ENERGY 2013

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

ISBN: 978-1-61208-259-2

March 24 - 29, 2013

Lisbon, Portugal

ENERGY 2013 Editors

Steffen Fries, Siemens, Germany

Petre Dini, Concordia University, Canada / China Space Agency Center, China

ENERGY 2013

Foreword

The Third International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies [ENERGY 2013], held between March 24 - 29, 2013 in Lisbon, Portugal, continued the event 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.

We take here the opportunity to warmly thank all the members of the ENERGY 2013 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 who dedicated much of their time and efforts to contribute to ENERGY 2013. 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 are grateful to the members of the ENERGY 2013 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ENERGY 2013 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the fields of smart grids, green communications and IT energy-aware technologies.

We are convinced that the participants found the event useful and communications very open. We also hope the attendees enjoyed the charm of Lisbon, Portugal.

ENERGY 2013 Chairs:
Stefan Mozar, CCM Consulting / CQ University - Sydney International Centre, Australia
Mehrdad (Mark) Ehsani, Texas A&M University - College Station, USA
Petre Dini, Concordia University, Canada / China Space Agency Center, China
Mardavij Roozbehani, Massachusetts Institute of Technology, USA
Dave Cavalcanti, Philips Research North America, USA
Marco Di Girolamo, Hewlett-Packard Company - Cernusco sul Naviglio, Italy
Thomas M. Overman, Boeing Energy Cyber Security, USA
Angelantonio Gnazzo, Telecom Italia - Torino, Italy

Dragan Obradovic, Siemens AG, Germany
Avi Mendelson, Microsoft Corporation, USA
Peter Müller, IBM-Zurich, Switzeland
Giorgio Nunzi, NEC Europe Ltd. - London, UK
Ritwik Majumder, ABB AB / Corporate Research Center - Vasteras, Sweden
Brian P. Gaucher, IBM Research Division - Yorktown Heights, USA
Gargi Bag, ABB Corporate Research, Sweden
Ken Christensen, University of South Florida, USA
Grzegorz Swirszcz, IBM Watson Laboratory, USA
Luciano Bertini, Fluminense Federal University, Brazil
Cathryn Peoples, University of Ulster, UK

ENERGY 2013

Committee

ENERGY Advisory Chairs

Stefan Mozar, CCM Consulting / CQ University - Sydney International Centre, Australia
Mehrdad (Mark) Ehsani, Texas A&M University - College Station, USA
Petre Dini, Concordia University, Canada / China Space Agency Center, China
Mardavij Roozbehani, Massachusetts Institute of Technology, USA

ENERGY Industry Liaison Chairs

Dave Cavalcanti, Philips Research North America, USA
Marco Di Girolamo, Hewlett-Packard Company - Cernusco sul Naviglio, Italy
Thomas M. Overman, Boeing Energy Cyber Security, USA
Angelantonio Gnazzo, Telecom Italia - Torino, Italy
Dragan Obradovic, Siemens AG, Germany
Avi Mendelson, Microsoft Corporation, USA

ENERGY Special Area Chairs on Nano-Grids

Peter Müller, IBM-Zurich, Switzeland

ENERGY Special Area Chairs on Efficient Energy Consumption

Giorgio Nunzi, NEC Europe Ltd. - London, UK

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 Publicity Chairs

Luciano Bertini, Fluminense Federal University, Brazil
Cathryn Peoples, University of Ulster, UK

ENERGY 2013 Technical Program Committee

Amir Abtahi, Florida Atlantic University - Boca Raton, US
Nizar Al-Holou, University of Detroit Mercy, USA
Lachlan Andrew, Swinburne University of Technology – Melbourne, Australia
Luca Ardito, Politecnico di Torino, Italy
Mehdi Bahrami, University of California - Berkeley, USA
Chakib Bekara, Fraunhofer Fokus Institute – Berlin, Germany
Fabio Luigi Bellifemine, Telecomitalia, Italy
Frede Blaabjerg, Aalborg 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
Mohamed Cheriet, ETS/GreenStar - Montreal, Canada
Howard Choe, Combat Systems - Plano, USA
Sangho Choe, The Catholic University of Korea, Korea
Delia Ciullo, INRIA 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, Electricity Authority of Cyprus, Cyprus
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
Harald Gjermundrod, University of Nicosia, Cyprus
Angelantonio Gnazzo, Telecom Italia - Torino, Italy
Manimaran Govindarasu, Iowa State University, USA
Mesut Günes, Freie Universität Berlin, Germany
Rune Gustavsson, KTH, Sweden
Helmut Hlavacs, University of Vienna, Austria
Chun-Hsi Huang, University of Connecticut - Storrs, USA
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

