the battery SOC drops below the pre-determined threshold (set at 30% for these simulations) as the name indicates, and continues to operate the generator to keep the battery at this threshold. Dynamic Programming based algorithm (DP) has knowledge of the complete driving profile, as determined already by the optimal routing algorithm, therefore starts to operate the generator much earlier during the profile to enable balanced operation of the battery and the generator as the energy source, and the battery reaches the pre-determined threshold significantly later during the driving profile. This contributes to the fact that battery is operating in a more efficient region for the duration of the trip, leading to lower fuel consumption, on the order of 3.7%.

VI. CONCLUSION

We have discussed a novel approach for combined route guidance and power management for hybrid and electric vehicles, carried out in an on-board/off-board coordinated computation architecture. This aspect has not previously been explored in literature, and is meant to serve as a starting point for further research where the interplay between route guidance and powertrain management will be explored in detail.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive feedback, which helped improve presentation of ideas and overall quality of the paper.

REFERENCES

[1] G. Rizzoni, L. Guzzella, and B. M. Baumann, "Unified modeling of hybrid electric vehicle drivetrains," IEEE/ASME Trans. Mechatronics, vol. 4, no. 3, 1999, pp. 246–257.

[2] A. Emadi, K. Rajashekara, S. S. Williamson, and S. M. Lukic, "Topological Overview of Hybrid Electric and Fuel Cell Vehicular

 ISBN: 978-1-61208-406-0

Power System Architectures and Configurations," IEEE Trans. Veh. Technol., vol. 54, no. 3, May 2005, pp. 763–770.

[3] C. Kurtulus and G. Inalhan, "Model Based Optimal Route Guidance for Hybrid and Electric Vehicles," unpublished.

[4] B. V. Cherkassky, A. V. Goldberg, and T. Radzik, "Shortest paths algorithms: Theory and experimental evaluation," Mathematical programming, vol. 73, no. 2, May 1996, pp. 129-174,

[5] R.E. Bellman, "On a routing problem," Quarterly of Applied Mathematics, vol. 16, 1958, pp. 87–90.

[6] L. R. Ford and D. R. Fulkerson, Flows in networks. Princeton: Princeton University Press, 1962

[7] E. F. Moore, "The shortest path through a maze," Proc. Internat. Sympos. Switching Theory 1957, Part II. Cambridge, Mass.: Harvard Univ. Press, 1959, pp. 285–292.

[8] A. Sciarretta and L. Guzzella, "Control of hybrid electric vehicles," IEEE Control Syst. Mag., vol. 27, no. 2, Apr. 2007, pp. 60–70.

[9] S. G. Wirasingha and A. Emadi, "Classification and Review of Control Strategies for Plug-In Hybrid Electric Vehicles," IEEE Trans. Veh. Technol., vol. 60, no. 1, Jan. 2011, pp. 111–122.

[10] [Online]. Available http://snap.stanford.edu/data/roadNet-CA.html [Accessed: 09-Jan-2015].

[11] X. Zhang and C. Mi, "Vehicle Power Management," London: Springer-Verlag, 2011, pp 179-208

[12] L. Guzzella and A. Sciarretta, Vehicle Propulsion Systems: Introduction to Modeling and Optimization. Berlin, Germany: Springer-Verlag, 2007, p. 350.

[13] S. Stockar, V. Marano, M. Canova, G. Rizzoni, and L. Guzzella, 'Energy-Optimal Control of Plug-in Hybrid Electric Vehicles for Real-World Driving Cycles', IEEE Transactions on Vehicular Technology, vol. 60, no. 7, Jan. 2011, pp. 2949–2962.

[14] [Online].Available:http://origin-www.audi.com/etc/medialib/ngw/product/a1/a1/my_2010_catalog_images/pdf.Par.0009.File.pdf/2012_07_ams1512_120.pdf [Accessed: 09-Jan-2015].

 ISBN: 978-1-61208-406-0

A Framework for Power Consumption Analysis of Green Cellular Networks with Separated Control and Data Base Stations

Yun Won Chung
School of Electronic Engineering, Soongsil University, Seoul, Korea
Email: ywchung@ssu.ac.kr

***Abstract*—In conventional green cellular network, network functionalities of Base Stations (BS) for network access and data service are tightly coupled. Although this joint control and data signaling approach is reasonable for homogeneous cellular networks, it is not appropriate for heterogeneous cellular networks with small cells overlaying macro cell, and a separated control and data signaling approach was proposed in the literature, where a macro cell is mainly responsible for control signaling to provide network access and a small cell is mainly responsible for data service. Although a detailed description of user equipment state transitions was provided previously, state transitions of a macro cell BS and a small cell BS have not been covered to the best of our knowledge. In this idea paper, we propose state transition diagrams for both the macro cell BS and the small cell BS, and present a framework for power consumption analysis of green cellular networks with separated control and data BSs.**

***Keywords*-power consumption, green cellular network, state transition diagram, base station**

I. INTRODUCTION

Recently, interest on green cellular networks has been increased and the basic idea of green cellular network is to turn off power consuming components of unnecessary Base Stations (BSs), especially during late night. To this end, numerous works on reducing power consumption of BSs of cellular networks have been carried out actively [1] [2]. In these works, it is generally assumed that network functionalities of BSs for network access and data service are tightly coupled and BSs manage both control signaling for network access and data signaling for data service.

Although the joint control and data signaling approach is reasonable for BSs in homogeneous cellular networks, it is not appropriate for heterogeneous cellular networks with small cells overlaying macro cell, since the joint approach is not efficient to deal with the issues such as high control overhead for growing traffic, lack of flexible topology adaptation accommodating traffic variations, and large control signaling overhead and frequent handovers [3]. To address these issues, a separated control and data signaling approach has been proposed in the literature [3][4]. In the separation approach, a macro cell is mainly responsible for control signaling to provide network access and a small cell is mainly responsible for data service. By separating network access and data service functionalities, a small cell BS can be put into sleep mode or turned off in order to reduce power consumption when there is few active mobile users under the coverage of a small cell BS, while idle mobile users under the coverage of a small cell BS can be served by a macro cell BS.

Xu et al. [3] separated network layer as control network layer and data network layer. The control network layer is responsible for supporting network access for mobile users and providing low rate data services. On the other hand, the data network layer is responsible for providing high rate data services. By doing this, transmission component of small cell BSs can be turned off if there is no requirement for high rate data service from any mobile users under the coverage of small cell BSs, and power saving can be achieved. In order to verify the support of the all the User Equipment (UE) activities in the separation approach, a state transition diagram for UE, which consists of detached, idle, c-active, and d-active states was proposed [3]. In c-active state, UE is served by control network layer and low rate data service is provided by a macro cell BS. In d-active state, UE is served by both control network layer and data network layer, and high rate data service is provided by a small cell BS.

Although a detailed description of UE state transitions was provided previously [3], state transitions of a macro cell BS and a small cell BS have not been covered to the best of our knowledge, which are essential for analyzing the power consumption of a macro cell BS and a small cell BS in the functionally separated cellular networks. In this idea paper, we propose state transition diagrams for both the macro cell BS and the small cell BS, and present a framework for power consumption analysis of green cellular networks with separated control and data BSs.

II. PROPOSED IDEA

In this idea paper, we define states for both the macro cell BS and the small cell BS, and propose state transition diagrams for the macro cell BS and the small cell BS. In each state of the state transition diagram, BS consumes different power and power consumption of a BS can be derived by calculating the weighted sum as follows:

$$P_{BS} = \sum_{i} Prob(i) \times P_i, \quad (1)$$

 ISBN: 978-1-61208-406-0

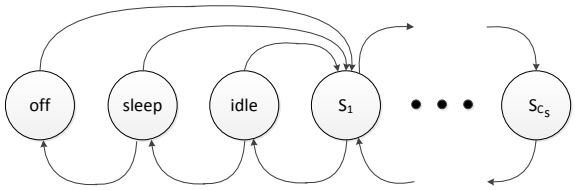


Figure 1. State transition diagram for a small cell BS.

where $Prob(i)$ denotes the steady state probability of state i and P_i denotes the power consumption of state i. Therefore, defining the state of a macro cell BS and a small cell BS and modelling the state transition diagram appropriately are essential to analyze the power consumption of the macro cell BS and the small cell BS.

Figure 1 shows a state transition diagram for a small cell BS which supports high rate data service from UE under the coverage of a small cell only area. In idle state, a small cell BS is ready for providing high rate data services to UEs and it moves to S_1 state after accepting high rate data service from a UE. If another high rate data service is accepted, the state moves to S_2, and similar transition occurs from S_{i-1} to S_i until i reaches to C_S, which is defined as the total number of data channels of a small cell BS. In idle state, if there is no high rate data service request from any UE until the expiration of idle timer, the state moves to sleep state and the small cell BS turns off some of components related with data transmission to save power. In idle state, if there is high rate data service request from UE, it moves to S_1 state. In sleep state, if there is high rate data service request from UE, it moves to S_1 state. Although a small cell BS firstly moves to idle state and stays there temporarily for the preparation of data service, and then moves to S_1 state in this case, we assume a direct transition from sleep state to S_1 state for simple analysis. In sleep state, if there is no high rate data service request until the expiration of sleep timer, the state moves to off state for more power saving. The transition from off state to S_1 state can occur if there is any high rate data service request, with the support of a macro cell BS, and the detailed algorithm of turning on the small cell BS in off state when a high rate data service is requested from UE is not covered in this idea paper since it is beyond the scope of this paper.

Figure 2 shows a state transition diagram for a macro cell BS which supports low rate data service from UE under the coverage of a macro cell and high rate data service from UE under the coverage of a macro cell only area. Total number of C_M channels are assumed in a macro cell BS, where it is assumed that one channel is required for a low rate data service and h channels are required for a high rate data service. Total number of L_M low rate data channels are reserved for low rate data service and total number of H_M high rate data channels are reserved for high rate data

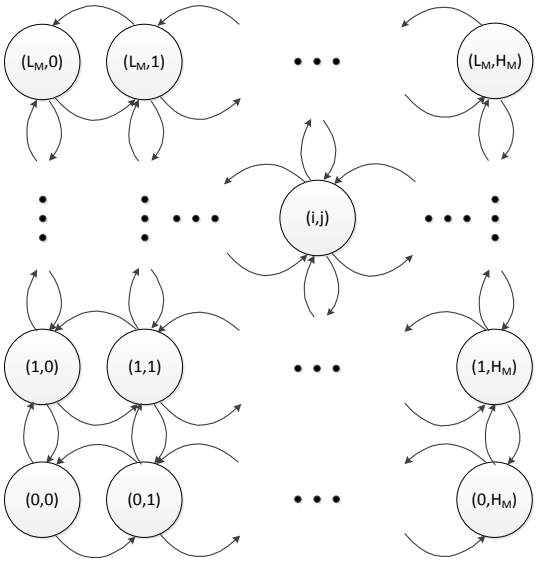


Figure 2. State transition diagram for a macro cell BS.

service, for simplicity. In the state transition diagram for a macro cell BS, there is no power saving state such as off and sleep state since the macro cell should support the network access for all the UEs under the coverage of a macro cell and thus awake all the time. In state(i,j) i denotes the number of low rate data service and j denotes the number of high rate data service; the following transitions occur depending on the events:

- (i,j) $\rightarrow$ (i+1,j): an arrival of low rate data service;
- (i,j) $\rightarrow$ (i-1,j): a departure of low rate data service;
- (i,j) $\rightarrow$ (i,j+1): an arrival of high rate data service;
- (i,j) $\rightarrow$ (i,j-1): a departure of high rate data service.

Since the power consumption of a BS consists of base power consumption (a) and traffic load proportional power consumption ($b\rho$), that is, $a + b\rho$ [1], we can obtain the power consumption of each state by calculating traffic load ρ using the channel occupied. For example, traffic load of state (i, j) for a macro cell BS can be defined as $\frac{i+hj}{C_M}$, where h denotes the number of required channels for a high rate data service. Then, the power consumption of the macro cell BS and the small cell BS is obtained as the weighted sum of power consumption of each state with the weight of the steady state probability of each state, as in eq. (1).

III. Future Plan

In our future work, we will derive the steady state probability of each state based on a detailed probability distributions of transition events between states which depend on mobility and traffic characteristics of UE, cell layout, call admission control scheme, etc. Power consumption of the macro cell BS and the small cell BS in the separation

 ISBN: 978-1-61208-406-0

approach will be derived and compared with that of the joint approach. Then, an efficient power saving scheme will be drawn for a macro cell BS and a small cell BS, based on an interworking between them. Finally, an integrated power saving scheme which considers the power consumption of UEs as well as BSs will be proposed and analyzed.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1037728).

REFERENCES

[1] L. Budzisz, F. Ganji, G. Rizzo, M. A. Marsan, M. Meo, Y. Zhang, G. Koutitas, L. Tassiulas, S. Lambert, B. Lannoo, M. Pickaet, A. Conte, I. Haratcherev, and A. Wolisz, "Dynamic resource provisioning for energy efficiency in wireless access networks: a survey and an outlook," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 2259-2285, 2014.

[2] J. Wu, Y. Zhang, M. Zukerman, and E. K. Yung "Energy-efficient base stations sleep mode techniques in green cellular networks: a survey," *IEEE Communications Surveys & Tutorials*, DOI:10.1109/COMST.2015.2403395 (to appear)

[3] X. Xu, G. He, S. Zhang, Y. Chen, and X. Xu, "On functionality separation for green mobile neteworks: concept study over LTE," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 82-90, 2013.

[4] Y. Kishiyama, A. Benjebbour, T. Nakamura, and H. Ishii, "Future steps of LTE-A: evolution toward integration of local area and wide area systems," *IEEE Wireless Communications*, vol. 20, no. 1, pp. 12-18, 2013.

 ISBN: 978-1-61208-406-0

Power-Aware Cooling Control Architecture for Container Data Center

Hiroyoshi Kodama, Masatoshi Ogawa, Hiroshi Endo, Toshio Sugimoto, Hiroyuki Fukuda, Masao Kondo, and Jun Tanaka
Fujitsu Laboratories Ltd.
Kanagawa, Japan
Email:{hkodama, ogawa.masatoshi, endo-hiroshi, tsugi, fukuda.hiro, condor, tanaka.jun.777}@jp.fujitsu.com

***Abstract*—In this paper, we propose a power-aware cooling control architecture for container data centers. In leading edge data centers, the Computer Room Air Conditioning (CRAC) unit controls the temperature and humidity. The CRAC uses large fans to take in hot air and blow out cold air. On the other hand, the server cooling is autonomously controlled with built-in server fans via many temperature sensors in the server. That is, both a large fan outside the server and a small fan in the server contribute to cooling the server. We consider that, if a large CRAC fan works mainly with the server cooling fan, the total electric power needed to cool the server can be reduced. For the same air flow, the fan speed of a big fan is less than that of a small fan. Our combined cooling system is based on keeping the temperature of the Central Processing Unit (CPU) fixed. The CRAC fan is controlled following the Model Predictive Control (MPC) using the CPU temperature. The server fan is programmed to a lower speed than the one that would be used by a standard controller by issuing commands to the Intelligent Platform Management Interface (IPMI). We verified the proposed system by using an actual container data center. The results show that the proposed system realizes power savings of more than 30% compared to the standard control system. In particular, the power-saving effect of the proposed system is large when the cold aisle is 25°C or more.**

Keywords- power-aware cooling system; container data center; CRAC; MPC; IPMI; CPU temperature.

I. INTRODUCTION

A variety of cloud computing services have been proposed in the last few years [1][2]. The Information and Communication Technology (ICT) equipment to provide cloud computing services is stored in leased facilities called Data Centers (DCs), which are strictly managed. The number of DCs increases as cloud computing spreads and develops. Therefore, power consumption increases, and lowering the power consumption of data centers is one of the pressing issues of the global environment.

A new cooling system for Container Data Centers (CDCs) that incorporates fan-less servers and a fresh air cooling method to minimize power consumption has been proposed [3]. As a result, a 22.8% energy saving was achieved with this system compared with the conventional container servers with built-in fans. However, the new system was not practical because it required special conditions, for example, fresh air cooling and the use of fan-less servers.

The typical server (x86 server) has built-in fans and cools itself automatically. For instance, the Baseboard Management Controller (BMC) can control the fan so that the fan speed in the server will increase as the temperature of the CPU increases. There are various places where the temperature has to be observed, for example, the CPU, the memory, the Power Supply Unit (PSU), and the exhaust of the server.

The CRAC controls the temperature, humidity, and flow of air using fans in the CDC. Two types of fans (built-in fans in the server and larger fans in the CDC) exist for the purpose of cooling the server. For the same air flow volume, the power consumption of a large fan is less than that of a small fan. However, these two types of fans are controlled independently. Therefore, we thought that we could obtain a further power saving by cooperatively controlling the two types of fans.

In this paper, we propose a cooling control architecture for the server and CRAC. The CRAC fan and the server fan are controlled by a manager server to bring the temperature below a constant temperature (for example, the CPU temperature) with the proposed architecture.

The paper is structured as follows. In Section II A, we explain power consumption and the control of a standard server fan. In Section II C, we explain our cooling control architecture. Section III discusses the experimental results. Finally, Section V concludes this study.

II. COOLING CONTROL ARCHITECTURE

A. Power consumption and control of server fan

The server has two or more internal cooling fans. These fans are controlled by an algorithm saved to the firmware based on the value of the temperature sensor in the server. In the case of the RX200S7 [4], which is made by Fujitsu, there are 6 tandem fans in the front. Figure 1 shows the relation between the server fan speed and the power consumption. In this case, the server fan speed was changed by us regardless of the sensor temperature, and the power consumption of the server was measured. When the server fan speed was raised to 80%, the power consumption increased by about 60 W. The power consumption of a fan is proportional to the cube of the rotation speed. When fan duty reaches 100%, as much as about 100 W is consumed by the fans only. The speed of the built-in fan of a standard server is controlled by the value of the temperature sensor in the server. The control algorithm is programmed into the BMC and executed. The rotation speed of the fan can be controlled with IPMI. However, the server vendor has not opened to the public the IPMI

 ISBN: 978-1-61208-406-0

command for fan control. In this paper, we controlled the built-in server fan by using a IPMI vendor-specific command.

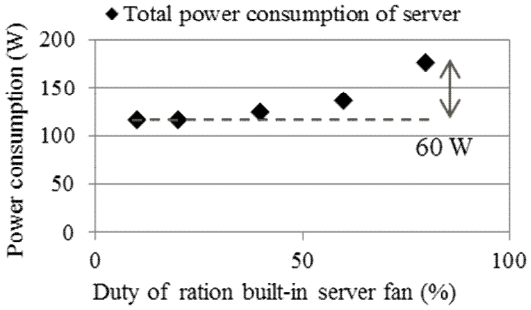


Figure 1. Power consumption of built-in server fan.

B. *Cooling Control method of CRAC fan*

The CRAC has a certain cooling capacity when the computer room is fully loaded with servers. The CRAC controls the air temperature and the air flow rate. However, when the cooling space is as narrow as that of the container data center, the air flow rate greatly affects the cooling capability. We have already proposed a cooling control method based on the Model Predictive Control (MPC) [5] for a container data center directly utilizing fresh air [6]. In our MPC method, the CRAC fan was controlled so that the highest CPU temperature in the rack might be made the predetermined CPU temperature.