Young Sun Kim, Korea Electrotechnology Research Institute (KERI), Korea
Thierry E. Klein, Bell Labs / Alcatel-Lucent, France
Dejan Kostic, EPFL, Switzerland
Paul J. Kühn, University of Stuttgart, Germany
DongJin Lee, University of Auckland, New Zealand
Eugene Litvinov, ISO New England, USA
Jaime Lloret Mauri, Universidad Politécnica de Valencia, Spain
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
Jean-Marc Menaud, Ecole des Mines de Nantes, France
Avi Mendelson, Microsoft R&D, Israel
George Michailidis, University of Michigan, USA
Nicolas Montavont, Telecom Bretagne, France
Daniel Mossé, University of Pittsburgh, USA
Gero Mühl, Universität Rostock, Germany
Dragan Obradovic, Siemens AG - München, Deutschland
Jacob Østergaard, Technical University of Denmark, Denmark
Cathryn Peoples, University of Ulster, UK
Mirek Piechowski, Meinhardt (Vic) Pty Ltd. - Melbourne, Australia
Philip W. T. Pong, The University of Hong Kong, Hong Kong
Enrique S. Quintana-Orti, Universidad Jaume I, Spain
Shankar Raman, Indian Institute of Technology - Madras, India
Andrew Rice, University of Cambridge, UK
Jan Richling, TU Berlin, Germany
Jan Ringelstein, Fraunhofer Institut für Windenergie und Energiesystemtechnik IWES, Germany
Darren Robinson, University of Nottingham, UK
Sebnem Rusitschka, Siemens AG - München, Germany
Eliot Salant, IBM Haifa Research Labs / Haifa University, Israel
Dave Saraansh, University of Bristol / Toshiba Research Europe Limited, UK
Dirk Uwe Sauer, ISEA / RWTH - Aachen University, Germany
Andrey V. Savkin, The University of New South Wales - Sydney, Australia
Harald Schrom, Technische Universitaet Braunschweig , Germany
Sandra Sendra, Universidad Politécnica de Valencia, Spain
Gerard Smit, University Of Twente - Enschede, The Netherlands
Fernando Solano, Warsaw University of Technology, Poland
Pavel Somavat, CCS Haryana Agricultural University, India
Grzegorz Swirszcz, IBM Watson Laboratory, USA
Nick A. Tahamtan, Vienna University of Technology, Austria
Zhibin Tan, East Tennessee State University, USA
Bernard Tourancheau , University Joseph Fourier of Grenoble, France
Lefteri H. Tsoukalas, Purdue University - West Lafayette, USA
Jean-Philippe Vasseur, Cisco Systems, Inc., France
Stéphane Vialle, SUPELEC, France
Matthias Vodel, Technische Universitaet Chemnitz, Germany
Le Yi Wang, Wayne State University, USA

Guanghua Yang, The University of Hong Kong, Hong Kong
Francis Zavoda, Hydro-Quebec, Canada
Jianqing Zhang, Intel, USA
Albert Zomaya, University of Sydney, Australia

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

Rafael Cividanes, Nelson Uto, Bruno Botelho, Sergio Ribeiro, Christiane Cuculo, Danilo Suiama, and Heron Fontana

Simulation Tool For Energy Consumption and Production

The development of a simulation tool for measuring the impact of a smart grid on a building

Michael Nysteen, Henrik Mynderup, Bjarne Poulsen, Chresten Træholt

Department of Informatics and Mathematical Modelling & Department of Electrical Engineering
Technical University of Denmark
Kongens Lyngby, Denmark
{s062413, s062373}@student.dtu.dk, bjp@imm.dtu.dk, ctr@elektro.dtu.dk

Abstract—In order to promote adoption of smart grid with the general public it is necessary to be able to visualize the benefits of a smart home. Software tools that model the effects can help significantly with this. However, only little work has been done in the area of simulating and visualizing the energy consumption in smart homes. This paper presents a prototype simulation tool that allows graphical modeling of a home. Based on the modeled homes the user is able to simulate the energy consumptions and compare scenarios. The simulations are based on dynamic weather and energy price data as well as a controller unit of the user's choice. The results of the simulations can be compared using a dynamic reporting window that allows the user to create custom charts of the data. The application has been designed such that it can easily be extended with additional controller units, price and weather data as well as appliances and other electrical components used in the modeled homes.