C. *Concept Architecture*

For the same air volume, using the large fan rather than the small one saves energy. The air flow of CDC is ideal. In this case, Newton's law of cooling model is applicable for a closed space like the CDC. The equation for Newton's law of cooling is shown below.

$$\frac{dQ}{dt} = h \times A \times \left(T(t)_w - T_f\right) \qquad (1)$$

where Q is the thermal energy, h is the heat transfer coefficient, A is the heat transfer surface area, T_w is the temperature of the parts of the servers, and T_f is the ambient temperature of the CDC. If the air volume of the CRAC fan is increased, the CPUs of the servers can be cooled. This situation was confirmed with a real machine. A server (RX200 S7) of 30 1U types was installed, and it had two CPUs [E5-2650, 2.0 GHz, thermal design power (TDP) 95 W] [7]. In this paper, the CPU temperature refers to the temperature measured by the Digital Temperature Sensor (DTS). The Platform Environment Control Interface (PECI) temperature of this CPU is 89°C. That is, if the CPU temperature reaches 89°C, it begins to lower the clock frequency with the PECI. The Intel Power Thermal Utility (PTU) was used for the load to the CPU. Figure 2 shows the CPU temperature map in the server rack. The power consumption of the entire rack and fan rotation speed of CRAC is shown at the top of the rack map in Figure 2. On the left, the numbers 15-42 indicate the location of the server in the rack. Setting the fan of the CDC for minimum rotation is shown in Figure 2 (a), and setting the rotation of the fan to 4360 rpm is shown in Figure 2 (b). The CPU temperature becomes low when the rotation of the CDC fan increases to 4360 rpm. It can be easily seen that the CPU temperature color of the entire rack changed. The power consumption of the server has decreased from 8.8 kW to 8.6 kW. As the rotation of the CRAC fan increased, the rotation of the server fan decreased, which led to a decrease in power consumption. The rotation of the server fan had slowed because the CPU temperature decreased in this experiment. The CPU temperature decreased from 73°C to 66°C for the U20 server, and the power consumption decreased from 328 W to 320 W. The rotation speed of the U20 server fan decreased from 7920 rpm to 6120 rpm. In addition, if the server fan can be slowed down further to maintain the CPU temperature, the power consumption of the server can be reduced. This is one advantage of this architecture. The CPU temperature at U20 is 66°C owing to the excess cooling. Our cooling algorithm aims to keep the CPU temperature constant with both a large CDC fan and a small server fan. The electric power of CRAC and the electric power of the built-in fan are compared and controlled so that the total electric power consumption can be reduced.

Figure 3 (a)-(b) shows the details of the proposed

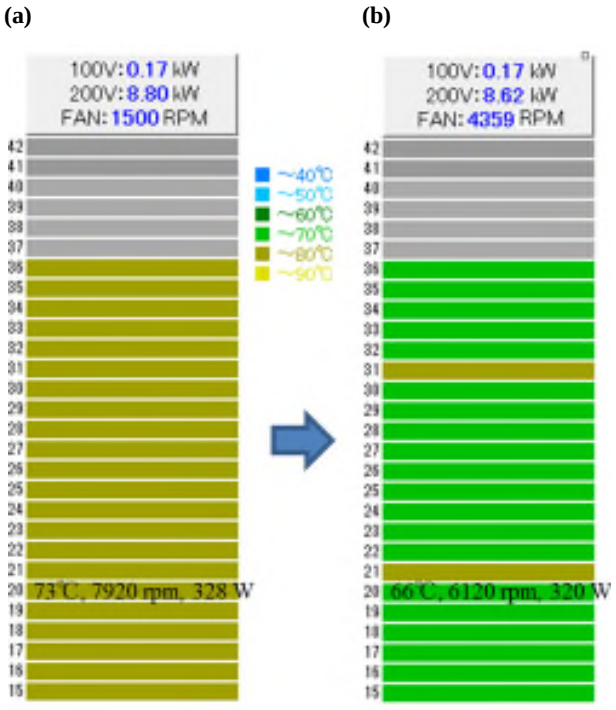
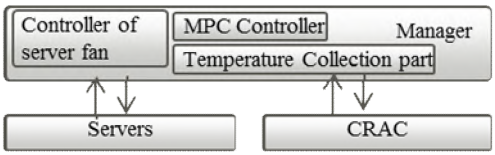


Figure 2. CPU Temperature map in server rack. CPU temperature, server fan speed, and power consumption of server at U20 are shown. Total power consumption and CDC fan speed are shown at top of rack map. (a) CDC fan speed of 1500 rpm. (b) CDC fan speed of 4360 rpm.

 ISBN: 978-1-61208-406-0

a) Block diagram



b) Flow chart

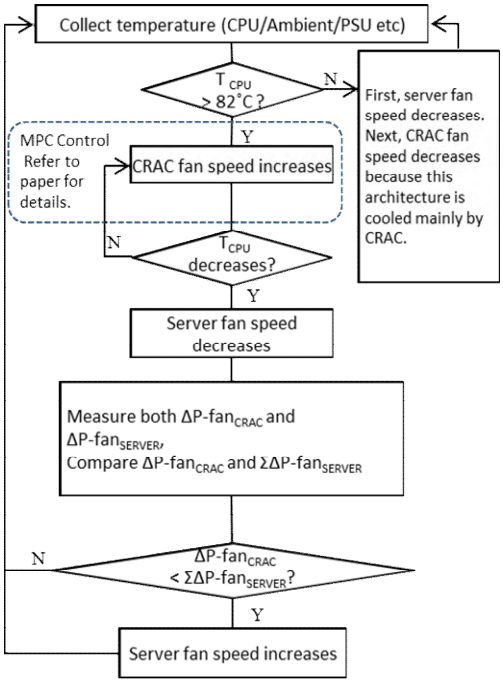


Figure 3. Block diagram of the proposed architecture.

architecture. The manager collected the temperatures of the CPUs and power consumption from the servers with IPMI. Of course, the manager collected the ambient temperature and other component temperatures from the BMC with IPMI. The server fan speed was controlled by the manager so that it was not based on the load of the server but was set to keep the CPU temperature at a fixed temperature. We called this method the server power-saving system (SEPOSS). The manager controlled the server fan speed with SEPOSS, and the CRAC fans were controlled by the MPC method [6]. The manager performed power-saving control by both SEPOSS and MPC and kept the CPU temperature constant at any load. In this architecture, the CPU temperature was regulated at 82°C. Figure 3(b) shows the processing of the case where the temperature of the CPU (T_{CPU}) is near 82°C (regulated temperature). The CRAC fan speed increases with MPC control. If T_{CPU} decreases, the manager decreases the server fan speed. The increment of electric power to increase the speed of the CRAC fan is assumed to be $\Delta P\text{-fan}_{CRAC}$. The increment of electric power to increase the speed of the server fan is assumed to be $\Delta P\text{-fan}_{server}$. The manager compares $\Delta P\text{-fan}_{CRAC}$ and $\Sigma\ \Delta P\text{-fan}_{server}$. The manager chooses the fan that does not have much power consumption.

III. EXPERIMENTS WITH ACTUAL EQUIPMENT

A. *Results of regulating CPU temperature control*

The CPU temperature control that we propose was experimented using an actual CDC [8]. Figure 4 shows the layout of our CDC. In this experiment, 26 servers were installed in rack 7 and rack 5, and the facilities used CRAC_A and CRAC_B. To adjust the maximum temperature of the CPUs of the servers installed in rack 7 to 82°C, CRAC_A was controlled by MPC. Similarly, CRAC_B was controlled at the maximum temperature of the CPUs of rack 5. The heaters were installed in the rack so that the quantity of heat per set of racks could be set to about 18 kW. Moreover, the partition was set up at the center of the CDC. The temperature of the Supply air (SA) of CRAC was 20°C.

(a) Layout of IT equipment and CRACs

(b) Appearance

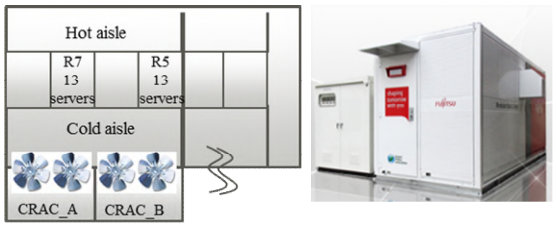


Figure 4. The CDC used for experiment.

The CPU load given to a server changed every 30 min in this order: 50%, 100%, idle (0%), and 80%. The highest temperature of the CPUs in rack 5 is shown in Figure 5 (a). The red line is the standard control, and the blue line is the regulated control. In standard control, the temperature of the CPU with loads of 50%, 100%, 0%, and 80% was 73°C, 78°C, 42°C, and 76°C, respectively. On the other hand, in regulated control, the CPU temperature was almost regulated to 82°C. When CPU loading was switched from 50% to 100%, some overshooting was observed. The temperature of the CPU was measured every 10 seconds, and overshooting was observed for 20 seconds. The air flow volume of the CRAC_A fan is shown in Figure 5 (b). In regulated control, when the loads of the server were 100% and 80%, the air flow volume increased. The comparison of the total power consumption of the standard control and cooperative control is shown in Figure 6. The regulated control reduces the power consumption by about 30% compared with the standard control. This is the maximum case, and it changes by the air specification of the CRAC fan. For instance, when the CRAC fan is changed from 10,000 m^3/hour to 6000 m^3/hour, the reduction rate is thought to be 10% or more. However, it is clear that the proposed method is able to reduce the power consumption. Therefore, the effectiveness of the proposed method is confirmed.

 ISBN: 978-1-61208-406-0

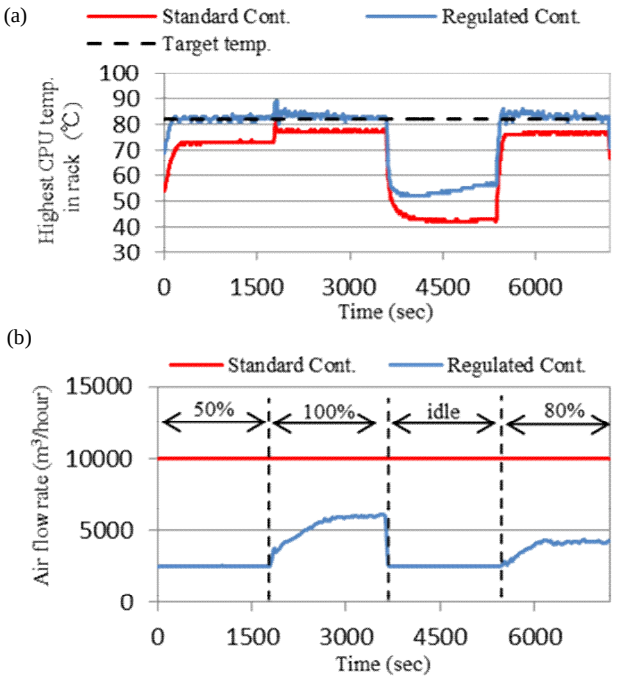


Figure 5. Comparison of standard control and regulated control. (a) Maximum CPU temperature at various loads. (b) Air flow volume of CRAC at various loads.

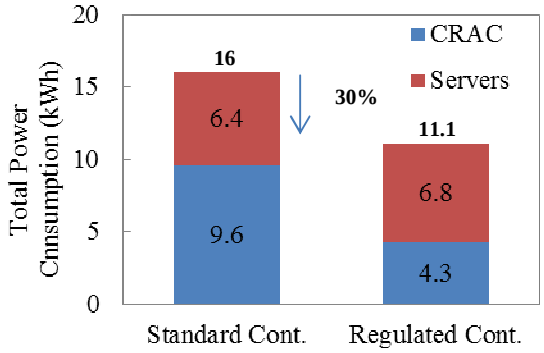


Figure 6. Comparison results of total power consumption

B. *Operation by class of A3 environment*

The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) shows the guidelines of data centers in TC9.9 [9]. In these guidelines, the equipment environment specification of 5°C-35°C is defined as class 3 (A3). In the environment of 35°C, the power consumption of the CRAC decreases compared with the environment of 20°C. On the other hand, a standard server begins to raise the server fan speed when the ambient temperature increases, and the server fan rotates at the highest speed at 35°C. ASHREA reported on page 13 of ASHRAE TC 9.9 [9], "if inlet temperature increases to 35°C, the IT equipment power could increase in the range of 7 to 20% compared to operating at 15°C." That is, even if it sets the environment temperature to a high temperature, the result that improves the power consumption of a server may be caused instead by the power consumption of the CRAC decreasing. In particular, in the regulated control that we propose, the effect of SEPOSS control is that it can perform power-saving in such high temperature environments. The temperature of supply air (SA) in the CDC was changed from 17°C to 35°C, and the standard control was compared with SEPOSS control. The CPU load was not given to the server. Figure 7(a) shows the relation between the environment temperature and power consumption in the CDC. Moreover, the relation of the environment temperature and the server fan speed is shown in Figure 7(b). In standard control, power consumption went up as environmental temperature became high, but the increase in power consumption was suppressed in SEPOSS control. This is based on the inhibiting effect of the server fan speed. In this case, a reduction in power consumption of 16% was confirmed by SEPOSS control. SEPOSS control supervised the temperature of the parts inside the server and did not control fans in accordance with the intake air temperature. We confirmed that the temperature of each part in the server was below the regulated value and also under SEPOSS control. The temperature in the server with an environmental temperature of 35°C is shown in Table I. It does not become a problem because it is below the alarm temperature though the temperature of the CPU when SEPOSS is controlled is about 4°C (CPU1) higher than usual. The part whose alarm

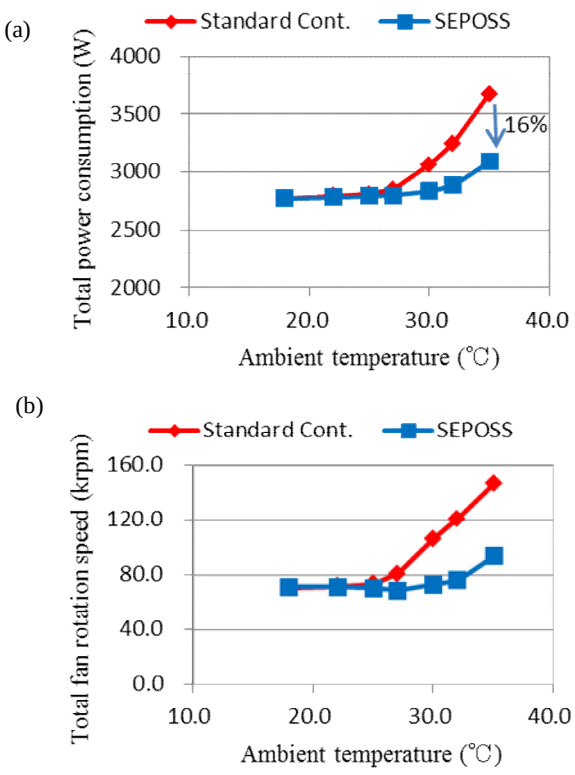


Figure 7. Comparison results of SEPOSS control under high ambient temperature. (a) Total power consumption. (b) Total server fan speed.

temperature is the lowest in the server is the PSU inlet. It is understood that the temperature of the PSU inlet is below the alarm temperature when SEPOSS is controlled.

TABLE I. INTERNAL TEMPERATURE OF SERVER AND ALARM TEMPERATURE OF SEPOSS CONTROL AND STANDARD CONTROL.

[℃]

	Ambient	CPU1	CPU2	Memory	PSU inlet	PSU	System board
Standard	34.6	46.4	47.0	40.9	43.0	62.3	48.2
SEPOSS	34.3	51.0	50.4	42.6	41.4	62.3	51.3
Alarm	37	88	88	78	52	90	75

IV. RELATED WORK

Researchers have investigated the optimization of the temperature control of CRAC in the datacenter. In 2012, T. Hayashi et al. mentioned energy saving systems [10]. This paper presents the energy-saving potential of coordinated control based on experiments. The results indicate that additional energy can be saved by controlling the rotation of fans and temperature settings among multiple CRAC units in response to heat generation due to operation of ICT equipment. In 2011, Wei Huang et al. mentioned power optimization techniques for servers and DCs [11]. They demonstrated a run-time optimization technique that reduces the aggregate server fan power and processor leakage power of a server system. In 2014, S. Yeo et al. mentioned a system where CPU power capping and CRAC control were combined [12]. They proposed a technique which trades off between performances and reliably.

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a power-aware control architecture for the server and CRAC. The CRAC fan and the server fan were controlled by a manager server to bring the temperature below a constant temperature (for example, the CPU temperature) with the proposed architecture. The CRAC fan was controlled by the MPC method, and the server fan was controlled by the SEPOSS method. Our proposed architecture was examined in an actual CDC. As a result, a power saving of 30% was confirmed compared with the standard control case. We would like to fine-tune the MPC and the SEPOSS control in the future. To reduce temperature overshooting during regulated controlling, we need to improve it first. In this experiment, the temperature was controlled from the management server by IPMI commands. The problem of this method is that the number of the controlled servers is limited. When the number of controlled servers exceeds 1000, a delay is expected to be caused in the control due to the management server. We want to give only the target temperature from the management server and improve the control directly by programming the BMC. This method has been actually used with various workloads though the server load remained the same as that of this study.

REFERENCES

[1] J. G. Koomey, “Estimating total power consumption by servers in the U.S. and the world”, Final report, Lawrence Berkeley National Laboratory, Palo Alto, CA, February 15, 2007.
http://hightech.lbl.gov/documents/DATA_CENTERS/svrpwrusecompletefinal.pdf [Retrieved: 4-2015]

[2] R. Brown et al., “Report to Congress on Server and Data Center Energy Efficiency: Public Law 109-431”, Lawrence Berkeley National Laboratory, 2008.

[3] H. Endo, H. Kodama, H. Fukuda, T. Sugimoto, T. Horie, and M. Kondo “Cooperative control architecture of fan-less servers and fresh-air cooling in container servers for low power operation”, Proceedings of the Workshop on Power-Aware Computing and Systems (HotPower 2013). Nov. 2013. doi: 10.1145/2525526.2525844.

[4] Fujitsu, “PRIMERGY RX200 S7 SERVER, Upgrade and Maintenance Manual”, pp152, 2013.
http://manuals.ts.fujitsu.com/file/10630/rx200s7-umm-en.pdf [Retrieved: 4-2015]

[5] J. M. Maciejowski: “Predictive Control with Constraints” Pearson Education, 2002. ISBN: 9780201398236

[6] M. Ogawa et al., “Cooling control based on model predictive control using temperature information of IT equipment modular date center utilizing fresh-air” , 13th International Conference on Control, Automation and Systems(ICCAS 2013), pp. 1815-1820. Oct. 2013, doi: 10.1109/ICCAS.2013.6704235.

[7] Intel, “Intel Xeon Processor E5-1600/E5-2600/E5-4600 v2 Product Families”, p. 12, March 2014.
http://www.intel.com/content/dam/www/public/us/en/documents/datasheets/xeon-e5-v2-datasheet-vol-1.pdf [Retrieved: 4-2015]

[8] Fujitsu, Datacenter Product Modular Data Center, http://jp.fujitsu.com/platform/server/container/ [Retrieved: 4-2015]

[9] White paper prepared by ASHRAE Technical Committee (TC)9.9, “2011 Thermal Guidelines for Data Processing Environments -Expanded Data Center Classes and Usage Guidance”, p. 13, 2011.
http://ecoinfo.cnrs.fr/IMG/pdf/ashrae_2011_thermal_guidelines_data_center.pdf [Retrieved: 4-2015]

[10] T. Hayashi, T. Tominaga, K. Saigo, and P. Gemma, “Minimum data set for controlling data center equipment for energy saving management” Power and Energy Society General Meeting, 2012 IEEE, July 2012, doi: 10.1109/PESGM.2012.6344974.

[11] Wei Huang, et al., “TAPO: Thermal-Aware Power Optimization Techniques for Servers and Data Centers”, Green Computing Conference and Workshops (IGCC), July 2011, doi: 10.1109/IGCC.2011.6008610.

[12] S. Yeo, M. M. Hossain, Jen-Cheng Huang, and Hsien-Hsin S. Lee, “ATAC: Ambient Temperature-Aware Capping for Power Efficent Datacenters”, Proceedings of the ACM Symposium on Clud Cpmputing 2014 (SoCC ’14), Nov. 2014, doi: 10.1145/2670979.2670996.

 ISBN: 978-1-61208-406-0

QoS-aware Interference Control in OFDMA Femtocell Networks

Chiapin Wang, Te-Sheng Tsai, and Hsing-Jung Li
Department of Electrical Engineering
National Taiwan Normal University
Taipei, Taiwan
e-mail: chiapin@ntnu.edu.tw; tkyamato@gmail.com; ted30209@gmail.com

***Abstract*—In this paper, we present a Quality-of-Service (QoS)-aware interference control scheme in femtocell networks. The proposed scheme dynamically adjusts the transmission power of Home Evolved Node B (HeNB) based on its "troubling" property and "servicing" property so as to balance the provisioning of QoS for users and the avoidance of interference to neighboring HeNBs. The simulation results illustrate that our approach can effectively improve the efficiency of resource utilization and aggregate throughput of the femtocell networks in comparison with other schemes.**

Keywords-3GPP Long Term Evolution (LTE); femtocell; interference mitigation; power control

I. INTRODUCTION

The Long Term Evolution (LTE) access technologies developed by the Third Generation Partnership Project (3GPP) group [1][2] are promising Broadband Wireless Access (BWA) solutions of providing high data-rate transmissions, extensive-area coverage, and high-speed mobility for a variety of wireless applications. However, with the growing number of mobile devices and various application demands, the traffic congestion problem will still be inevitable due to limited frequency spectrums even with the BWA technology. Frequency reuse might be the most suitable solution to increase the total capacity of an area. Moreover, once mobile users enter into a building, the radio signal strength can drop tremendously due to a large path loss especially when the building is made up of the concrete and steel. Therefore, Home Evolved Node B (HeNB), also called femtocell is expected to prevail in the LTE networks in the future because of certain large amounts of indoor users and further for its capability to solve the problem of spectrum insufficiency by frequency reuse.

In femtocell networks, the interference problem can be severe if no proper power control or resource management scheme is presented. As a matter of fact, the interferences in femtocell networks can be classified as cross-tiered and co-tiered interferences. When the macrocell and femtocell have an overlap area and use both the same frequency spectrum and time slot for data deliveries, the cross-tiered interference would be incurred. On the other hand, if a number of HeNBs are deployed close to each other, the interference between neighboring HeNBs is so called the co-tiered interference [3]-[10].

In this paper, we tackle the co-tiered interference problem between neighboring HeNBs in the LTE networks and propose simple and practical interference control and resource allocation schemes in a distributed manner. The proposed scheme dynamically adjusts the power of HeNBs according to its "troubling" property and "servicing" property to simultaneously reduce the interference to neighboring HeNBs and improve the overall transmission efficiency of the femtocell networks. Our key idea is based on the observation that each HeNB simultaneously acts as a "servicing node" for data deliveries to the associated users, and acts as a "troubling node" with respect to the neighboring HeNBs. Thus the power control of each HeNB should simultaneously consider the two roles so as to improve its individual transmission efficiency, as well as to lessen the interference in femtocell networks. We conducted simulations to compare the performance of our proposed power control algorithm with that of other approaches. The results demonstrate that our approach can effectively improve the efficiency of resource utilization and aggregate throughput of the femtocell networks in comparison with other schemes.

The rest of this paper is organized as follows. Section 2 introduces the existing studies of interference mitigation in wireless networks. The problem description is presented in Section 3. Section 4 illustrates the proposed power control algorithms. Section 5 explains the simulation setup and results. Finally, the conclusion is given in Section 6.

II. RELATED WORK

There have been many efforts on the mitigation of interferences and/or the cooperation of resource allocation in previous studies [4]-[10]. Sun et al. [4] proposed a subcarrier allocation scheme based on the auction algorithm for solving the cross-tiered interference between macrocells and femtocells. Hoteit et al. [7] proposed a game-theoretic resource allocation scheme in cooperative femtocell OFDMA networks. From the evaluation results, it is shown that their proposed game-theoretic scheme outperforms other proposals in terms of throughput, fairness, and computation time. The studies in [8] proposed a cognitive-radio based scheme which uses cognitive relays to improve the cross-tier interference performance between macrocells and femtocells. However, these centralized schemes can be complex and unpractical. When the femtocell networks are expected to deployed in indoor environments such as home and enterprise buildings, it is hard to adopt a centralized management scheme for resource allocation or power control because the femtocell management system will maintain a growing data base for global interference information with the growth of femtocell deployment, and the computational complexity will dramatically increase and remain a challenge issue.

 ISBN: 978-1-61208-406-0

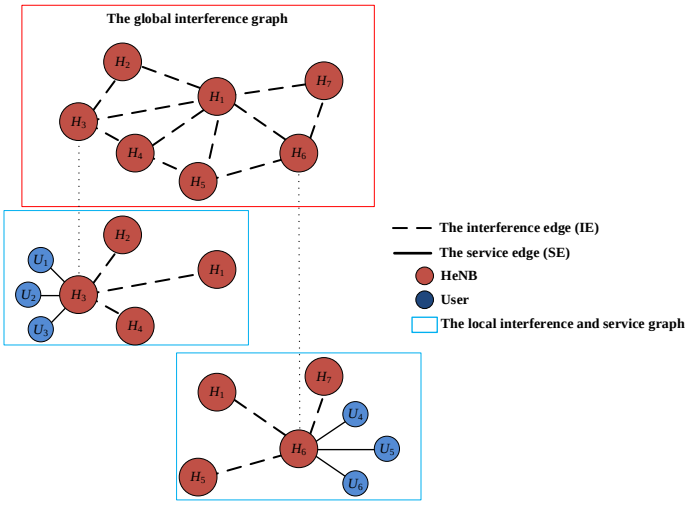


Figure 1. Example of network model in a co-tier interference scenario

III. PROBLEM DESCRIPTION

To solve the optimization problems of interference control and resource allocation, we model the service and interference scenarios in the downlink of femtocell networks as a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V}$ is the set of vertices and $\boldsymbol{E}$ is the set of edges. Figure 1 shows an example of the modeled graph. There are two types of vertices, namely, the HeNB and UE vertices. The HeNB vertex is connected with two kinds of edges including Service Edges (SE) and Interference Edges (IE), which represent the connection with the served users and interfered neighboring femtocells, respectively. For HeNB H_i, let $\boldsymbol{S_i}$ denote the set of served users and $|\boldsymbol{S_i}|$ is the number of served users; let $\boldsymbol{I_i}$ denote the set of interfered neighboring femtocells and $|\boldsymbol{I_i}|$ is the number of interfered femtocells. In the downlink transmission scenario, $\boldsymbol{I_i}$ consists of the femtocells which have users being interfered by HeNB H_i. Thus, if $H_j \in \boldsymbol{I_i}$, we have

$$\frac{RSS(H_j, U_k)}{RSS(H_i, U_k)} < SINR_{th}, U_k \in \boldsymbol{S_j}, \quad (1)$$

where $RSS(H_i, U_h)$ denotes the Received Signal Strength (RSS) of UE U_h from the serving HeNB H_i; $SINR_{th}$ is the threshold of Signal to Interference plus Noise Ratio (SINR) for data transmission.

For HeNB H_i, let $SE_{H_i,U_h}, U_h \in \boldsymbol{S_i}$ denotes the weight edge between HeNB H_i and user U_h, representing the degree of service for U_k given by H_i. In this paper, we adopt the QoS class identifiers (QCIs) specified in 3GPP [2] to quantify the service degree $SE_{H_i,U_h}, U_h \in \boldsymbol{S_i}$. As Table I shows, each QCI is associated with a specific resource type, priority and packet delay budget. Denote QCI_x as the QCI with index x. Let $\boldsymbol{QCI_{GBR}}$ denote the set of GBR-type QCIs and $\boldsymbol{QCI_{Non\text{-}GBR}}$ denote the set of Non-GBR-type QCIs. It is shown in Table I that $\boldsymbol{QCI_{GBR}}$ consists of $\text{QCI}_1, \text{QCI}_2, \text{QCI}_3$, and QCI_4, while $\boldsymbol{QCI_{Non\text{-}GBR}}$ consists of $\text{QCI}_5, \text{QCI}_6, \text{QCI}_7, \text{QCI}_8$, and QCI_9.

To reflect the service degree according to the priority of QCI, we define the weight of QCI_x, w_{QCI_x} as 10 - p_{QCI_x}, where

TABLE I. QCI CHARACTERISTICS [2]

QCI	Resource Type	Priority	Packet Delay Budget	Example Services
1	GBR	2	100 ms	Conversational Voice
2		4	150 ms	Conversational Video (Live Streaming)
3		3	50 ms	Real Time Gaming
4		5	300 ms	Non-Conversational Video (Buffered Streaming)
5	Non-GBR	1	100 ms	IMS Signalling
6		6	300 ms	Video (Buffered Streaming), TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		7	100 ms	Voice, Video (Live Streaming), Interactive Gaming
8		8	300 ms	Video (Buffered Streaming), TCP-based (e.g., www, e-mail, chat, ftp, p2p file, sharing, progressive video, etc.)
9		9		

p_{QCI_x} is the priority of QCI_x. Thus, we have $w_{\text{QCI}_1} = 8$, $w_{\text{QCI}_2} = 6$, $w_{\text{QCI}_3} = 7$, $w_{\text{QCI}_4} = 5$, $w_{\text{QCI}_5} = 9$, $w_{\text{QCI}_6} = 4$, $w_{\text{QCI}_7} = 3$, $w_{\text{QCI}_8} = 2$, $w_{\text{QCI}_9} = 1$, respectively. Assume the service type of $U_h, U_h \in \boldsymbol{S_i}$ is with QCI_x, then SE_{H_i,U_h} can be expressed as

$$SE_{H_i,U_h} = w_{\text{QCI}_x}, U_h \in \boldsymbol{S_i}. \quad (2)$$

Based on the LTE standards [2], Resource Block (RB) is the smallest transmission unit in the LTE system, which consists of 2 time slots (1 ms) in time domain and 12 sub-carriers (180 KHz) in frequency domain. Consider an interference femtocell network consisting of K HeNBs, which share F RBs in downlink transmissions. For femtocell i, let σ_i^f, f = 1, 2, …. F, denote a binary variable which is 1 if RB f is allocated to femtocell i, and 0 otherwise. For RB f, the PHY data rate r_f can be expressed as [11],

$$r_f = \frac{R_f^{(c)} \log_2(M_f)}{T_s N_{RB}} DSC_{RB}, \quad (3)$$

where $R_f^{(c)}$ is the code rate associated with the Modulation and Coding Scheme (MCS) used by RB f; M_f is the constellation size of the MCS used by RB f; T_s is the OFDM symbol duration; N_{RB} is the number OFDM symbols in a RB, which can be 14 or 12 depending on whether a normal or extended cyclic prefix is used; DSC_{RB} is the number of data-carrying subcarriers for a RB.

Define the RB allocation problem as a constrained optimization problem which aims to maximize the overall spectrum efficiency of femtocell networks. The optimization problem can be formulated as follows:

$$\max \sum_{1 \le i \le K} \sum_{1 \le f \le F} \sigma_i^f \cdot r_f \quad (4)$$

subject to

$$P^{\min} \le P_i \le P^{\max}, \forall i \quad (5)$$

$$\sigma_i^f \sigma_j^f = 0, \forall H_j \in \boldsymbol{I_i}, \forall i \quad (6)$$

 ISBN: 978-1-61208-406-0

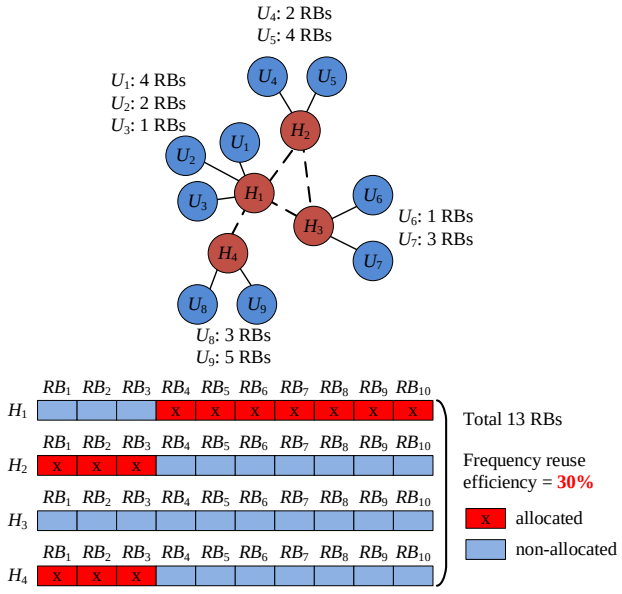


Figure 2. Example of RB allocation in a co-tier interference scenario

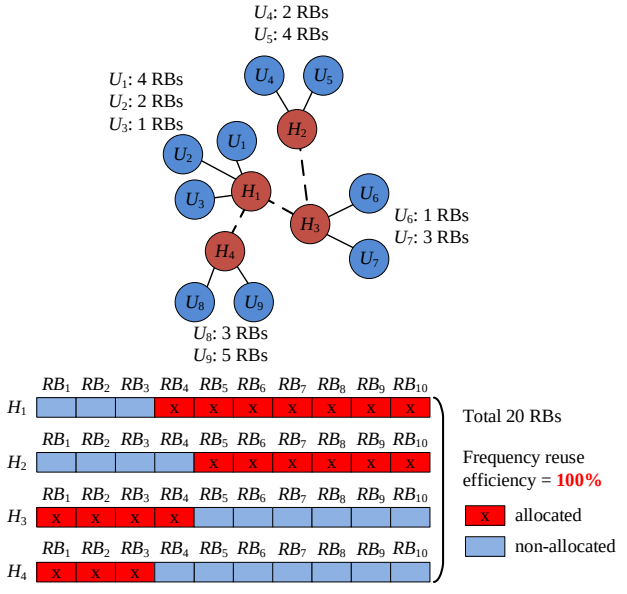


Figure 3. Example of RB allocation with the proposed power adjusting scheme

$$a_{U_h} = \gamma_{U_h}, \text{ for } \mathrm{QCI}(U_h) \in \boldsymbol{QCI_{GBR}}, \forall U_h \quad (7)$$

$$1 \le a_{U_h} \le \gamma_{U_h}, \text{ for } \mathrm{QCI}(U_h) \in \boldsymbol{QCI_{Non\text{-}GBR}}, \forall U_h \quad (8)$$

$$\sum_{1 \le f \le F} \sigma_i^f = \sum_{U_h \in \boldsymbol{S_i}} a_{U_h}, \forall i \quad (9)$$

where P_i is the transmission power of HeNB i; $P^{\min}$ and $P^{\max}$ are the specific minimum and maximum transmission power of LTE femtocell, respectively. γ_{U_h} is the amount of RBs requested by U_h, and a_{U_h} is the amount of allocated RBs. Equation (5) shows the constraints of transmission power for femtocells, and (6) states that the neighboring femtocells which are in the interference domain cannot use the same bands for data transmissions. Equation (7) expresses that the amount of allocated RBs is guaranteed for GBR-type users, whereas (8) expresses that there is no guarantee for Non-GBR-type users [5]. Equation (9) indicates that the RBs allocated to a femtocell are shared by its served users for data transmission.

IV. Proposed Power Control Algorithm

To illustrate the key idea of our power control and resource allocation scheme, consider the example shown in Figure 2. H_1 has three users which request 7 RBs totally; H_2 has two users which request 6 RBs; H_3 has two users which request 4 RBs; H_4 has two users which request 8 RBs. The number of the total required RBs for the four HeNBs is 25. H_1 is regarded as a "troubling node" for its largest amount of interferences over the considered femtocell network (the number of the interference edges of H_1, H_2, H_3 and H_4 is 3, 2, 2, and 1, respectively). If we allocate all the required RBs for H_1 without adjusting its power, the number of the total allocated RBs in the femtocell network is 13, and there are 12 RBs unable to be allocated for H_2, H_3 and H_4 (3 RBs with H_2, 4 RBs with H_3 and 5 RBs with H_4). The frequency reuse efficiency is only 30%. In this example, if the transmission power of H_1 is decreased to a certain level, the overall transmission efficiency can therefore be improved. For instance, consider another situation when the transmission power of H_1 is lessened such that H_1 has only 2 interference edges, as shown in Figure 3. Hence, the number of the total allocated RBs increases as 20 (7 RBs for H_1, 6 RBs for H_2, 4 RBs for H_3 and 3 RBs for H_4) and the frequency reuse efficiency becomes as large as 100%. The example illustrates that a power restriction on the "troubling node" can provide improvements of the overall transmission efficiency in the femtocell network.

Consider K HeNBs in a femtocell network. For HeNB i, let δ_i denote the amount of RBs requested by its served users. δ_i can be expressed as:

$$\delta_i = \sum_{U_h \in \boldsymbol{S_i}} \gamma_{U_h}, \quad (10)$$

1: **for** each cycle of resource allocation
2: Input:
(1) the transmission power of HeNB $i, P_i, (1 \le i \le K), (P_1, P_2, \ldots. P_K)$
(2) the amount of RBs requested by HeNB $i, \delta_i \ (1 \le i \le K)$, $\delta_i = \sum_{U_h \in \boldsymbol{S_i}} \gamma_{U_h}$, γ_{U_h} is obtained with constraints in Eq. (7) and (8)
3: First-step co-channel RB assignment with constraints in Equation (6) and obtain $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots. \boldsymbol{\theta}_K)$, $\boldsymbol{\theta}_K$ is the set of RBs assigned for HeNB i
4: **while** $\sum_{i=1}^{K} |\boldsymbol{\theta}_i| > B$ && $\forall P_i \le P^{\min}$ // check whether the required amount of RBs by co-channel assignment exceeds the amount of allocated RBs, B, and the transmission power of HeNB is smaller than the specific minimum value $P^{\min}$.
5: Sort the K HeNBs in descending order of their power tuning parameter, μ_i, Assume that the order is $\mu_1, \mu_2, \ldots\ldots \mu_K$, i.e., $\mu_1 \ge \mu_2 \ge \cdots \mu_K$
6: $P_1' = \max(P_1 * b, P^{\min}), 0 < b < 1$ // power back off without being lower than the specific smallest power in LTE standards, $P^{\min}$
7: Output:
(1) the adjusted transmission power $(P_1', P_2, \ldots. P_K)$
(2) the new assignment of RBs $(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \ldots. \boldsymbol{\theta}_K')$
8: Replace $\boldsymbol{\theta}_i'$ as $\boldsymbol{\theta}_i$ and P_1' as P_1
9: **end while**
10: **end for**

Figure 4. The proposed QoS-aware power control algorithm (QAPC)

 ISBN: 978-1-61208-406-0

where γ_{U_h} is the amount of RBs requested by U_h. We assume that the co-channel aware scheme is used for RB allocation, i.e., the HeNBs in the interference domain will select different RBs as possible. Assume $\boldsymbol{\theta}_i$ is the set of RBs assigned for HeNB i by the co-channel aware assignment, and $|\boldsymbol{\theta}_i|$ is the number of RBs assigned for HeNB i. $|\boldsymbol{\theta}_i|$ can be expressed as:

$$|\boldsymbol{\theta}_i| = \sum_{U_h \in \boldsymbol{S}_i} a_{U_h}, \tag{11}$$

where a_{U_h} is the amount of RBs allocated to user U_h stated in (7) and (8) for GBR and Non-GBR users, respectively. According to the constraints stated in (7) and (8), we have

$$|\boldsymbol{\theta}_i|^{\max} = \sum_{U_h \in \boldsymbol{S}_i} \gamma_{U_h} = \delta_i, \tag{12}$$

$$|\boldsymbol{\theta}_i|^{\min} = \sum_{\text{all } U_h \in \boldsymbol{S}_i \text{ with QCI} \in \boldsymbol{QCI_{GBR}}} \gamma_{U_h} + \sum_{\text{all } U_h \in \boldsymbol{S}_i \text{ with QCI} \in \boldsymbol{QCI_{Non\text{-}GBR}}} 1, \tag{13}$$

$$|\boldsymbol{\theta}_i|^{\min} \le |\boldsymbol{\theta}_i| \le |\boldsymbol{\theta}_i|^{\max}. \tag{14}$$

Consider B as the total amount of RBs in the femtocell network. If the condition stated in (15) happens,

$$\sum_{i=1}^{K} |\boldsymbol{\theta}_i| > B, \tag{15}$$

the network capacity is insufficient to support the total service requirements of users due to interferences between HeNBs. In this case, we argue that the power level of troubling nodes should be decreased to improve the RB reuse efficiency. However, a trouble node also acts as a service node with respect to its users. Thus, we propose an adaptive power control algorithm which dynamically adjusts the power of HeNB according to both the interference factors and service factors. The interference factor of HeNB i is expressed as $|\boldsymbol{I}_i|$ which is the number of interfered femtocells. The service factor of HeNB i, ψ_i can be expressed as

$$\psi_i = \sum_{U_h \in \boldsymbol{S}_i} SE_{H_i,U_h} \cdot \gamma_{U_h} \tag{16}$$

where SE_{H_i,U_h} denotes the degree of service for U_h by H_i in terms of the QCI weight, which is illustrated in (2). The feasible transmission power of HeNB should be determined by both the interference factor $|\boldsymbol{I}_i|$ and service factor ψ_i. For example, when $|\boldsymbol{I}_i|$ is large, we should decrease the power to lessen the interference to the neighboring HeNBs. On the other hand, when ψ_i is large, we should increase the power to provide service qualities for users. Thus, we define a power tuning parameter, μ_i, which is the ratio of $|\boldsymbol{I}_i|$ to ψ_i. That is,

$$\mu_i = \frac{|\boldsymbol{I}_i|}{\psi_i}. \tag{17}$$

When μ_i is large, we should decrease the power to lessen the interference to the neighboring HeNBs; on the other hand, when μ_i is small, we should increase the power to improve service qualities for users. In case that the network resource is insufficient to support the total service requirements of users due to severe interferences, the power level of HeNBs with the largest μ_i should be decreased by multiplication with a factor b ($0 < b < 1$), That is,

$$i^* = \arg\max_{1 \le i \le K} \mu_i \tag{18}$$

$$P_{i^*} = \max(b \cdot P_{i^*}, P^{\min}). \tag{19}$$

Equation (19) states that the adjusted power cannot be lower than the specific minimum value $P^{\min}$.