Keywords-Smart grid; simulation; modeling; controller; smart home; software; graphical user interface; end-user application;

I. INTRODUCTION

According to a Danish workgroup [1], the government needs to increase the awareness of the smart grid technology with the consumers to smooth the transition to smart grid. Following the above mentioned recommendation a tool that helps people realizing the effects of smart grid integration in private homes is needed.

The power grid today is highly affected by variations in the use of power during a day. The energy use during the night is very low, but the use of energy increases significantly during a couple of hours in the morning, afternoon and evening (called peak hours) when people come home from work. Today the demand in these peak hours are met by powering up so called *peaker plants* that can be started up rapidly when more power is required [2]. These typically run on oil or natural gas and are very expensive to operate. Furthermore, they produce a significant amount of pollution [3]. In some countries the consumption in the peak hours is reaching, or even exceeding, the full capacity of the grid and results in frequent power outages [4]. A way of shaving these peaks in consumption would be to give the consumer a motivation for moving her energy consumption to other periods of the day. Such a motivation is likely to be of

economic character, e.g. by making energy cheaper outside of peak hours.

In order to maximize the economic gain of dynamic pricing, energy should be consumed at the cheapest possible time. Several systems [5, 6], so called *controller units*, have been developed that control the energy consumption in the home by automatically increasing consumption when the prices are low and decreasing consumption during peak hours. This includes the Control4 EMS 100 controller system [5] and Tendril Connect [6] among others. Both of these systems can control any energy consumption in the home such as heating, lights and appliances as well as charging electric vehicles.

A. Network Architecture of the Smart Grid

The smart grid is based on a tiered architecture of a number of networks that connects all the parts of the smart grid, from power plants to the private home. The communication goes both ways so that each partner in the smart grid can communicate with each other, in theory. The smart grid can be divided into three main networks; Wide Area Network (WAN), Neighborhood Area Network (NAN) and Home Area Network (HAN).

The solution described in this paper only focuses on the HAN, as this network is responsible for the communication between the appliances within the house. There are currently two overall approaches for constructing a Home Area Network without looking at which technologies to use. Common for both of the approaches is that a smart meter is connected to the HAN along with all the appliances.
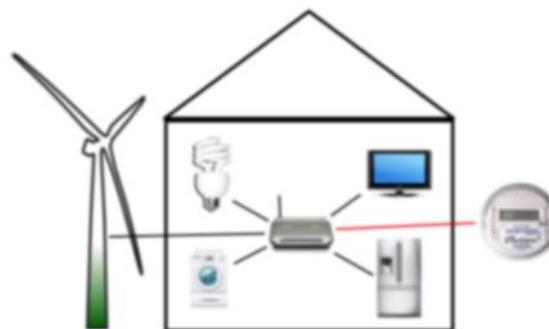

Figure 1.   An example of a centralized HAN.

- A Decentralized HAN solution is where each appliance on the network will be responsible for their own scheduling. Each appliance is communicating with the smart meter directly using the chosen standard and they contain embedded controller software that based on the current power price decides how the appliance should act. This price is the current market price for a kWh which the smart meter receives from the distribution network.
- A Centralized HAN network is a network where all the appliances are connected to a router which may be connected to the Internet. The smart meter and a control unit are also connected to this network. In this scenario each of the appliances, which are connected to the network, only provides a software interface that can be used to control the appliance. This interface can then be accessed by the controller unit on the HAN. The controller unit will control the appliance based on the power price information that the controller unit gets from the smart meter.