After the power of the trouble node is adjusted, the new assignment of RBs by the co-channel aware scheme, ($\boldsymbol{\theta}'_1$, $\boldsymbol{\theta}'_2$, …. $\boldsymbol{\theta}'_K$), can therefore be obtained. The pseudo code of the proposed QoS-aware Power Control (QAPC) is shown in Figure 4. The power control algorithm will be iteratively executed until the network resource is sufficient to support the total amount of RBs allocated for users, or the adjusted power is lower than the specific minimum value $P^{\min}$.

TABLE II. SIMULATION PARAMETERS

Parameter	Value
Map Size	100 m × 100 m
System Bandwidth	5 MHz
Number of cells	3 or 6
Cell distribution	uniform
Number of RBs	25
Frame Structure	FDD
One Subframe	1 ms
Femtocell Min/Max. Tx Power	10 dBm / 20 dBm
Femtocell Noise Figure	8 dB
Femtocell Min/Max. Radius	2 m / 20 m
Distance from Femtocell to User	2 m
Number of Users	5
Thermal Noise	−174 dBm / Hz
Path Loss	Dual-Stripe Model
Inner Wall	5 dB
Number of Penetrated Wall and Floor	1/0
$d_{2D,\text{ indoor}}$	0

V. PERFORMANCE EVALUATION AND DISCUSSION

In this section, we conduct simulation scenarios of the downlink transmissions in the LTE femtocell network to demonstrate the effectiveness of the proposed power control scheme. The simulation is constructed by the C++ programming and followed the LTE standard closely [2]. The LTE PHY parameters and their values are listed in Table II. The simulation setup assumes a 100 m ∗ 100 m square area where the HeNBs are randomly deployed. We assume that the occurrence of each QCI is uniform, while the resource request process is Poisson with a constant size of 1000 bytes. The total simulation period is 120 seconds while the results are provided with the average values over 20 times of simulations. We compare the performance of the proposed power control algorithm with that of fixed power and distributed power control algorithms presented by ElBatt and Ephremides [12]. For all these power control approaches, the co-channel assignment scheme is used for resource allocation (i.e., the assignment of RBs will prevent interference as well as possible) to clearly examine the achieved performances with

 ISBN: 978-1-61208-406-0

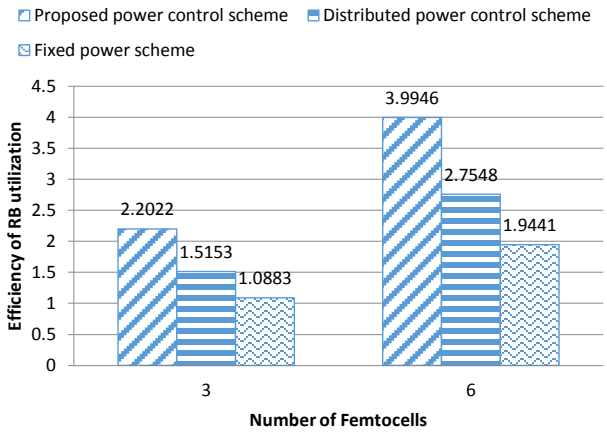


Figure 5. The efficiency of RB utilization with different power control schemes

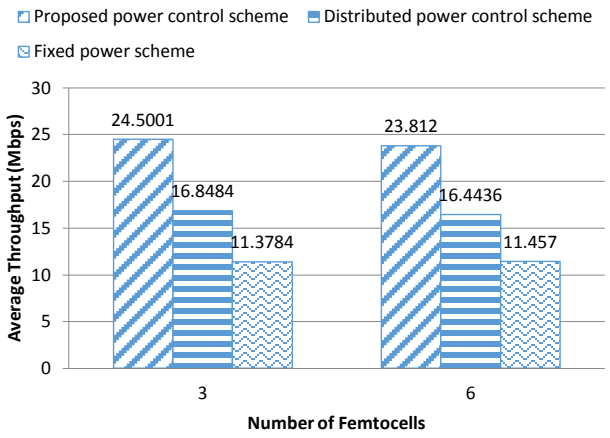


Figure 6. The average throughput with different power control schemes

different power control schemes. The performance metrics are indexed as the efficiency of RB utilization and average throughput.

First, we examine the efficiency of RB utilization with different power control schemes. It is shown in Figure 5 that the proposed CA scheme has the best performance in terms of RB efficiency in the two cases. In the 3-femtocell scenario, our proposed scheme can increase the RB efficiency by 45.33% and 102.35% in comparison with the distributed power control scheme and fixed power scheme, respectively. In the 6-femtocell scenario, our proposed scheme can even increase the RB efficiency by 45.01% and 105.47% with regard to the two schemes, respectively. Figure 6 shows the corresponding throughput performances with different power control schemes. As a result shown in Figure 6, our power control scheme averagely can increase the throughput by as large as 45.41% and 102.35% over the distributed power control and fixed power schemes, respectively, in the 3-femtocell scenario, and by 44.81% and 107.84%, respectively, in the 6-femtocell scenario. The simulation results demonstrate that our power control scheme which adjusts the transmitting power of troubling nodes to lessen interferences to neighboring femtocells can improve the RB utilization efficiency and overall system throughput in the LTE femtocell network.

VI. Conclusion and Future Work

This study tackled the interference problems in femtocell and present an interference control scheme. The proposed scheme dynamically adjusts the transmission power of HeNB based on its "troubling" property and "servicing" property so as to balance the provisioning of QoS for users and the avoidance of interference to neighboring HeNBs. We conducted simulations to compare the performance of our proposed scheme with that of other approaches. The results illustrate that our approach can effectively improve the efficiency of resource utilization and system throughput of the LTE femtocell networks. Our future research will investigate the joint optimization problem with the proposed power control algorithm and Discontinuous Reception (DRX) mechanism for energy-efficiency and QoS performances field testing.

Acknowledgment

This work was supported in part by Taiwan Ministry of Science and Technology under grant MOST 103-2221-E-003-007.

References

[1] 3GPP TS 36.213: "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," V10.4.0, Dec. 2011.

[2] 3GPP TS 23.203: "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Policy and charging control architecture," V10.6.0, March 2012.

[3] G. de la Roche, A. Valcarce, D. López-Pérez, and J. Zhang, "Access Control Mechanisms for Femtocells," IEEE Communications Magazine, vol. 47, no. 9, Sep. 2009, pp. 41 - 48.

[4] Y. Sun; R. P. Jover, and X. Wang, "Uplink Interference Mitigation for OFDMA Femtocell Networks," IEEE Transactions on Wireless Communications, vol. 11, no. 2, 2012, pp. 614 - 625.

[5] Y.-S. Liang, W.-H. Chung, G.-K. Ni, I.-Y. Chen, H. Zhang, and S.-Y. Kuo, "Resource Allocation with Interference Avoidance in OFDMA Femtocell Networks," IEEE Transaction on Vehicular Technology, vol. 61, no. 5, Jun. 2012, pp. 2243 -2255.

[6] N. Saquib, E. Hossain, L. B. Le, and D. I. Kim, "Interference management in OFDMA femtocell networks: issues and approaches," IEEE Wireless Communications, vol. 19, no. 3, 2012, pp. 86-95.

[7] S. Hoteit, S. Secci, R. Langar, G. Pujolle, and R. Boutaba, "A Bankruptcy Game Approach for Resource Allocation in Cooperative Femtocell Networks," Proc. IEEE GLOBECOM, 2012, pp. 1800 - 1805.

[8] W. Wang, G. Yu, and A. Huang, "Cognitive radio enhanced interference coordination for femtocell networks," IEEE Communications Magazine, vol. 51, no. 6, 2013, pp. 37 - 43.

[9] C. T. Do, D. N. M. Dang, T. LeAnh, N. H. Tran, R. Haw and C. S. Hong, "Power Control under QoS and Interference Constraint in Femtocell Cognitive Networks," Proc. IEEE International Conference on Information Networking (ICOIN), 2014, pp. 292 - 297.

[10] P. Lin, J. Zhang, Q. Zhang, and M. Hamdi, "Enabling the Femtocells: A Cooperation Framework for Mobile and Fixed-Line Operators," IEEE Transactions on Wireless Communications, vol. 12, no. 1, 2013, pp. 158 - 167.

[11] R. Kwan, C. Leung, and J. Zhang, "Proportional Fair Multiuser Scheduling in LTE," IEEE Signal Processing Letters, vol. 16, no. 6, 2009, pp. 461-464.

[12] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad hoc networks," IEEE Transaction on Wireless Communications, vol. 3, no. 1, Jan. 2004, pp. 74 - 85.

 ISBN: 978-1-61208-406-0

Constrained Application Protocol Profile for Robust Header Compression Framework

Mikko Majanen, Pekka Koskela and Mikko Valta

VTT Technical Research Centre of Finland Ltd
Email: firstname.lastname@vtt.fi

***Abstract*—Due to the extensive growth of Internet of Things (IoT), the number of wireless devices connected to the Internet is increasing and will continue to increase remarkably in the near future. In wireless networks, the available bandwidth is always restricted. Much of the bandwidth is consumed by protocol overheads, while the actual data payload may be only couple of bytes. In this paper, we propose to use header compression to minimize the protocol overhead, especially for IoT-based communication using the new lightweight Constrained Application Protocol (CoAP). We define a CoAP compression profile for Robust Header Compression (ROHC) framework and evaluate its performance in the name of compressed packet size, delay caused by the compression and decompression, behavior in (wireless) lossy links, and energy efficiency. In our tests, the packet size could be reduced by 91.2% at best by using the proposed header compression. The round-trip delay increased slightly due to the extra processing needed for compression and decompression; however in lossy links with bit error rates $\geq 10^{-3}$ the smaller packet size due to the header compression turned out to be extremely beneficial due to the smaller need for packet retransmissions. In single packet transaction in our test bed, the header compression increased the energy consumption by 2.5%. However, in lossy links energy may be saved due to the smaller need for packet retransmissions. Other possible scenarios for energy savings were also identified as future work items.**

Keywords–Constrained Application Protocol (CoAP); Robust Header Compression (ROHC); delay; energy efficiency.

I. INTRODUCTION

Due to the extensive growth of Internet of Things (IoT), the number of connected devices to the Internet is increasing and will continue to increase remarkably in the near future. For example, Gartner estimates that the IoT, which excludes PCs, tablets and smartphones, will grow to 26 billion units installed in 2020 representing an almost 30-fold increase from 0.9 billion in 2009 [1]. Most of the new IoT devices are wireless, communicating locally for example via Bluetooth Low Energy, IEEE 802.15.4, or IEEE 802.11ah. Usually, a gateway node is needed to connect this local wireless network into the Internet. The Internet connection is in many cases also wireless; it can be for example based on cellular networks (3G, 4G LTE) or satellite networks. In wireless networks, the available bandwidth is always restricted. Usually, the IoT devices are for example sensors that periodically report their data values to the cloud services in the Internet. Thus, the transmitted data may be only couple of bytes. The problem is the protocol overheads: at the network layer, IPv4 or IPv6 header takes 20 or 40 bytes, respectively, and TCP or UDP on the transport layer takes 20 or 8 bytes, respectively. On the application layer, Hypertext Transfer Protocol (HTTP), for example, can take easily over 40 bytes.

Constrained Application Protocol (CoAP) [2] is a new proposed standard for a lightweight application layer protocol. It can be thought as a lightweight HTTP that can connect, for example, IoT devices to the Internet. CoAP has a small header overhead and it works on top of UDP, so it has considerably smaller protocol overhead than, for example, HTTP running on top of TCP.

For the IoT devices using CoAP, for example, a sensor node, a very common behavior is to send periodically the sensor data to the server, or vice versa, the server (or user) periodically asks the data value from the sensor. The header part of the packet remains almost constant all the time while only the data part changes. Since there are possibly many wireless hops along the path from the sensor to the server or user, the header overhead consumes the wireless bandwidth much more than the actual transmitted sensor data.

In this paper, we propose to use header compression to further minimize the protocol overhead in wireless links. We define a CoAP compression profile for Robust Header Compression (ROHC) [3] and evaluate its performance in the name of compressed packet size and delay caused by the compression and decompression. We also study its behavior in (wireless) lossy links and its energy efficiency.

The rest of the paper is organized as follows: Section II explores the related work and Section III shortly presents the principles of ROHC. In Section IV we define the CoAP profile for ROHC and in Section V we evaluate its performance in a real test bed environment. Finally, Section VI concludes the paper with future work items.

II. RELATED WORK

Header compression is based on the redundancy between header field values within packets, and especially between consecutive packets belonging to the same flow. For example, many header fields remain constant between packets, or change according to a known pattern. Over a single link, not all that information is needed and part of it can be temporarily removed, i.e., the full IP packet has to be re-created on the receiving side of the link.

Header compression is not a new idea. The first header compression scheme was compressed TCP (CTCP) [4] developed in 1990. IP Header Compression [5] can compress IP, UDP and TCP headers. CRTP [6] adds RTP header compression ability to IP Header Compression.

6LoWPAN Header Compression [7] defines a scheme for header compression for IPv6 packets transmitted over IEEE 802.15.4 networks, i.e., packets transmitted in Low Power Wireless Personal Area Networks (6LoWPANs) [8]. The compression format relies on shared context to allow compression of arbitrary prefixes. The approach is able to compress IPv6 and UDP headers. Also, it is meant to be used only inside 6LoWPANs. [9] is an addition to 6LoWPAN

 ISBN: 978-1-61208-406-0

Header Compression that enables the compression of generic headers and header-like payloads next to IPv6 header, without a need to define a new header compression scheme for each new such header or header-like payload.

ROHC [3] is a general and extendable framework for header compression, on top of which profiles can be defined for compression of different protocol headers. Currently, there are ROHC compression profiles for RTP, UDP, IP, UDP-Lite, ESP and TCP protocols. Especially, RTP profile is widely used in multimedia transmissions. ROHC has been incorporated into 3G and WiMAX network specifications and standards. Its performance has been evaluated, for example, in [10] and [11]. As can be seen, ROHC does not support CoAP protocol yet. In the next section, we shortly summarize the principles of ROHC, and then we define a missing CoAP profile for ROHC in Section IV.

III. Robust Header Compression (ROHC) principles

ROHC consists of a compressor and a decompressor located on the ends of a link with limited capacity. Redundant information in packet headers is transmitted only in the first packets; in the next packets, only dynamic header parts are transmitted. Packets are classified into flows to take advantage of inter-packet redundancy. A compression profile defines the rules how the packet headers are compressed.

ROHC has three operational modes [12]:

- In **Unidirectional Mode** the packets are sent only in one direction, from compressor to decompressor. This mode is always used in the beginning.
- In **Bidirectional Optimistic Mode** the decompressor uses feedback channel to send error recovery requests and acknowledgments of significant context updates to the compressor.
- In **Bidirectional Reliable Mode** the feedback channel is used more intensively. It aims to maximize robustness against loss propagation and damage propagation, i.e., minimize the probability of context invalidation, even under heavy loss/error burst conditions.

The ROHC compressor has three states [12]:

- The compressor starts in the **Initialization and Refresh (IR)** state, in which it sends full header header information.
- When the compressor is fairly confident that the decompressor has received the static header information correctly, it may proceed to **First Order (FO)** state. In this state, the compressor sends all the irregularities in the packet flow.
- In **Second Order (SO)** state, the compression is optimal. The change patterns of all dynamic header fields are fully exploited.

The ROHC decompressor has three self-explanatory states: **No Context**, **Static Context**, and **Full Context**. The used ROHC packet type depends on the current state of the compressor and decompressor. An interested reader can take a look at [3] for more details on ROHC.

IV. Constrained Application Protocol (CoAP) profile for ROHC

The first step in the header compression consists of identifying and grouping packets together into different ”flows”, so that packet-to-packet redundancy is maximized in order to improve the compression ratio [3]. Grouping packets into flows is usually based on source and destination host (IP) addresses, transport protocol type (e.g., UDP or TCP), process (port) numbers, and potentially additional unique application identifiers, such as the synchronization source (SSRC) in RTP. The compressor and decompressor each establish a context for the packet flow and identify the context with a Context Identifier (CID) included in each compressed header.

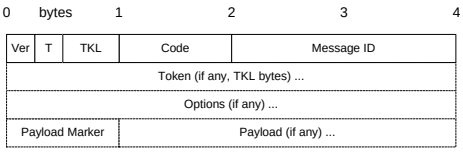


Figure 1. CoAP message format [2]

For the sensors using CoAP protocol, a very common behavior is to send periodically the sensor data to the server, or vice versa, the server (or user) periodically asks the data value from the sensor. CoAP PUT and GET request messages are used for these two, respectively. So, the PUT or GET request messages form one flow, and the corresponding responses to these requests (i.e., the ACK/RESPONSE messages) form another flow in the opposite direction.

The second step is to understand the change patterns of the various header fields. On a high level, header fields fall into one of the following classes [3]:

- INFERRED: These fields contain values that can be inferred from other fields or external sources, for example, the size of the frame carrying the packet can often be derived from the link layer protocol, and thus does not have to be transmitted by the compression scheme.
- STATIC: Fields classified as STATIC are assumed to be constant throughout the lifetime of the packet flow. The value of each field is thus only communicated initially.
- STATIC-DEF: Fields classified as STATIC-DEF are used to define a packet flow as discussed above. Packets for which respective values of these fields differ are treated as belonging to different flows. These fields are in general compressed as STATIC fields.
- STATIC-KNOWN: Fields classified as STATIC-KNOWN are expected to have well-known values, and therefore their values do not need to be communicated.
- CHANGING: These fields are expected to vary randomly, either within a limited value set or range, or in some other manner. CHANGING fields are usually handled in more sophisticated ways based on a more detailed classification of their expected change patterns.

The CoAP message format [2] is depicted in Figure 1. As can be seen, a CoAP message consists of a fixed length header information, followed by a (possibly zero-length) token, and optionally options and payload. The option format is depicted in Figure 2.

Version (Ver) is a 2-bit unsigned integer indicating the CoAP version number. Currently, it has to be set to 1, so it can be treated as STATIC-KNOWN in ROHC for now. In the

 ISBN: 978-1-61208-406-0

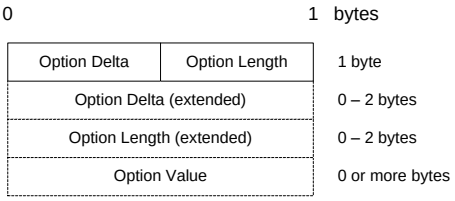


Figure 2. CoAP option format [2]

future, there might be other versions of CoAP, so the version number could then be STATIC.

Type (T) is a 2-bit unsigned integer indicating if this message is of type Confirmable (0), Non-Confirmable (1), Acknowledgement (2) or Reset (3). As said in the first step, the type of the packet is used to define the flow, so the Type could be treated as STATIC-DEF.

Token Length (TKL) is a 4-bit unsigned integer indicating the length (0-8 bytes) of the **Token** value. Token value is used for correlating requests and responses. The server shall use the same token in the response as it was in the request. The tokens currently in use for a given source/destination endpoint pair shall be unique. Thus, when sending the requests sequentially and having piggy-packed responses, the token can be kept as constant (or even as zero-length). In that case, the Token and Token Length can be considered as STATIC-DEF. In other cases, it should be treated as CHANGING.