A Home Area Network in itself does not provide any intelligence to the home. This intelligence comes from the controller units connected to it (or partly by the individual appliances). By making the consumer aware of the dynamic prices she can save a significant amount of energy. A report from California Energy Commission [7] states that the average household can save 34.5% through technology used to inform people of the varying prices and each units automatically adjusting their consumption according to the price. In order to maximize the benefit from this information, however, one needs something that can control energy consumption 24 hours a day. This is the so-called *controller unit*. A controller unit is placed in a centralized or decentralized HAN (here it is actually not a physical unit, but a unit distributed in all devices). The centralized approach can be divided into two types of controller units: A physical unit installed in the home that controls each unit based on the customer's preferences or an external service communicating with the household over e.g. the Internet.

B. State-of-the-Art Simulation

A number of simulation tools that examine the energy consumption in homes have been developed. Many tools such as GridLab-D [8] and EnergyPlus [9] examine the energy consumption with focus on heat generation and thermal load of a building. Very few tools examine the home in the context of a smart grid where the price of energy varies throughout the day. One of the few tools that were found is PowerMatcher Simulation Tool [10]. It is, however, only intended to demonstrate the PowerMatcher technology which is a home controller system and the tool is therefore limited in features. It does e.g. not provide any means of modeling accurate energy consumption patterns based on many independent appliances, but simply allows the user to create a static *energy demand* [11].

None of the existing simulation tools consider the effects of using different controller units for controlling the home's energy consumption and none of them provide a modeling environment that allows the user to create an accurate representation of her home to be used in the simulations. This means that the existing tools cannot properly show a consumer how turning her home into a smart home will be of value to her. These points are addressed in this work.

The remainder of this paper is organized as follows. Section 2 describes the identified requirements for the simulation tool and the targeted audience. Section 3 presents the core features of the prototype simulation tool and shows the Graphical User Interface (GUI). Finally, conclusion and future work are presented in sections 4 and 5. For an extensive explanation of the developed tool see the report that this paper is based on [12].

II. SOLUTION REQUIREMENTS

The simulation tool that was developed in this work can provide reasonable simulations of how a private household is (financially) affected by a smart grid implementation. The simulation tool focuses on allowing the user to create a virtual household and add electrical appliances (washing machines, dish washers, refrigerators etc.) as well as local energy sources (such as wind turbines, solar cells etc.) to the household using a user friendly drag-and-drop approach. The focus of the simulations is to see how different scenarios affect the power bill when the energy consumption is controlled using various controller units.

In order to simulate this, the application will use environmental variables which includes the weather (affecting power generated by wind turbines, solar cells etc.) and the prices of buying energy from external energy sources at any given time. Most of these variables will be importable from real sources or manually configurable. The application supports the creation of multiple home scenarios and lets the user compare these, giving the user the possibility of evaluating the value of potential appliances that might be added to the house in the future. The scenarios can also be executed without smart-scheduling, thereby allowing the user to compare the scenarios with and without smart-scheduling enabled to see the effect of using controller units to schedule the energy consumption.

A. Target Audience

Three groups of people have been identified as potential users of the application. These groups are *the consumer*, *energy advisers/salesmen* and *educators/educates*. Each of these groups has been analyzed and their preferences in four categories have been compared: *Focus, Interaction with Tool, Desired Output* and *Generic vs. Real Data*. *Focus* refers to what the user wishes to achieve when using the application. *Interaction with Application* indicates what is most important to the user in the way she interacts with the application. *Desired Output* is the kind of output that supports the focus of the user in the best way.
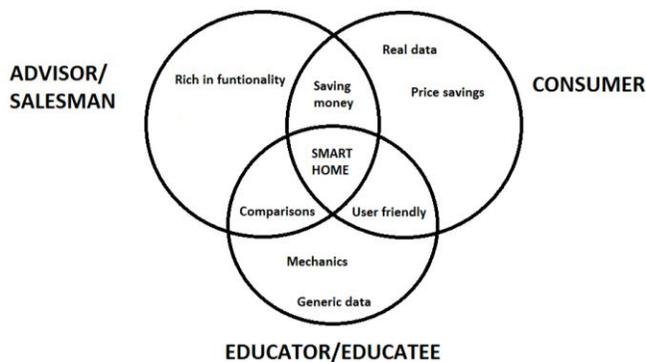
Figure 2.   Shared interests between target audiences.

Finally, *Generic vs. Real Data* refers to whether or not the user prefers realistic data about weather, energy prices, and more importantly manufacturer specific appliances or generic versions of the appliances and customizable weather and prices.