Code is a 8-bit unsigned integer indicating if the message carries a request (1-31) or a response (64-191), or is empty (0). In case of a request, the Code field indicates the Request Method; in case of a response a Response Code. Code can be used to define the flow, so it can be treated as STATIC-DEF.

Message ID is a 16-bit unsigned integer that is used for the detection of message duplication, and to match messages of type Acknowledgement/Reset and messages of type Confirmable/Non-confirmable. The same Message ID should not be re-used in the communication with the same endpoint within EXCHANGE_LIFETIME, so the Message ID should be treated as CHANGING. An Acknowledgement or Reset message related to a Confirmable message has to repeat the same Message ID as in the Confirmable message.

Option Delta is a 4-bit unsigned integer indicating the difference between the **Option Number** of this option and the previous option (or zero for the first option). In other words, the Option Number is calculated by simply summing the Option Delta fields of this and previous options before it. Values 13-14 together with the Option Delta extension field are used for deltas larger than 12, and the value of 15 is reserved for the payload start marker (see below).

Option Length indicates the length of the **Option Value**, in bytes. Normally Length is a 4-bit unsigned integer allowing value lengths of 0-12 bytes. For longer options, values 13-14 together with the extension field are used. Length of 15 is not allowed.

Value is a sequence of exactly Option Length bytes. The length and format of the Option Value depends on the respective option. Option Value may be an unsigned integer (uint), a string, opaque, or empty.

Options include for example the Uri-Host, Uri-Port, Uri-Path, Uri-Query, Content-Type, and some matching options (If-Match, If-None-Match) [2]. Uri-Path and Uri-Query specify the path to the resource requested, which can be treated as STATIC-DEF in the flow. Content-Type is usually constant in the flow (e.g., sensor sending periodically the same measurement data), so it can be treated as STATIC-DEF. Also other options can usually be treated as STATIC-DEF in the flow. The only exception is perhaps the Location-Path that is normally used in a response to POST method and that defines the location path for the newly created resource, and thus it is CHANGING. But remember that POST is only used when creating the resources in the beginning. After that, the resources are updated by using PUT method. So POST is used much more seldom than PUT, and it is perhaps not wise to create a flow for POST messages at all, only for PUT and GET messages and their responses. Thus, basically all the options in the CoAP PUT and GET flows can be treated as STATIC-DEFs. Moreover, all the options can be treated as one combined bundle; there is no need for handling every single option as separate in compression/decompression.

Payload is preceded by the **Payload Marker** (0xFF). This marker can be treated as STATIC-KNOWN. It can be handled together with the options. Payload naturally changes and it is not part of the packet header, so header compression is not used for the payload.

Our CoAP profile was implemented for the open source ROHC library [13]. A version 1.6.1 of the ROHC library was used as the basis for the implementation. Basically, the CoAP profile had to implement the CoAP context for the flow, profile and flow identification function, and rules for compressing and decompressing the CoAP header part of the packet. The CoAP context for the flow basically consists of one CoAP header belonging to the flow's packet, in addition to the CID. The profile and flow was identified by using IP addresses, UDP ports, and CoAP Type and Code fields, i.e., the fields defined as STATIC-DEFs earlier. The compression of CoAP header part was based on the static and dynamical header parts discussed earlier in this section. The UDP and IP header parts of the packet were compressed as in the existing UDP/IP profile.

V. Evaluation of the CoAP profile

The performance of the developed CoAP profile was tested in a real test bed environment. First, we tested the compression ability, i.e., how much the packet headers could be compressed by using the CoAP profile. Secondly, we measured the delay caused by the compression and decompression. Thirdly, we studied the performance of header compression in lossy (wireless) links. Finally, we studied the energy efficiency of the header compression.

A. Test bed and testing scenario

Our test bed (see Figure 3) consisted of an old laptop connected to the VTT's CNL laboratory's Willab network via WiFi and a Raspberry Pi (RPi) computer connected to the Willab using Ethernet. An UDP-tunnel was established between the laptop and the RPi by using the tunnel application provided by the ROHC library for easy testing of the library. It is to be noted that the tunneling was used only for practical testing reasons since we do not have a real ROHC implementation integrated in the protocol stack in the both ends of the link. The laptop served as a CoAP server running the example CoAP server software provided by the C-based libcoap-4.0.3 library [14]. The RPi acted as a CoAP client requesting periodically the 'time' resource from the server

 ISBN: 978-1-61208-406-0

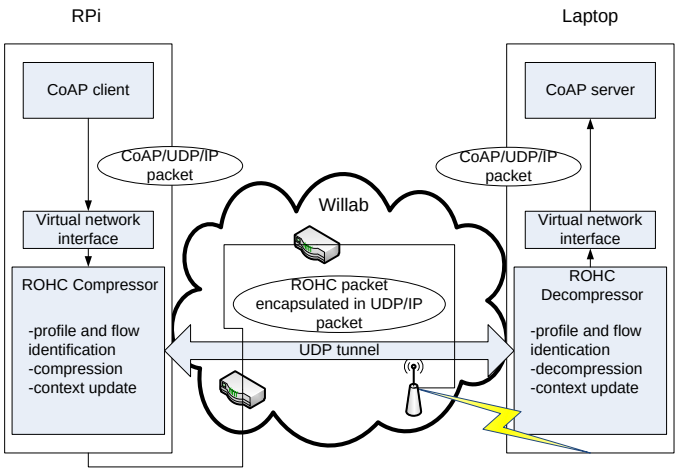


Figure 3. The test bed

using libcoap's example CoAP client software. Thus, it created a GET request flow to the server. The server responded to the requests by sending acknowledgements piggy-packing the CoAP RESPONSEs. The laptop had a 1.73 GHz Intel Pentium M processor and 1 GB of RAM. RPi was a Model B having a 700 MHz ARM CPU processor and 512 MB of RAM.

In the energy-efficiency studies, we used 2 RPis, one as a server and another one as a client. RPis were communicating via an IEEE 802.11g WLAN access point using 54 Mbit/s bit rate. We used D-Link DIR-300 as a wireless access point. The client RPi was connected to the access point with D-Link DWA-121 USB-WLAN dongle and the server RPi with Ethernet connection. The current as a function of time was measured from the client RPi with a current probe (AM503B amplifier) connected to the oscilloscope (Tektronix TDS3032B). The client RPi was Model B+, which is an improved version of prior Model B, having a lower power consumption, for example.

B. *Compressed packet size*

The created CoAP GET request message consisted of the following CoAP header fields:

- version = 1
- type = 0 (confirmable)
- tkl = 0 (no token)
- code = 1 (GET request)
- message ID = 58468 (changing number)
- option Uri-Path = 'time' in TLV format = 0xb474696d65; Type=11, Length=4, Value='time'
- no payload

In the beginning, the compressor starts in the ROHC Initialization and Refresh (IR) state and transmits full headers. The uncompressed packet size was 37 bytes consisting of IPv4 header of 20 bytes, UDP header of 8 bytes, and CoAP header of 9 bytes. One extra byte was needed for the CID in ROHC, so the compressed packet size was in the beginning 38 bytes. However, after couple of packets when the context was created and the compressor changed into First Order (FO) state, only the dynamical (changing) header parts were transmitted and the compressed packet size decreased to 15 bytes with the CoAP profile. The only dynamical part of the CoAP header is the Message ID that takes 2 bytes. So the 9-byte CoAP header was compressed to only 2 bytes, i.e., the savings in the CoAP header part were 77.8%. The rest of the compressed packet was used by the UDP and IPv4 headers. The savings for the whole packet were 59.5%. When the compressor changes into Second Order (SO) state, also the change patterns of the dynamical header parts are taken into account. Since the CoAP Message ID may vary totally randomly, there were no further saving possibilities in the CoAP header part. However, further compression was still available in the other (UDP and IPv4) headers, and the packet could be compressed to only 5 bytes at best. Thus, the savings for the whole packet were 86.5%. The achieved gains are equivalent to the gains of about 85% reported, e.g., in [11] and [15] for RTP profile.

The same packet flow was also compressed with the UDP/IP and IPv4-only profiles. With these profiles, the compressed packet sizes decreased to 22 and 28 bytes, respectively, in the FO state. The savings were 40.5% and 24.3%, respectively. UDP/IP profile compresses only the UDP and IP header parts, while the IP-only profile compresses only the IP header part. These profiles do nothing for the CoAP header part. In SO state, the IPv4 compression profile compressed the packet to 18 bytes at best, so the savings were only 51.4%.

The server's acknowledgement message piggy-packed also the response, and its length was 51 bytes uncompressed. IPv4 header took 20 bytes, UDP header 8 bytes, CoAP header 8 bytes, and CoAP payload 15 bytes. The CoAP header part consisted of the following header fields:

- version = 1
- type = 2 (ACK)
- tkl = 0 (no token)
- code = 69 (2.05 Content)
- message ID = 58468 (changing number)
- options Content-Format (Length = 0) and Max-Age (length 1 and value 1) in TLV format = 0xc02101;
- Payload Start Marker = 0xFF
- Payload = 'Oct 14 19:45:32'

For the ack/response packet, the header parts could be compressed as much as in the request messages, but the overall savings were smaller because the payload part could not be compressed. Some of the messages also contained ROHC feedback data. With the CoAP profile, the compressed packet size (including the payload of 15 bytes) was 53 or 58 bytes in IR state (depending on the amount of feedback data), 31 or 22 bytes in FO state, and 21 bytes in SO state. So the savings in the total packet size were 58.8% at best in SO state. With IPv4-only profile, the packet size could be compressed to 33 bytes at best, meaning 35.3% savings in total packet size.

We also studied the packet sizes of the same CoAP request and ack/response messages when using IPv6 protocol. IPv6 protocol header takes 40 bytes, i.e., 20 bytes more than IPv4. Thus, the uncompressed CoAP request packet size was now 57 bytes and the corresponding ack/response packet 71 bytes. In the IR state, the compressed packet sizes (with ROHC header and feedback information) were 61 or 68 bytes for the CoAP request packet and 75 bytes for the ack/response packet. In the next state, when only dynamical header parts were transmitted, the compressed packet sizes decreased to 12 and 27 bytes, respectively. Finally, in the SO state, the compressed packet sizes were only 5 and 20 bytes, respectively, with the developed CoAP profile. As percentages, these mean total savings

 ISBN: 978-1-61208-406-0

of 91.2% and 71.8% in the request and ack/response packet sizes, respectively. With the UDP profile, the SO state packet sizes were 12 and 26 bytes, respectively, meaning 78.9% and 63.3% savings. With IPv6-only profile, the SO state packet sizes were 18 and 32 bytes, respectively, meaning 68.4% and 54.9% savings. Compared to the IPv4 packet sizes, IPv6 results in larger packets in the beginning of the compression, but in the end, the packets can be compressed to about the same sizes (or even smaller) as with IPv4. Thus, the percentile savings in the packet sizes are actually bigger with IPv6 than with IPv4.

So, even if the CoAP header is designed to have only a small overhead, header compression can still make it even smaller. The used CoAP GET message was practically almost as small as a CoAP packet can be and the savings were 77.8% in the header part. In theory, if the CoAP packet does not have any options or token, the minimum size would be 4 bytes. This could be compressed to 2 bytes, so the savings are always at least 50% in the CoAP header part. If the CoAP Message ID field had a certain known change pattern, the compression could be enhanced still by 2 bytes.

C. Delay measurements

First, we measured the compression and decompression processing delay, using the same traffic as in packet size measurements, but now sent locally over loopback interface at the laptop and the RPi. So the laptop and RPi both ran the CoAP server and the CoAP client, and the communication was only inside the device. Tables I and II represent the processing delays due to the compression and decompression for different ROHC packet types at the laptop and RPi, respectively. As can be seen, the processing delay is bigger in the beginning when the context is created for the flow. After reaching the SO state and the smallest ROHC packet sizes, the delay is much smaller. It is to be noted that the IR state with longer delay is needed only in the beginning of the flow, or if the context needs to be updated for some reason (e.g., because of corrupted context due to the dropped packets). All the first 6 IR packets are reported individually since the packet size varies a little bit due to the feedback information, and also because the very first GET and RESP packets need more processing time due to the initialization of the flow context. There are two UOR-2 packets in each flow before switching to the SO state; the processing time does not differ between them. Also, in the SO state, the delay remains in the same level all the time. Each reported measurement in the tables is an average of at least 5 packets of that type. As can be seen, the total processing delay in the SO state is less than 0.3 ms at the laptop and 0.9 ms at RPi.

Next, we measured the round-trip delay from CoAP request sent from the RPi to the CoAP response from the laptop received at the RPi. As the ROHC library needs the tunnel application between the communicating devices, there is some extra delay caused by the tunnel itself. For that reason, we also measured the round-trip delay without the tunnel and ROHC by using only CoAP client and server to see what is the overhead caused by the tunneling. The measurements were done for both IPv4 and IPv6. The results are based on the average (Avg) of 20 packets and are depicted in Tables III and IV. Minimum (Min) and maximum (Max) values are also reported in the tables. As can be seen, if ROHC is used, the delay is the shortest with the uncompressed profile, i.e., when the packet is not compressed at all. This is because of the processing

TABLE I. ROHC COMPRESSION AND DECOMPRESSION PROCESSING DELAY AT LAPTOP.

ROHC state	*packet type ROHC/CoAP*	*packet size (bytes)*	*comp (ms)*	*decomp (ms)*	*total (ms)*
IR	IR/GET	38	0.322	0.365	0.687
IR	IR/RESP	58	0.215	0.283	0.498
IR	IR/GET	44	0.248	0.265	0.512
IR	IR/RESP	53	0.143	0.217	0.360
IR	IR/GET	38	0.175	0.235	0.409
IR	IR/RESP	53	0.143	0.228	0.371
FO	IR-DYN/GET	15	0.148	0.191	0.339
FO	IR-DYN/RESP	31	0.130	0.167	0.297
FO	UOR-2/GET	6	0.172	0.211	0.384
FO	UOR-2/RESP	22	0.112	0.164	0.276
SO	UO-0/GET	5	0.139	0.145	0.254
SO	UO-0/RESP	21	0.101	0.138	0.216

TABLE II. ROHC COMPRESSION AND DECOMPRESSION PROCESSING DELAY AT RASPBERRY PI.

ROHC state	*packet type ROHC/CoAP*	*packet size (bytes)*	*comp (ms)*	*decomp (ms)*	*total (ms)*
IR	IR/GET	38	1.506	1.618	2.976
IR	IR/RESP	58	1.837	1.009	2.846
IR	IR/GET	44	0.703	0.803	1.627
IR	IR/RESP	53	0.603	0.723	1.326
IR	IR/GET	38	0.578	0.688	1.266
IR	IR/RESP	53	0.511	0.811	1.321
FO	IR-DYN/GET	15	0.507	0.616	1.124
FO	IR-DYN/RESP	31	0.476	0.611	1.087
FO	UOR-2/GET	6	0.545	0.617	1.162
FO	UOR-2/RESP	22	0.428	0.584	1.012
SO	UO-0/GET	5	0.397	0.427	0.824
SO	UO-0/RESP	21	0.370	0.469	0.838

delay needed for compression and decompression. If any compression is used, we can see that it is then worth to compress all the packet headers, since the CoAP profile seem to have a smaller delay than UDP/IP or IP-only profiles. There is no difference whether we use IPv4 or IPv6, the order of the profiles remains the same. Without the tunneling between the laptop and the RPi, and without any ROHC profile, the delay is about 8 ms shorter than with uncompressed profile with IPv4, and about 3 ms with IPv6. So, that is the overhead caused by the tunneling and ROHC without any compression. It is to be noted that tunneling would not be needed if ROHC was implemented directly on both sides of the link; the tunnel is meant only for easy testing of the ROHC library.

TABLE III. ROUND-TRIP DELAY WITH DIFFERENT ROHC PROFILES (IN MS), IPV4.

Profile:	*CoAP*	*UDP*	*IPv4*	*uncompressed*	*no tunnel*
Avg:	20.194	22.368	21.511	18.824	10.627
Min:	6.769	7.480	7.261	5.962	2.209
Max:	37.577	34.496	43.166	34.596	25.325

TABLE IV. ROUND-TRIP DELAY WITH DIFFERENT ROHC PROFILES (IN MS), IPV6.

Profile:	*CoAP*	*UDP*	*IPv6*	*uncompressed*	*no tunnel*
Avg:	20.085	22.366	21.089	16.289	13.365
Min:	8.428	8.070	8.493	5.851	2.389
Max:	46.282	34.950	37.598	39.564	59.472

D. Performance over lossy links

We also studied the effect of smaller packet size to the delay and throughput in lossy (wireless) links. Every real

 ISBN: 978-1-61208-406-0

wireless link has bit errors. For example, for 3G links bit error rates 10^{-3} or even higher are possible [10]. The smaller packet has a smaller probability to have bit errors, so it has a bigger probability to go through the lossy link without errors, meaning less retransmissions, shorter delay, and greater throughput.

The tunnel application provided by the ROHC library supports setting a given bit error rate (BER) to the tunnel. We used it to generate bit errors with a given rate according to a uniform distribution. The tunnel application generates the bit errors on the sender side of the tunnel, and packets with bit errors are not actually transmitted through the tunnel. Retransmissions are then triggered at the CoAP client application because of a missing ACK to the sent message.

With the bit error rate of 10^{-4} (and less than that), the uncompressed profile was still slightly better than compressed profiles, since there were not so many errors. The round-trip delays were almost the same as in Table III, being 20.961 ms for the CoAP profile and 18.503 ms for the uncompressed profile. Tables V - VI represent the round-trip delays with the bit error rate of 10^{-3}. The results for each profile are based on the average (Avg) of 20 sent packets. Also minimum (Min) and maximum (Max) values are reported in the tables. With the uncompressed profile, 5 packets out of 20 were dropped totally due to the exceeding the number of maximum CoAP retransmissions (= 4 times); with other profiles, this did not happen due to the smaller packet size and hence less errors and retransmissions. As can be seen, with high bit error rates such as 10^{-3}, it is useful to use header compression, since it results in shorter delay, less dropped packets, and hence better throughput. Note that the delay increases from milliseconds to seconds! The difference between different profiles is also clear: it is beneficial to compress all the headers to get the packet size as small as possible. There was also a small difference between unidirectional and bidirectional ROHC modes: bidirectional mode (i.e, with feedback) seems to recover faster from corrupted flow context caused by dropped packets and hence it seem to have a slightly shorter delay.

TABLE V. ROUND-TRIP DELAY (S) AND PACKET LOSS WITH DIFFERENT ROHC PROFILES IN LOSSY LINKS, BER=10^{-3}, BIDIRECTIONAL MODE

Profile:	*CoAP*	*UDP*	*IPv4*	*uncompressed*
Avg:	1.040	1.715	2.873	22.342
Min:	0.008	0.007	0.007	0.019
Max:	8.048	8.275	14.324	44.452
#Dropped:	0	0	0	5

TABLE VI. ROUND-TRIP DELAY (S) AND PACKET LOSS WITH DIFFERENT ROHC PROFILES IN LOSSY LINKS, BER=10^{-3}, UNIDIRECTIONAL MODE

Profile:	*CoAP*	*UDP*	*IPv4*	*uncompressed*
Avg:	1.373	2.399	3.250	21.512
Min:	0.007	0.013	0.008	0.015
Max:	8.562	17.549	20.738	44.395
#Dropped:	0	0	0	5

The effect of BER to ROHC has also been studied, e.g., in [10] and [16]. They both show significant decrease in packet loss when the BER is high and ROHC is used. In both of these, RTP profile was used with video and voice applications, respectively.

E. Energy efficiency

In energy efficiency point of view, compression and decompression processing will consume extra energy. On the other hand, as a result of compression, the packet size is reduced, which will save energy during packet transmission.

Figure 4 depicts the current consumption on the client RPi during a single CoAP GET message sending and receiving a response from the server (as in earlier studies). The current was measured for both the CoAP and the uncompressed profiles. With the CoAP profile the packet size was 5 bytes, while with the uncompressed profile it was 38 bytes. The total energy consumption for the whole operation was 156.4 mJ for the CoAP profile and 153.0 mJ for the uncompressed profile. For this calculation, we used a time interval that was started when the current rises above 260 mA at about 14 ms and stopped when the current drops below 260 mA at about 115 ms (uncompressed) or 117 ms (CoAP profile). Thus, the CoAP profile seems to increase the energy consumption about 2.2% compared to the uncompressed profile. However, everything interesting (compression, decompression, transmission, reception) happens during the highest peak in the figure. Thus, if we only observe that part of the figure, we can say that it takes 15.9 ms from the uncompressed profile and 18.2 ms from the CoAP profile. The difference is 2.3 ms and it is because of the compression and decompression processing needed at the client and the server side. This is in line with our delay measurements, where it took little bit less than 1 ms for RPi to compress and decompress a packet. The difference in energy consumption is then roughly $2.3\ \mathrm{ms}{*}340\ \mathrm{mA}{*}5\ \mathrm{V} = 3.9\ \mathrm{mJ}$, which means a 2.5% increase. Note that the extra peak at 77 ms is due to the 802.11 beacon reception. That is why we used 340 mA in the calculation.

The energy savings due to shorter transmission time are insignificant because of the high speed of transmission. In case of uncompressed packet of 38 bytes and compressed one of 5 bytes, the theoretical transmission times with the speed of 54 Mbit/s are $5.6\ \mu$s and $0.7\ \mu$s, respectively, so the saving in transmission time is less than $5\ \mu$s. However, in lower speed networks, the packet size and transmission time may play a bigger role. Studying this remains as one of our future work items.

So, in these tests, the energy needed for compression and decompression processing was bigger than the possible energy saving due to the shorter transmission time. However, the increase in the energy was only about 2.5%. In lossy links with bit errors, this amount of energy can be easily consumed due to the retransmissions because of bit errors. So, let E be the energy needed for one transmission. For X transmissions, the needed energy is $X * E$. If the E increases 2.5% due to the header compression, then by solving X from $X * 1.025E \leq (X + 1) * E$, we get $X \leq 40$, which means that if there is a need for one retransmission at CoAP level in every 40 packets or less, then it turns out to be energy efficient to use the CoAP profile. In the future, we will test this in a real test bed.

In these tests, we observed only one sender and receiver. The smaller packet size due to the header compression saves the bandwidth in the network. This may also turn out to be energy efficient when there are more than one sender and more traffic, since smaller packets may result in less queuing and less packet collisions. Studying this will be one of our future

 ISBN: 978-1-61208-406-0

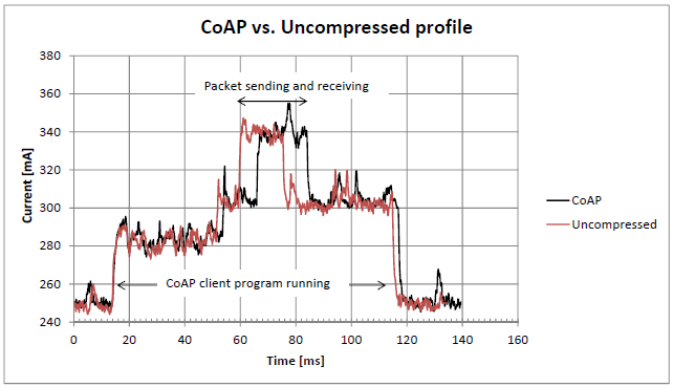


Figure 4. Current as a function of time at the client

work items, too.

VI. Conclusion and Future Work

This paper defined a CoAP profile for ROHC. Its feasibility was studied by measuring the compressed packet size and delay caused by the compression/decompression related processing, by studying its performance in lossy links, and by studying its energy efficiency. With CoAP profile and ROHC, the packet size could be reduced by 91.2% at best. It was found that the delay increased slightly due to the extra processing required for the compression and decompression. The total delay was less than 1 ms in SO state and less than 3 ms in IR state for a single packet compression and decompression at RPi. Due to the smaller packet size, the performance (delay, throughput) in lossy (wireless) links with bit error rates $\geq 10^{-3}$ was much better with the header compression than without. The energy efficiency was also studied. Due to the extra processing needed for compression and decompression, the total energy consumption increased 2.5% in a single CoAP message transaction. So, the savings in the shorter transmission time were not enough to beat the consumption of extra processing. However, in lossy links with bit errors, energy will be saved due to the smaller need for retransmissions. In our test bed, the threshold for energy efficiency was estimated to be around one CoAP level retransmission in every 40 packets.

Our future work items include energy efficiency measurements with different hardware and radio technologies. Especially interesting is to study the energy efficiency with radio technologies with much slower bit rates, such as IEEE 802.15.4-based XBee or ZigBee radios. We are also planning to study the energy efficiency and the performance of header compression in real lossy wireless links instead of using generated bit errors by the tunnel application. We are also thinking about possibilities to integrate ROHC directly to the both ends of the wireless link for avoiding the use of the tunnel application and its overhead. We will also study the effects of smaller packet size to the energy efficiency in a larger test bed with more nodes and traffic. The other ROHC profiles have been standardized, so standardization of the developed CoAP profile may be included in our future work, too.

Acknowledgment

This work was supported by TEKES as part of the Internet of Things program of DIGILE. The comments from the anonymous reviewers are also greatly acknowledged.

References

[1] "Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020," 2013, URL: http://www.gartner.com/newsroom/id/2636073 [accessed: 2015-04-01].

[2] Z. Shelby, K. Hartke, and C. Bormann, "The Constrained Application Protocol (CoAP)," in RFC 7252. IETF, Jun. 2014, URL: https://datatracker.ietf.org/doc/rfc7252/?include_text=1 [accessed: 2015-04-01].

[3] K. Sandlund, G. Pelletier, and L.-E. Jonsson, "The RObust Header Compression (ROHC) Framework," in RFC 5795. IETF, Mar. 2010, URL: https://datatracker.ietf.org/doc/rfc5795/?include_text=1 [accessed: 2015-04-01].

[4] V. Jacobson, "Compressing TCP/IP Headers for Low-Speed Serial Links," in RFC 1144. IETF, Feb. 1990, URL: http://tools.ietf.org/html/rfc1144 [accessed: 2015-04-01].

[5] M. Degermark, B. Nordgren, and S. Pink, "IP Header Compression," in RFC 2507. IETF, Feb. 1999, URL: https://tools.ietf.org/html/rfc2507 [accessed: 2015-04-01].

[6] S. Casner and V. Jacobson, "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links," in RFC 2508. IETF, Feb. 1999, URL: http://tools.ietf.org/html/rfc2508 [accessed: 2015-04-01].

[7] J. Hui and P. Thubert, "Compression Format for IPv6 Datagrams over IEEE 802.15.4-Based Networks," in RFC 6282. IETF, Sep. 2011, URL: https://tools.ietf.org/html/rfc6282 [accessed: 2015-04-01].

[8] G. Montenegro, N. Kushalnagar, J. Hui, and D. Culler, "Transmission of IPv6 Packets over IEEE 802.15.4 Networks," in RFC 4944. IETF, Sep. 2007, URL: https://tools.ietf.org/html/rfc4944 [accessed: 2015-04-01].

[9] C. Bormann, "6LoWPAN Generic Compression of Headers and Header-like Payloads (GHC)," in draft-ietf-6lo-ghc-05. IETF, Sep. 2014, URL: https://tools.ietf.org/html/draft-ietf-6lo-ghc-05 [accessed: 2015-04-01].

[10] A. Couvreur, L.-M. Le Ny, A. Minaburo, G. Rubino, B. Sericola, and L. Toutain, "Performance Analysis of a Header Compression Protocol: The ROHC Unidirectional Mode," Telecommunication Systems, vol. 31, no. 1, Jan 2006, pp. 85–98.

[11] F. Fitzek, S. Rein, P. Seeling, and M. Reisslein, "Robust header compression (rohc) performance for multimedia transmission over 3g/4g wireless networks," Wireless Personal Communications, vol. 32, no. 1, 2005, pp. 23–41. [Online]. Available: http://dx.doi.org/10.1007/s11277-005-7733-2

[12] C. Bormann (editor) and et.al., "RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed," in RFC 3095. IETF, Jul. 2001, URL: http://tools.ietf.org/html/rfc3095 [accessed: 2015-04-01].

[13] "ROHC library," 2015, URL: http://rohc-lib.org/ [accessed: 2015-04-01].

[14] "libcoap: C-Implementation of CoAP," 2015, URL: http://sourceforge.net/projects/libcoap/ [accessed: 2015-04-01].

[15] M. Tömösközi, P. Seeling, and F. Fitzek, "Performance Evaluation and Comparison of RObust Header Compression (ROHC) ROHCv1 and ROHCv2 for Multimedia Delivery," in Globecom 2013 Workshop - Control Techniques for Efficient Multimedia Delivery, Dec. 2013.

[16] Effnet and Intel, "Effnet ROHC (Robust Header Compression) Performance on Intel Core Microarchitecture-Based Processors," in Technology Evaluation White Paper, 2007, URL: http://www.effnet.com/pdf/uk/19350_EFFNET_Final.pdf [accessed: 2015-04-01].

 ISBN: 978-1-61208-406-0

On Investigating the Benefits of TTCN-3-Based Testing in the Context of IEC 61850

Georg Panholzer, Christof Brandauer
Advanced Networking Center
Salzburg Research Forschungsgesellschaft mbH
Salzburg, Austria
email: {firstname.lastname}@salzburgresearch.at

Stephan Pietsch
Testing Technologies IST GmbH
Berlin, Germany
email: pietsch@testingtech.com

Jürgen Resch
Ing. Punzenberger COPA-DATA GmbH
Salzburg, Austria
email: JuergenR@copadata.com

***Abstract*—This paper is concerned with approaches to testing in the context of the International Electrotechnical Commission (IEC) standard IEC 61850, which is gaining momentum in general power utility automation tasks. We outline the current state-of-the-art in IEC 61850 testing and argue that an approach based on the Testing and Test Control Notation (TTCN-3), as has been used successfully in other industries, would provide several advantages. The test specification and execution language TTCN-3 is briefly introduced and potential benefits are discussed. We describe how we used TTCN-3 to implement conformance tests for an IEC 61850 client and run those tests against a certified server. Finally, we present a performance test case for synchrophasor measurements on the basis of distributed TTCN-3.**

Keywords–IEC 61850; TTCN-3; conformance testing; performance testing.

I. INTRODUCTION

The world's energy systems are currently undergoing radical changes that affect individuals, as well as the society as a whole. The change is provoked by a multitude of factors like the limited availability of conventional energy sources, increased demand, increased energy costs, new legal regulations, and the renunciation of energy monopolies.

At the European level the '20-20-20' goals are strived to be reached until 2020: saving 20% of the EU's primary energy consumption, a binding target of 20% reduction of greenhouse gas emissions and 20% renewable energies. The increased integration of renewable energies poses new challenges. Until recently, electricity networks have been "one-way" streets, due to the unidirectional flow of energy from production via transmission and distribution to customers. Through the decentralized integration of renewable energy sources, volatile in their nature, bidirectional energy flows become a reality and the task of controlling these flows becomes much more challenging.

To cope with these challenges, the concept of a next generation electric power system, a so called "smart grid", has emerged. It is characterized by the heavy use of information and communication technologies (ICT), e.g., digital processing and communications, so that data flows and information management become an integral part of the future grid. Indeed, the "intelligence" of the smart grid as implemented in sophisticated energy applications is based on the (near) real-time exchange of measurement and control data amongst a large number of devices located throughout the whole grid [1].

The inter-dependent processes of data acquisition, communication and (often automated) control of the smart grid are highly complex and stringent requirements in terms of correctness, standard conformance, interoperability, performance, and security are posed on the components involved. The topic of testing plays a major role in meeting these requirements.

Our studies on the current state of testing are focused on the multi-part standard IEC 61850. Its development started in 1994 with the goal of developing a comprehensive world-wide standard for the design and operation of substation automation. The main requirements were to advance beyond a growing set of proprietary, incompatible and non-comprehensive approaches to communication solutions in substation automation and to define a global standard that facilitates interoperability and integration. While the initial focus was on the electric substation, the standard has been extended in several directions (substation to substation, substation to control center, hydroelectric power plants, distributed energy resources, wind turbines, etc.) and is nowadays employed more broadly for general power utility automation tasks.

This paper is divided in the following sections: Section II provides a short overview over the multi-part standard IEC 61850, Section III describes the International Organization for Standardization (ISO)/International Telecommunication Union (ITU)/European Telecommunications Standards Institute (ETSI) approach to testing and briefly introduces TTCN-3. In Section IV, we describe how we used TTCN-3 for conformance testing of a IEC 61850 client and for performance testing of a IEC 61850-90-5 based synchrophasor transmission system. In the last section, we present our conclusion and plans for future work.

II. IEC 61850

IEC 61850 must not be seen as yet another communication protocol (like, e.g., many fieldbus protocols that have been invented over the course of time). In fact, IEC 61850 communications are based on existing and matured networking technologies (Ethernet, TCP/IP). A major strength lies in the information model and the device and vendor agnostic configuration and description language that are built upon it.

The information model is an object oriented model that contains a rich set of hierarchical classes which are used to describe physical devices, their functions, services, status information, measured values etc. in a standardized way (IEC 61850-7-3 / 61850-7-4). The information model enables the exchange of semantically well-defined information using common naming conventions. The way the information exchange takes place is abstractly defined in the form of an Abstract

 ISBN: 978-1-61208-406-0

TABLE I. TEST PROCEDURE 'CLIENT ASSOCIATION 1'.

cAss1	Associate and force a client to release a TPAA
Expected result:	
	1. SUT accepts Associate.response+ from server 2. SUT returns to "state" where it is able to start a new TPAA with the same server
Test description:	
	1. Set-up a TPAA with one server 2. Force SUT to release or abort TPAA 3. Repeat step 1 and 2, 10 times server

Communication Service Interface (ACSI, IEC 61850-7-2). The ACSI defines various types of communication services (e.g., directory services, read/write data (datasets), activate group settings, transmission of reports, file transfer, etc.) and furthermore specifies which of these services can be used with the specific elements of the information model. As an example, the `getServerDirectory` ACSI service can (naturally) only be used in combination with the `Server` object of the information model.

The definition of how the abstract services make use of communication protocols to actually transmit data on the network is defined in the so called Specific Communication Service Mappings (SCSM). By following this two-tiered approach the IEC 61850 communication services are decoupled from their concrete networking implementation. This affords a long-term stable interface towards applications without being locked into a specific networking technology. The first standardized mapping for client-server communication is based on the Manufacturing Message Specification (MMS) protocol [2][3], which makes use of OSI protocols on top of TCP/IP.

A. Testing in IEC 61850

Part 10 of the IEC 61850 standard *Conformance testing methodology and framework* is dedicated to conformance and performance testing. A conformance test has to include documentation and version control (IEC 61850-4), configuration (IEC 61850-6), data model (IEC 61850-7-3 and -7-4) and mapping of ACSI models and services (IEC 61850-7-2). Moreover, the testing "ecosystem" is explained including how to certify a tester. Part 10 identifies the areas to be tested (e.g., association, reporting, etc.) and introduces 167 server test cases for the ACSI mapping, all of which are mandatory if supported by the tested system. The positive and negative test cases are briefly described. Based on this framework, the testing subgroup of the Utility Communication Architecture International user group (UCAIug), a *"not-for-profit corporation of utility users and supplier companies"* [4] elaborated more detailed test procedures for servers as well as clients, performance of fast event distribution with Generic Object Oriented Substation Events (GOOSE) and extended server reports.

As a common denominator, all these test cases are described in prose using the table template as required by IEC 61850-10. As an example, Table I shows the description of a simple client association test case.

Such brief prose descriptions naturally leave a lot of room for interpretation and many aspects of the test, some of which have certainly been in mind by the test developers, are not present in the test cases. As an example, it is not specified how to verify that the association has indeed been established. Where are the Points of Observation and Control (POC)? What exactly are the criteria against which the expected Response+ reply has to be matched? Which protocol fields are required? Do some of them need to contain specific values? Which are irrelevant for the given test? It is obvious that the test cases, while still remaining on the abstract level, could be significantly enhanced by making use of a (more) formal description. We have thus studied other approaches to conformance testing with a special focus on communication protocols.

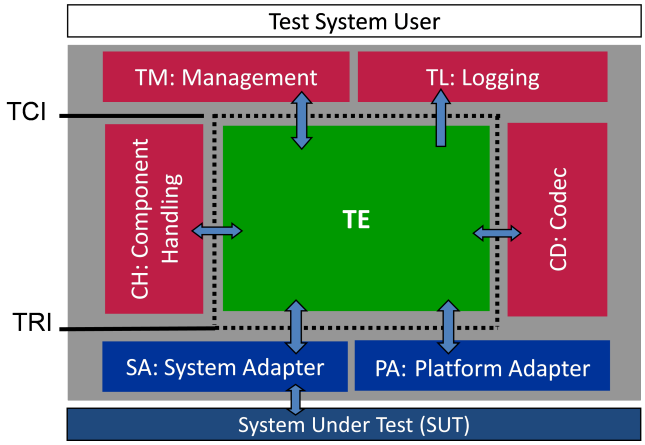


Figure 1. The TTCN-3 test system architecture.

III. THE ISO/ITU/ETSI APPROACH TO (CONFORMANCE) TESTING

Organizations like the ISO, ITU or ETSI are concerned with ensuring that different implementations of a specification conform to the recommendation or standard they are based upon. To this end they have developed the Conformance testing methodology and framework (CTMF) which is published as the ISO/IEC 9646 [5] and ITU-T X.290 to X.296 series of standards.

The CTMF is the standard method for testing OSI communication protocols and it is thus thematically close to IEC 61850 which also partly relies on OSI communication protocols.

A. Testing and Test Control Notation

Part 3 [6] of the ISO 9646 standard series introduces the tree and tabular notation (TTCN) for a formal description of abstract test cases. It is actively developed by ETSI and ITU-T. The current version is TTCN-3 [7] (now called Testing and Test Control Notation).

TTCN-3 has been successfully employed over a decade for major industrial testing efforts in the context of IPv6, SIP, WiMAX, LTE, TETRA, AUTOSAR, and others [8][9]. Initially, TTCN(-3) has been used exclusively to test communication protocols but it has since become a universal testing language that has established itself also in the automotive and medical domain.

TTCN-3 is a formal language which is specifically tailored to testing. As such it provides constructs that are not found

 ISBN: 978-1-61208-406-0

in general purpose programming languages. An example are language elements for data matching that are commonly required for comparing incoming data against test expectations. The language provides an extensive type system and constructs for message-based and procedure-based communication. The concurrent execution of test components is a natural fit to the language. As the language is standardized and has a well-defined semantic for each and every element, a tool vendor lock-in situation is prevented. TTCN-3 is not "just" a language but instead provides a complete test system architecture with standardized interfaces among its components as depicted in Figure 1. There are TTCN-3 extensions related to, e.g., performance and real-time testing, as well as interfaces with continuous signals.

The abstract TTCN-3 test cases are compiled into executable code, which, in combination with a TTCN-3 runtime system, form the core Test Executable (TE). Codecs (CD) provide for the transformation between TTCN-3 data types and the native data format of the System Under Test (SUT). The Test Runtime Interface (TRI) [10] provides functions for communicating with the SUT via the System Adapter (SA) and the platform that hosts the test system via the Platform Adapter (PA). Implementation of the SA and PA components are needed to obtain an executable test system for a concrete SUT.

B. Potential benefits of TTCN-3-based testing in IEC 61850

From a technical point of view, we argue that there are several significant benefits in providing abstract test cases in TTCN-3 compared to prose descriptions. The standardized TTCN-3 language with its well-defined semantic helps eliminate the ambiguities of English (or any other language for that matter) and, consequently, a lot of room for interpretation. It allows for precise specifications of the expected static and dynamic behavior and provides a full-featured so called template system that enables brief and concise data matching statements. As TTCN-3 can make use of ASN.1 and XML Schema types, as well as IDL-based interface definitions, the type specification of standards can often be used directly in (or at least imported into) the TTCN-3 environment. In IEC 61850, the MMS standard – the currently used SCSM mapping for client-server communication – makes use of message types defined in ASN.1.

If abstract TTCN-3 tests were provided to the community, they would serve as a common, reusable starting point for any testing activities. Still, it would not restrict testers with respect to the choice of testing tools and/or implementation languages, respectively.