The three user groups have a number of shared interests that they find important; Money savings, user friendliness and scenario comparisons. From the three main focus areas of the program it seems realistic to develop an application that can be used by all three user groups. Therefore the target audience is a combination of the three user groups.

B. Electric Components in the Home

In order to give the user the best possible results, the program needs to be as realistic as possible when modeling scenarios that they want simulated. Therefore the program has to contain realistic and user-customized content.

In the program a home can consist of a lot of components such as lights, refrigerators, freezers, computers, TVs, washing machines etc. Some of the components, such as washing machines, can be scheduled based on dynamic prices while others, such as TVs, computers and lights, cannot. While you can still implement energy saving initiatives for non-schedulable components such as automatically turning off the components when you're not home, the work described in this paper focuses on how components can be scheduled based on dynamic prices during a day. Therefore components will be placed in two groups; schedulable and non-schedulable components. Schedulable components will, as the name indicates, be scheduled by the simulation based on the dynamic prices, whereas the non-schedulable components will simply be run during a user-defined period each day regardless of the electricity price.

III.   SOLUTION PRESENTATION

When the developed application is started the user is met by the main window. This window is shown in Fig. 3 and the four main regions of the program are marked.
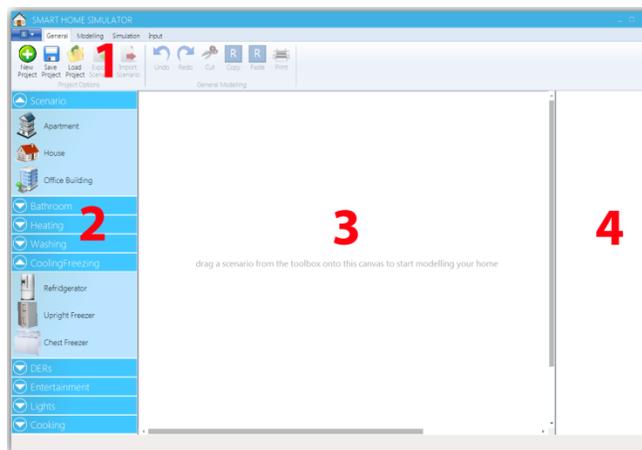
Figure 3.   The main window of the application with each of the primary regions marked.

1. **Ribbon navigation**: This region provides the user with various options to manage the project including created models and simulations.
2. **Toolbox**: This region contains all the elements that can be used when modeling a home.
3. **Canvas**: The elements from the toolbox can be dragged onto this region to create the desired models of houses.
4. **Summary**: This region provides a brief overview of the yearly power consumption distribution in the modeled homes.
A. **Input**: The application provides data about power prices and weather to be used during the simulations.
B. **Simulation**: The simulation component of the program simulates the consumption of the modeled homes and provides the results to the user. This is not shown in Fig. 3.
C. **Reporting Window:** The reporting window allows the user to analyze the results of the simulations provided by the simulation component. This can be seen in Fig. 8.

The user can now start adding content to the Canvas (3) by dragging elements from the Toolbox (2) onto the Canvas or he can go through the options within the ribbon navigation (1). Initially the text on the Canvas informs the user that he can drag elements onto the canvas to start modeling. If the users choose to add content to the Canvas (3) the text on the canvas will change to inform the user of the next step, i.e. adding components. Once the user adds components to the scenario the Summary (4) will update. The summary region will contain a sum up of the content currently on the canvas, with information about their estimated yearly consumption and their overall part of the consumption of the house (Scenario).

A. Modelling - Toolbox

In the left side of the program's main window is a toolbox containing all the elements that can be dragged onto the centered canvas. The toolbox is divided into logical categories where each category contains the elements related to it. The *Cooking* category could e.g. contain an oven and a stove as well as other cooking appliances. Clicking the expander arrow of a category unfolds it and clicking it again collapses the category.

It is possible for the user to create new, delete or edit existing toolbox elements. This allows them to create exactly the elements that they use within their home and reuse them in another project. When the user wants to edit or delete elements they can do this from two places; from the modeling tab in the Ribbon and from a context menu that appear when right clicking on an element. Creating new elements can only be done from the modeling tab in the Ribbon.

When choosing to delete or edit an element through the Ribbon the users will be prompted with a simple window where they can choose the element to edit or delete. When the user chooses to delete the element from either the window above or from the context menu, the element is deleted and will be removed from the toolbox.

When choosing to edit an element the user is prompted with a new window which contains all the possible properties which can be edited on the selected element. Before saving the changes the input values are verified in order to ensure that they are of the correct format. It could be that numbers do not contain letter and so on.

In Fig. 4 a screenshot of the edit window can be seen, containing fields and values related to a Light Bulb element. As seen in the figure two buttons are placed in the bottom of the window. These buttons allow the user to cancel changes and close the window or save changes and close the window.

The window for creating a new element is very similar to the edit window. The main difference is that when creating a new element you need to choose which type of element you want to create before the related fields will occur. When new elements are added to the toolbox it is automatically refreshed to include the new element.

B. Modeling - Canvas

The canvas is a crucial part of the modeling in the program. It is responsible for containing and presenting the result of the users modeling actions. Graphically the canvas is only a white background, but the elements which are shown on it are represented by an icon and some small buttons used for control and information.

The estimated yearly consumption of the component is shown in the top right corner of its icon (see Fig. 5). The color indicates how high the consumption is; the darker red, the higher consumption. The component can be edited by right clicking it or through the controls that appear when hovering the mouse over it.

When modeling a home the components added to the home are connected via colored lines. The colors of these lines represent the category the component belongs to as seen in the toolbox.
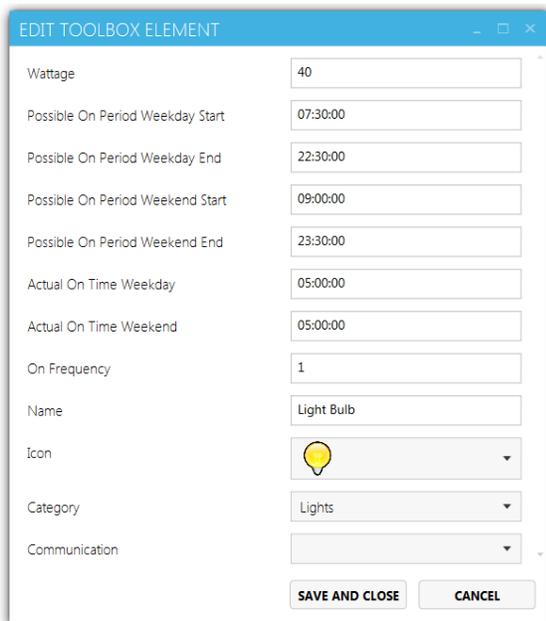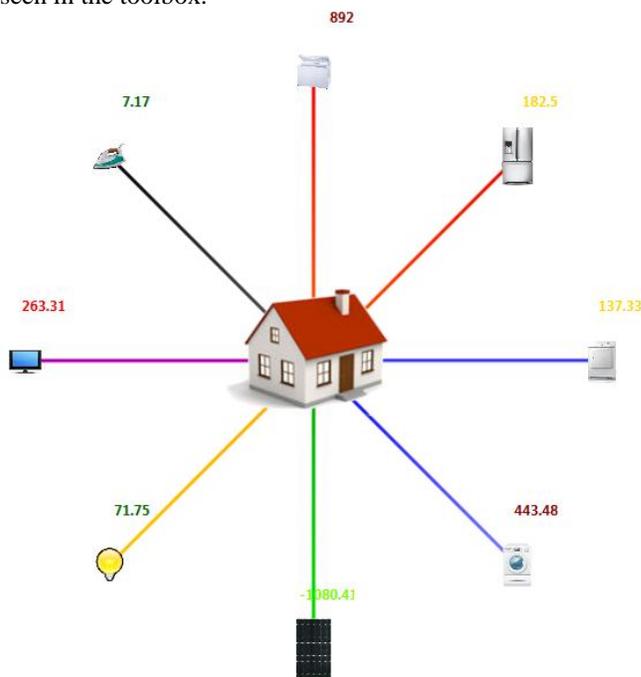
Figure 4.   The Edit Toolbox Element window that allows the user to edit properties of an element in the toolbox.