From a non-technical point of view, we argue that an open standard of a formal test description and execution language is preferable to prose test specifications and their derived proprietary and closed test implementations. Particularly for critical infrastructure like energy networks a review of test results is indispensable. In order to verify and reproduce those test results, it is necessary to make the test methodologies and surrounding conditions publicly available. Standardized TTCN-3-based test specifications readily fulfill a lot of these requirements and thus can significantly improve trust in the test results.

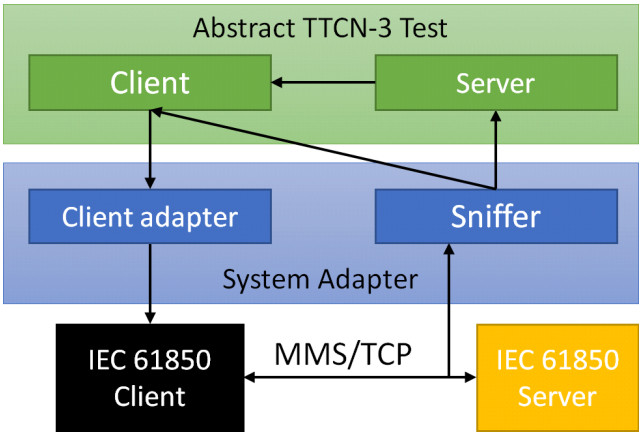


Figure 2. Setup for the IEC 61850 client conformance tests.

IV. IEC 61850 TESTING WITH TTCN-3

Interestingly, TTCN-3 has so far been seldom used in the context of energy networks [11][12]. As a consequence of the arguments listed above, the authors of this report are investigating the applicability of TTCN-3 for testing components and systems based on IEC 61850.

A. IEC 61850 client conformance tests

To start off, we implemented TTCN-3 test cases for a subset of the IEC 61850 client tests as defined by the UCAIug [13]. All tests were run against a certified server. The client has an API to trigger certain actions and query its status. We developed a system adapter that makes this API accessible via TTCN-3. On the server side, a network sniffer captures all network traffic and forwards it to the test runtime. The setup is depicted in Figure 2 and Listing 1 shows the relevant parts of an abstract test case for a client association test in TTCN-3.

Two *components*, `v_Server` and `v_Client` act as TTCN-3 proxies for the server and client respectively. They do not implement any real functionality but simply exchange messages within the runtime environment or with the actual implementations via a system adapter using so called *ports*. In our setup we have three ports, `pt_ObservationPort`, `pt_ControlPort` and `pt_CommPort`.

Both components receive messages from the sniffer via `pt_ObservationPort`. The sniffer's system adapter uses the source IP address to forward the messages to the correct component. Messages with the source IP of the client are sent to the client component and messages with the server's source IP are sent to the server component. The system adapter also maps from the on-wire data formats to the TTCN-3 data types.

The client component can access the client's API via `pt_ControlPort`. The client's system adapter takes the `requestPDU` (e.g., a *MMS initiate-RequestPDU* in case of the client association test case using the MMS SCSM) and calls the correct client API function.

The last port, `pt_CommPort`, connects the server component and the client component within the TTCN-3 environment. The server component uses that port to send the packets it received from the sniffer to the client.

Neither the server component nor the comm port are actually necessary, as the observation port could be used for

 ISBN: 978-1-61208-406-0

```
testcase Ass1_Associate() runs on MTC system TSI {
    clientAddress := {ipAddress := "192.168.1.1"}
    serverAddress := {ipAddress := "192.168.1.2"}
    f_configClientServer(clientAddress, serverAddress);

    v_Server.start(f_Replay());
    v_Client.start(f_Associate(requestPDU, responsePDU));
    f_waitAndGuard();

    f_deconfigClientServer();
}

function f_Associate(in 61850Pdu requestPdu, template 61850
     Pdu responsePlusPdu) runs on ClientComponent  {
    pt_ObservationPort.send(Command:startTrace);
    pt_ControlPort.send(requestPdu);

    // wait for the association request packet
    alt {
        [] pt_ObservationPort.receive(requestPdu) {
            setverdict(pass);
        }
        [] pt_ObservationPort.receive(61850Pdu:?) {
            log("Ignoring Message");
            repeat;
        }
    }
    // wait for the association response packet
    alt {
        [] pt_CommPort.receive(responsePdu) {
            setverdict(pass);
        }
        [] pt_CommPort.receive(negativeResponsePdu) {
            setverdict(fail);
        }
        [] pt_CommPort.receive(61850Pdu:?) {
            log("Ignoring Message",);
            repeat;
        }
    }

    pt_ObservationPort.send(Command:stopTrace);
}
```

Listing 1. Section of a TTCN-3 implementation for a client association (cf. Table I).

traffic in both directions. However, separating the traffic has the advantage that it is easier to understand the code during the implementation phase and the logging output of the execution.

The function `f_configClientServer` sets up the client and server and initializes the ports, `f_deconfigClientServer` releases the ports and resets the client and server.

After configuration, the server component runs `f_Replay`, a function that simply receives messages from the observation port and forwards them to the client component via the comm port. The client component runs the function `f_Associate`. It takes two arguments, the `requestPDU` is the association request message that has to be sent by the client and the template `responsePDU` to match the expected response. Internally, it also uses `negativeResponsePdu`, a globally defined template for negative responses. The function first starts the sniffer by sending the command `startTrace` via the observation port and triggers the association by sending the request message via the control port. Subsequently, it waits until it receives a message that matches the request PDU on the observation port and a message that matches the response PDU template or the negative response PDU on the comm port. Any other 61850 messages are simply ignored.

The test is successful (the verdict is *pass*) if matching messages are received in the correct order; the verdict is *fail* if the second message is a negative response. Otherwise the test will remain in one of the `alt`-statements. Finally the function `f_waitAndGuard` implements a timeout. It waits until either the client component is done with the execution of `f_Associate` or a predefined timeout is reached. If that happens it terminates the test case and the verdict is set to *inconclusive*.

It is very easy to extend this setup for most, if not all, other client tests that do not require any direct interaction between the test system and the IEC 61850 server, i.e., test cases for 'normal' operation. We have in fact successfully implemented several test cases including unbuffered and buffered reporting and direct as well as select-before-operate control models.

Testing the client's behaviour in irregular circumstances requires a non-conformant server. As our server implementation was part of a (conformance certified) SCADA system we were unable to use it for such test cases. The alternative is using a (mock-up) server implementation in TTCN-3 that simply sends (preconfigured) messages to emulate certain behaviour. However, implementing such a server was beyond the scope of our experiments.

B. IEC 61850-90-5 synchrophasor performance tests

As we were convinced that TTCN-3 can be used for all kinds of client and server interoperability tests we extended our investigations towards performance tests for the transmission of synchrophasors according to the technical report TR IEC 61850-90-5 [14].

Synchrophasors are measured by Phasor Measurement Units (PMU). A PMU is a device that measures voltage magnitude and phase angle and current magnitude and phase angle relative to a known time-reference. Additionally, the frequency and frequency drift (rate of change of frequency (ROCOF)) are estimated. Compared to conventional remote terminal units much higher sampling rates of up to 50/60 per second are supported (sometimes even up to 120/s). PMUs are a key sensor to establish a modern wide area monitoring system and they enable a multitude of energy applications [15][16].

We selected the use case 'Under voltage load shedding' (Section 5.9.4 of [14]): An intelligent electronic device (IED) receives synchrophasor measurements from multiple PMUs. It detects a voltage collapse by observing an unusual continuous voltage deviation and generates a GOOSE event that instructs a circuit breaker to trip a line. All communication delays must not exceed 100ms and the synchrophasor measurement timing error must be less than 50µs.

Our test case setup can be seen in Figure 3. PMU 1 and the PMU tester are installed in Salzburg while the rest is located in a testbed in Berlin. Obviously the PMU tester is used to generate unusual voltage deviations and the WAN emulator emulates a WAN connection by adding delay, jitter, packet loss and bit errors. Additionally, GPS is used as a very precise time source at each site.

We developed the following test cases:

- Communication delay: For varying emulated network delays (0-100ms) the duration from measuring a phasor until it is processed by the IED is measured. A test passes if the duration is below 100ms.

 ISBN: 978-1-61208-406-0

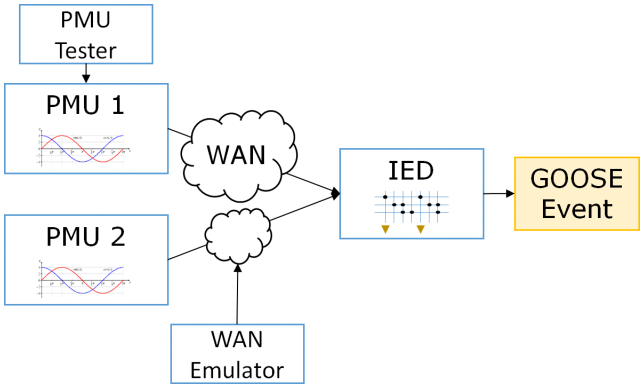


Figure 3. Setup for the IEC 61850 'Under voltage load shedding' performance test case.

- Time to trip: For varying network conditions (delay, loss, bit errors) the time between a significant voltage deviation and the GOOSE event is measured. A test passes if the GOOSE message is sent within a predefined threshold.

These test cases are much more complex than the previously conducted client tests, and so is the test setup. Naturally each device that has to be controlled during the test needs a system adapter. But the biggest difference is clearly that these devices are running at two separate sites. Fortunately TTCN-3 can also run distributed tests.

A distributed TTCN-3 test system consist of one test management (TM) and component handling (CH) instance at the master test component (MTC) and a number of parallel test components (PTC), each with it's own test executable, codec, system and platform adapters. The general structure can be seen in Figure 1. However, TTCN-3 handles the distribution and from an implementers point of view there is no difference between local and remote test components.

V. Conclusion and Future Work

Abstract test cases in TTCN-3 are much more formal than the currently used prose descriptions, while still being easy to understand without extensive training as a TTCN-3 developer. Special language constructs like the `alt`-statement or the powerful templating and data matching features of TTCN-3 offer huge benefits over general purpose programming languages.

By using TTCN-3 we were able to define abstract test cases for conformance testing of an IEC 61850 client, import the data types from an ASN.1 definition and execute these test cases in a TTCN-3 runtime environment to test a concrete client against a certified server. By implementing proper initialization and cleanup functionality in the system adapters a whole test suite can be executed automatically without any human intervention. Furthermore we were able to define performance tests for IEC 61850-90-5 synchrophasor transmissions where the SUT is distributed over two sites.

Our next goals are the definition and implementation of several other test cases for synchrophasor transmission, including Phasor Data Concentrators (PDC) and a Data Archiver (DA). We will also investigate how TTCN-3 can be used to test security mechanisms, e.g., key updates from a key distribution center.

Acknowledgment

This work is partially funded by FFG (Österreichisches Forschungsförderungsgesellschaft) through the ERA-SME project OTITOS.

References

[1] K. C. Budka, J. G. Deshpande, and M. Thottan, Communication Networks for Smart Grids: Making Smart Grid Real, 1st ed., ser. Computer Communications and Networks. Springer, London, England, 2014.

[2] Industrial Automation systems - Manufacturing Message Specification - Part 1: Service Definition, ISO Std. ISO/IEC 9506-1:2003, 2003.

[3] Industrial Automation systems - Manufacturing Message Specification - Part 2: Protocol Specification, ISO Std. ISO/IEC 9506-2:2003, 2003.

[4] "UCA International User Group," [Online]. Available from: http://www.ucaiug.org/aboutUCAIug/default.aspx, 2015.03.10.

[5] Information technology - Open Systems Interconnection - Conformance testing methodology and framework - Part 1: General concepts, ISO Std. ISO/IEC 9646-1:1994, 1994.

[6] Information technology - Open Systems Interconnection - Conformance testing methodology and framework - Part 3: The Tree and Tabular Combined Notation (TTCN), ISO Std. ISO/IEC 9646-3:1994, 1994.

[7] Methods for Testing and Specification (MTS); The Testing and Test Control Notation version 3; Part 1: TTCN-3 Core Language, ETSI Std. ETSI ES 201 873-1 V4.6.1, 2014.

[8] "TTCN-3 Standardized ETSI Test Suites," [Online]. Available from: http://www.ttcn-3.org/index.php/downloads/publicts/publicts-etsi, 2015.03.10.

[9] "TTCN-3 Standardized 3GPP Test Suites," [Online]. Available from: http://www.ttcn-3.org/index.php/downloads/publicts/publicts-3gpp, 2015.03.10.

[10] Methods for Testing and Specification (MTS); The Testing and Test Control Notation version 3; Part 5: TTCN-3 Runtime Interface (TRI), ETSI Std. ETSI ES 201 873-5 V4.6.1, 2014.

[11] S. Schwabe, Q. H. Nguyen, and T. Vassiliou-Gioles, "Synchronized Distributed Testing using TTCN-3 and GPS," ETSI TTCN-3 User Conference 2007, Stockholm, Sweden, 2007.

[12] S. Gröning, C. Lewandowski, J. Schmutzler, and C. Wietfeld, "Interoperability Testing based on TTCN-3 for V2G Communication Interfaces," in Proceedings of the International Conference on Connected Vehicles and Expo (ICCVE), Beijing, China, 2012, pp. 298–303.

[13] M. Flohil and R. Schimmel, "Conformance Test Procedures for Client System with IEC 61850-8-1 interface. Revision 1.1," KEMA Consulting, 2009.

[14] Communication networks and systems for power utility automation – Part 90-5: Use of IEC 61850 to transmit synchrophasor information according to IEEE C37.118, IEC Std. IEC 61 850-90-5, 2012.

[15] M. Gavrilas, "Recent Advances and Applications of Synchronized Phasor Measurements in Power Systems," in Proceedings of the 9th WSEAS/IASME International Conference on Electric Power Systems, High Voltages, Electric Machines, Budapest, Hungary, 2009.

[16] M. Patel et al., "Real-Time Application of Synchrophasors for Improving Reliability," North American Electric Reliability Corporation, Princeton, NJ, USA, 2010.

 ISBN: 978-1-61208-406-0

Potential Impacts of 9-150 kHz Harmonic Emissions on Smart Grid Communications in the United States

Maria Arechavaleta, S. Mark Halpin, Adam Birchfield, Wendy Pittman, W. Eric Griffin, Michael Mitchell
Department of Electrical and Computer Engineering
Auburn University, AL USA
e-mails: mza0036@auburn.edu, halpism@auburn.edu, abb0017@auburn.edu,
wcp0002@auburn.edu, weg0003@auburn.edu, mdm0018@auburn.edu

Abstract— **A key enabling component of the Smart Grid is communications. Of particular interest is power line communications between distributed smart meters and some central relay point. In the very vast majority of cases, smart meters will be located in the low voltage environment and therefore must be designed to operate properly in the presence of disturbance levels bounded by established compatibility levels. Numerous smart meter products are designed to communicate in the 2-150 kHz frequency band range. Communication failures, thought to be due to higher-frequency harmonics, have been reported in the literature and demonstrated in tests conducted in Europe. All of this information is being considered by the International Electrotechnical Commission Technical Committee 77, Sub-Committee 77A, Working Group 8, which is presently tasked with developing compatibility levels for disturbances in this band. However, only limited (if any) work has been done in North America. It is unknown if compatibility levels developed based on the European low-voltage environment are applicable in North America where the environment is much different in several key aspects. Emission evaluation results from product testing in the 120 V three-wire low-voltage environment commonly found in North America are presented in this paper. Results of initial tests conducted to evaluate 9-150 kHz disturbance propagation through North American low-voltage systems are also presented.**

Keywords-smart grid, power line communication, high-frequency harmonics, electromagnetic compatibility.

I. INTRODUCTION

Smart meters are typically connected directly at the low-voltage (LV) point of service for an end user. These meters make the traditional direct measurements of voltages and currents and compute power and energy consumption for billing purposes. Of course, smart meters are also capable of providing many other characterizations of the quantities measured such as harmonic content and voltage excursions. All of these evaluations/calculations are done locally at the meter point. When these meters and their enhanced capabilities are integrated into a coordinated communication and control system, the smart grid is born. Without data sharing and communications, there would be no smart grid.

Smart meter communication approaches can be broadly divided into two categories: wired and wireless. Recognition and tolerance of the background environment and disturbances for wireless systems are covered by numerous applicable standards for wireless communications. Similar standards for wired systems exist for higher frequencies (generally above 150 kHz) and lower frequencies (generally below 2 or 3 kHz), but no consensus presently exists over the range 2-150 kHz [1]. This frequency range includes the bands used by most smart meter manufacturers for communications via the power line (PLC). Lacking existing compatibility levels and limits in the 9-150 kHz range has resulted in smart meter development without regard to standardized background emission levels and problems are beginning to appear [2]. Other problems in addition to PLC failures have also been attributed to harmonics in this range [3]-[5]. All industry stakeholders recognize the need for rapid standardization and numerous activities are underway across Europe. Concerns related to PLC systems are most often the main focus [2][6-8].

Working Group (WG) 8 of the International Electrotechnical Commission (IEC) Technical Committee (TC) 77, Sub-Committee (SC) 77A is specifically charged with reviewing and evaluating the results of these ongoing activities and developing consensus compatibility levels in the 2-150 kHz frequency range [9]. This task is complicated by the fact that numerous end-use products produce emissions in this frequency range, usually due to the common use of various high switching frequency power converter designs required to meet energy efficiency requirements [6][10]. These high-frequency emissions are produced by end-use equipment categories ranging from entertainment (e.g., televisions and displays) to lighting (compact fluorescent and LED ballasts and controls). Other sources of high-frequency emissions include voltage-source inverters used in motor drives and a number of alternative energy (e.g., photovoltaic system) interfaces with the public network. All of these emissions combine in the LV network serving the end-user facility and the cumulative disturbance levels could reach values such that interference with smart meter communications occurs. Alternatively, these high-frequency emissions could essentially circulate between local-area direct-connected devices, resulting in very little impact on the supply system [11]-[13].

To date, all testing, research, and evaluation of high-frequency harmonic product emissions has been focused on products used in European LV networks [11][14]. While these networks are certainly important, smart meter manufacturers stand to benefit from a truly international specification that can be used on a global scale. To accomplish this, information on products used in North American LV networks is required. The testing and measurement approach used in this work to help provide

 ISBN: 978-1-61208-406-0

information on North American LV networks is described in Section II. Emission testing results for some 120 V products in common use in North America are presented in this paper in Section III. Also in Section III, results of tests to assess high-frequency disturbance propagation in 120 V, three-wire LV systems are presented so that postulates can be developed regarding how multiple disturbance-producing (in the 9-150 kHz range) products might summate at the point of service where the smart meter is located.

II. TEST AND MEASUREMENT APPROACH

All measurements were carried out on 120 V equipment and systems using a 100 MHz Tektronix digitizing oscilloscope with built-in signal processing functions including Fourier analysis. Spectral analysis was also conducted off-line using digitized data transferred from the oscilloscope to a local computer. The tests were carried out using a 120 V supply taken directly from the local public network. Equipment was connected to the 120 V public supply source using a standard three-wire cable (#12 AWG solid copper "romex" cable commonly used in North America) rated for continuous operation at 120 V, 15 A. Measurements are taken at two locations: (1) the supply terminals, M_1, and (2) the load equipment connection point, M_2. Voltage signals were the only quantities that were measured; current emissions were not considered in this work. The setup is shown in Figure 1 and mimics a standard 120V service to end-use equipment. Note that the required filters, described later in this paper, are connected at the terminals of measurement points M_1 and M_2 prior to the connection of the measurement equipment.

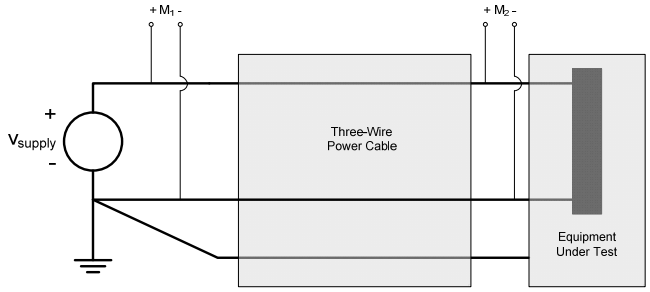


Figure 1. Test and Measurement Setup

As required by electrical codes in the United States, the public supply point is grounded at the point of service only. The equipment under test (EUT) is grounded by connection back to the single ground point at the service; this ground conductor is specifically required to be isolated from the normal current-carrying conductors. Because smart meters will communicate using the power conductors and not the ground, the emission levels produced at the EUT terminals and the supply voltage terminals are measured between the two power conductors rather than from either conductor to the ground.

Measurements are made at the points M_1 and M_2 as shown in Figure 1. The measured emission levels at M_2 with and without the EUT in operation can be used to evaluate the emissions due solely to the operation of the EUT. The measured levels at M_1 with and without the EUT in operation can be used to evaluate the propagation of emissions from the EUT to the supply point. Of course this propagation is a direct function of the frequency response of the power cable and the impedance of the supply system along with any other connected equipment [11]-[13]. While theoretical models can be developed for simple cases, this task can be become difficult and inaccurate for realistically complex systems and is best assessed via direct measurement of input and output characteristics (e.g., at M_2 and M_1).

Typical disturbance levels in the range 9-150 kHz are on the order of a few millivolts (mV) and are commonly expressed in the units "decibel-microvolt" (dBμV) where 20 dBμV= 10*1 μV=10 μV, 40 dBμV=100*1 μV=100 μV, etc. Expressed in this common dBμV unit, typical disturbance levels in the frequency range of interest will be around 80-100 dBμV (10-100 mV). Of course, these disturbance levels will only be encountered at the specific frequencies at which they are produced; much lower levels, typically 40-60 dBμV, will be present over the majority of the frequency range of interest. In order to resolve these small spectral components with sufficient accuracy using a typical oscilloscope/spectrum analyzer, it is necessary to remove the power frequency component from the measured signal before it is processed by the spectrum analyzer. This removal process requires an analog filter of band-stop or high-pass design to eliminate the power frequency signal or pass without attenuation the higher frequencies of interest, respectively.

A high-pass design was chosen for the measurements reported in this work and a custom design was conceived and implemented to avoid any over-dependence on commercial products. In addition, it is difficult to select any particular commercially-available product based on accuracy, performance, or other criteria because no standardized interface coupling exists [14]. The analog filter design is shown schematically in Figure 2.

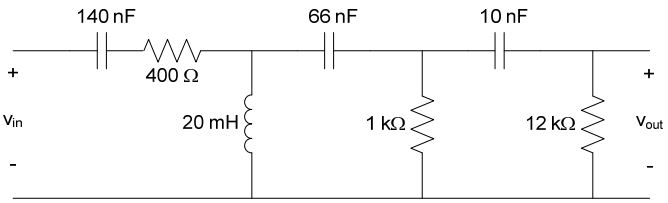


Figure 2. Filter Design and Parameters

Because the filter is a custom design, it is necessary to validate the expected high-pass frequency characteristics and verify that the power frequency component, in this case at 60 Hz, will be sufficiently attenuated. The frequency response characteristic of the filter is shown in Figure 3.

 ISBN: 978-1-61208-406-0

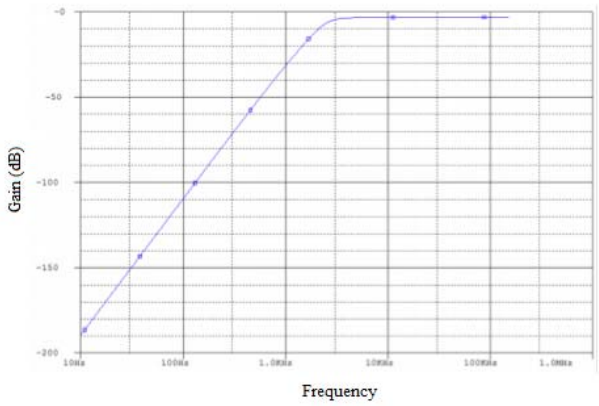


Figure 3. Simulated Frequency Response of Custom Filter

It is clear from Figure 3 that the filter delivers a significant attenuation at the power frequency and corresponding low-frequency harmonics whereas the response is essentially flat with minimal attenuation in the frequency range of interest (9-150 kHz).

III. MEASUREMENT RESULTS

Measurements were conducted at M_1 and M_2 in Figure 1 using the high-pass filter of Figure 2, the digitizing oscilloscope, and off-line computer-based (Matlab) spectral analysis. All measurements were taken in a university office/laboratory environment. Measurements were initially performed at M_1 with no EUT operating in order to establish a baseline condition. To recognize and evaluate expected variations in background disturbance levels over time, the baseline evaluations were conducted over a 72 hr period including a normal workday, multiple nighttime periods, an end-of-week day, and a holiday. These results are shown in Figure 4 averaged over a period defined by a particular date and hour-of-day range as shown in the figure. It is clear from Figure 4 that the variations in background disturbance levels are not overly significant. It is equally clear that there are significant background disturbances in the frequency range 60-70 kHz.

The background disturbance levels in Figure 4 can be used during the emission assessment of various operating EUTs. These longer-time background levels are also useful for evaluating potential measurement errors; erroneous measurements would likely deviate significantly from the established background levels. It is important to note that the particular background levels shown are relevant only to the particular measurement sets of this paper and should not be directly transferred or assumed applicable to another location or time period. An example of this time dependence can be seen in some of the measurement results that follow where "before" and "after" measurements are shown.

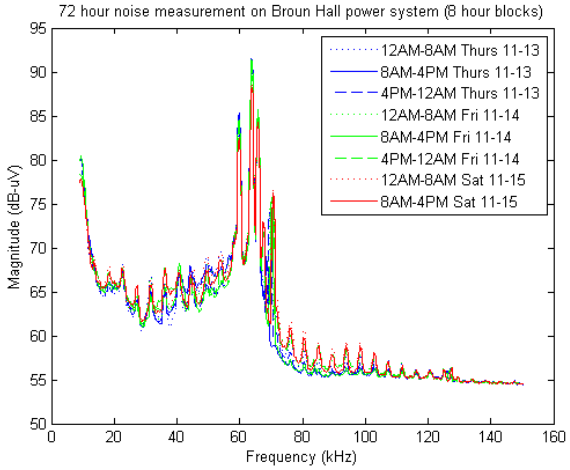
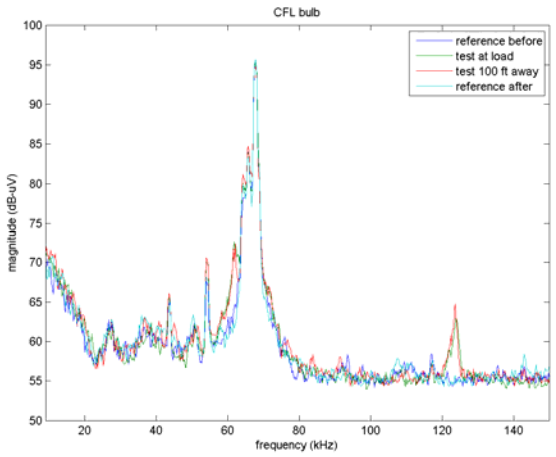


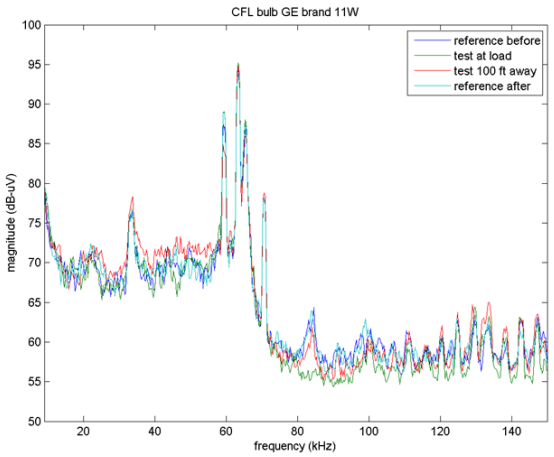
Figure 4. 72 hr Background Disturbance Levels

Two major categories of consumer products were tested in this work: lighting and televisions (displays). Measurements were taken on both ends of the supply cable impedance (approximately 30.5m long power cable as previously described) with and without the EUT in operation. For the cases with the EUT disconnected, measurements were made both before and after the EUT connection and operation so that the reference levels immediately before and after each test could be known and, for validation purposes, compared to the longer-time results of Figure 4 as appropriate.

The results of two compact fluorescent lamp (CFL) tests are shown in Figure 5 (a) and (b). These results clearly show that one of the CFLs produces a noticeable emission around 120 kHz whereas the other tested lamp provides an attenuating affect around 80 kHz at the EUT terminals but not at the supply terminals. From these two tested lamps, it does not appear reasonable to make generalizations.



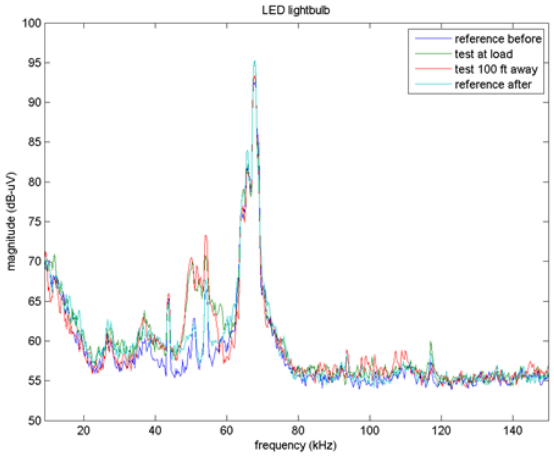
(a)



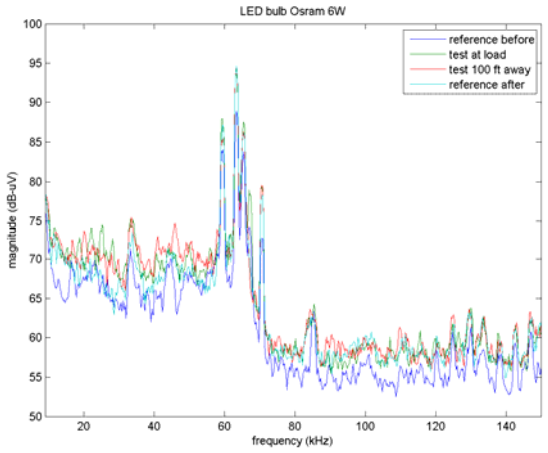
(b)

Figure 5. Emissions from Compact Fluorescent Lamps

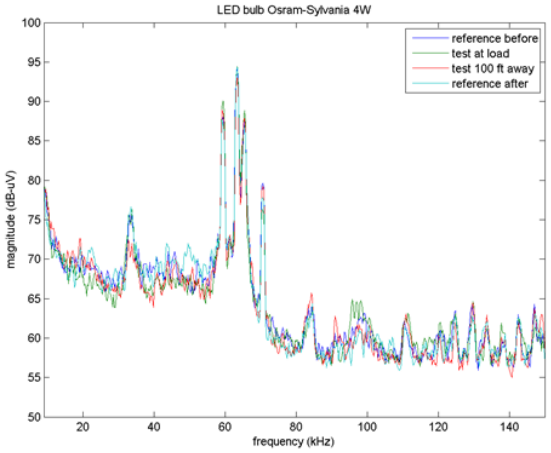
The results of four LED lamp tests are shown in Figure 6 (a)-(d). These results show higher emissions around 50 kHz (a), the effects of an increasing change in background disturbancess (b), a general change with some increases and some decreases (c), and the effects of a decreasing change in background disturbances (d). For all the tested LED lamps, there does not appear to be a significant impact on disturbance levels relative to the background levels at either the source or load terminals.



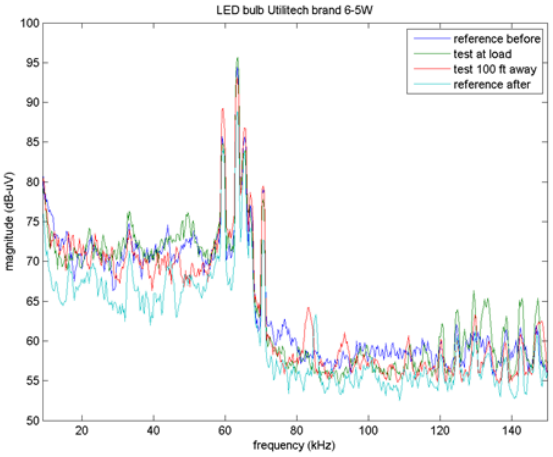
(a)



(b)

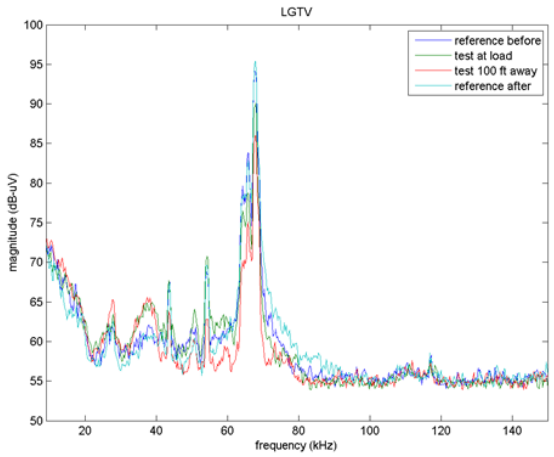


(c)

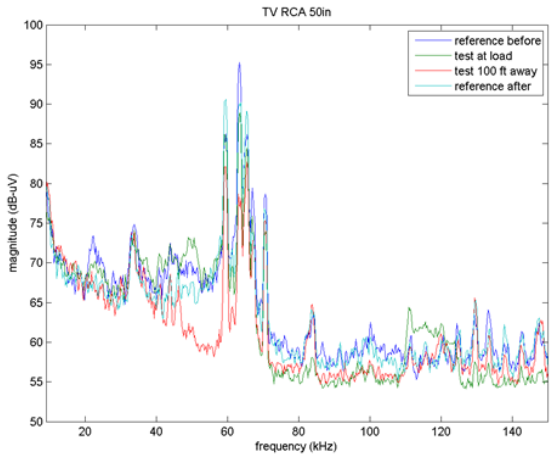


(d)

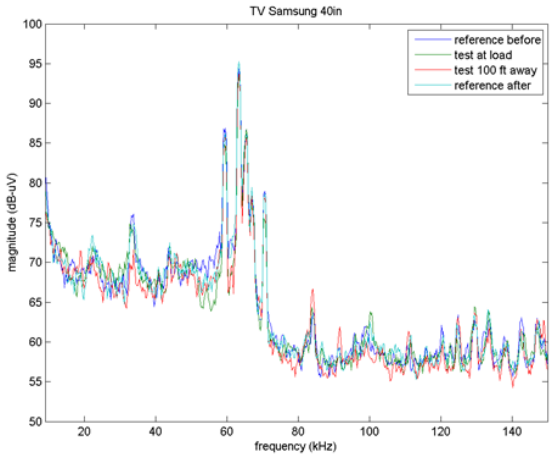
Figure 6. Results of LED Lamp Tests



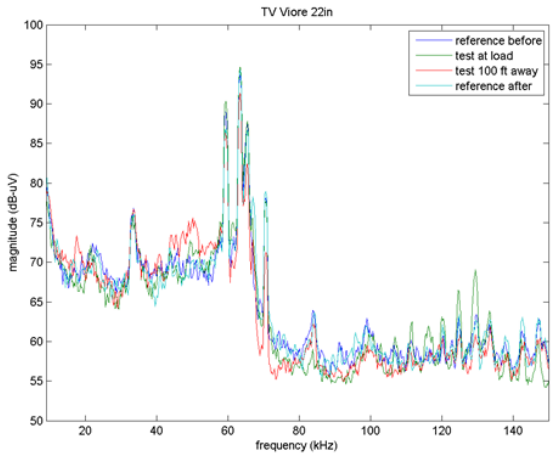
(a)



(b)



(c)



(d)

Figure 7. Results of Television/Monitor Tests

The results of four television/display tests are shown in Figure 7 (a)-(d). Tests (b) and (d) show some amplification and attenuation effects of the power cable, particularly around 110-120 kHz (b) and 120-130 kHz (d). The other two tests do not appear to have any single dominant features but it is clear that the disturbance levels change with and without the EUT in operation in all cases.

IV. CONCLUSIONS

Emission testing of 120 V consumer products that are in common use in the United States has shown that typical products do in fact present the potential to appreciably change the high-frequency harmonic disturbance levels in locations in close (electrical) proximity to the equipment under test. However, the measurement results reported in this paper are clearly different from one piece of (similar) equipment to the next and no obvious generalizations can be made. Furthermore, connecting a public-supply source to the tested equipment using standard power cable conductors has shown that the power cabling can have amplifying or attenuating effects depending on the specific situation.

All of the tests conducted as a part of this work support the general conclusion that typical consumer equipment can be expected to increase the disturbance level by 5-10 dBμV (more in some cases) over relatively narrow ranges of the general band of interest 9-150 kHz. Combined with the long-duration evaluation of the background disturbance levels present in the public supply source, these increases can help to establish a realistic compatibility level for harmonics in the 9-150 kHz band. Given these levels, smart meter manufacturers can offer improved communications hardware that can more effectively communicate using public power system conductors as the channel.

 ISBN: 978-1-61208-406-0

REFERENCES

[1] R. Burkart and J.W. Kolar, "Overview and Comparison of Grid Harmonics and Conducted EMI Standards for LV Converters Connected to the MV Distribution Grid," Proceedings of the 1st Power Electronics South America Conference and Exhibition, September 2012.

[2] P. Pakonen, S. Vehmasvaara, M. Pikkarainen, B.A. Siddiqui, and P. Verho, "Experiences on Narrowband Powerline Communication of Automated Meter Reading Systems in Finland," Proceedings of the Electric Power Quality and Supply Reliability Conference (PQ), June 2012, pp. 1-6.

[3] M. Pikkarainen, S. Vehmasvaara, B.A. Siddiqui, P. Pakonen, and P. Verho, "Interference of touch dimmer lamps due to PLC and other high frequency signals," Electric Power Quality and Supply Reliability Conference (PQ), June 2012, pp. 1-6.

[4] G.A. Franklin, "A Practical Guide to Harmonic Frequency Interference Affecting High-Voltage Power-Line Carrier Coupling Systems," IEEE Transactions on Power Delivery, pp. 630-641, April 2009.

[5] B. Gustavsen, "Study of Transformer Resonant Overvoltages Caused by Cable-Transformer High-Frequency Interaction," IEEE Transactions on Power Delivery, April pp. 770-779, 2010.

[6] S.M. Bashi, "Effects of High Power Electronics Converters on PLC Signals," Journal of Applied Sciences, pp. 1888-1891, January 2006.

[7] S. K. Ronnberg, M.H.J. Bollen, and M. Wahlberg, "Interaction Between Narrowband Power-Line Communication and End-User Equipment," IEEE Transactions on Power Delivery, pp. 2034-2039, August 2011.

[8] E. de Jaeger, H. Renner, and K. Stockman, "Special Report-Session 2: Power Quality and Electromagnetic Compatibility," 22nd International Conference on Electricity Distribution (CIRED), June 2013.

[9] D. Heirman, "EMC Standards Activity," IEEE Electromagnetic Compatibility Magazine, January 2014, pp. 94-96.

[10] J. Luszcz, "High Frequency Harmonics Emission of Modern Power Electronic AC-DC Converters," 8th International Conference on Compatibility and Power Electronics, June 2013, pp. 269-274.

[11] E.O.A Larsson, C.M. Lundmark, and M.H.J. Bollen, "Distortion of Fluorescent Lamps in the Frequency Range 2-150 kHz," 7th International Conference on Harmonics and the Quality of Power, September 2006.

[12] S. Ronnberg, M. Wahlberg, M. Bollen, A. Larsson, and M. Lundmark, "Measurements of Interaction Between Equipment in the Frequency Range 9 to 95 kHz," 20th International Conference on Electricity Distribution (CIRED), June 2009.

[13] D. Frey, J.L. Schanen, S. Quintana, M. Bollen, and C. Conrath, "Study of High Frequency Harmonics Propagation in Industrial Networks," 2012 International Symposium on Electromagnetic Compatibility, September 2012, pp. 1-5.

[14] M. Coenen and A. van Roermund, "Conducted Mains Test Method in the 2-150 kHz Band," 2014 International Symposium on Electromagnetic Compatibility," September 2014, pp. 601-604.

 ISBN: 978-1-61208-406-0