Figure 5.   Example of a modeled house with eight connected components.
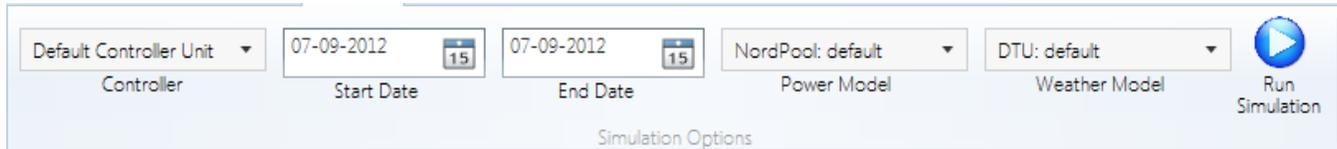
Figure 6   The simulation tab of the ribbon navigation menu.

C. Modeling - Summary

As mentioned above a graphical summary of the estimated consumption of the scenarios on the canvas is shown in a view to the right of the canvas (see Fig. 7). The summary contains a section for each scenario on the canvas in which the consumptions of the components connected to the scenario is summed up and grouped by their category. The estimated consumptions are also displayed as a pie chart along with an estimated total yearly consumption for the scenario. The color of the categories is the same as the colors used by the lines connecting the components with the home. This makes it a lot easier for the user to identify which components are members of what category such that the reasons for a high consumption in a particular category can be identified quickly.
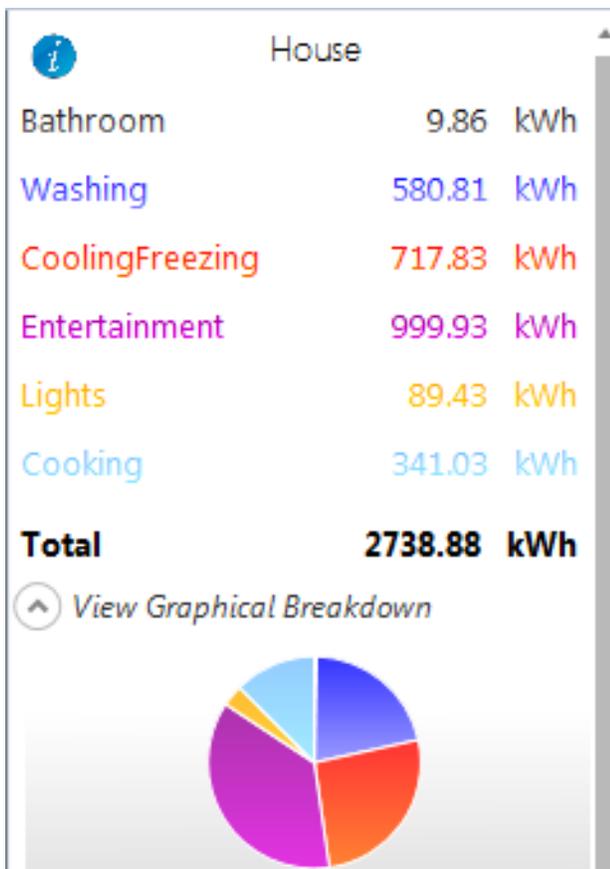
Figure 7.   The summary displaying a breakdown of the consumption of a scenario.

D. Input Data

There are currently two types of input windows in the solution, one for weather data and one for power data. They are two very similar windows in which it is possible to retrieve data from a selected data provider. The window can be accessed through the buttons from the input tab in the ribbon navigation. In the top of the window is a drop down menu containing the providers you want to retrieve data from. Multiple different weather and power providers can be added to the application and selected here. Next to the drop down menu are two calendars where the user can select the start and end date for the period that he want to retrieve data for. Clicking the "Get Data" button retrieves the specified data and displays it in the window. In the current implementation no CO_2 data is retrieved although it is part of the simulations as there are currently no accessible sources for this information.

To create custom data models the user can add or edit values in the fields by selecting the field with the cursor and then typing the new value into the field. This enables the user to create her own examples of data for special events like storms or increasing oil prices and use these when performing simulations.

E. Simulation

To perform a simulation the user has to navigate to the Simulation tab, which can be seen above in Fig. 6. In this tab the user can either inspect previously performed simulations or perform a new simulation, as described in this section. Before starting a new simulation the user has to consider the values in the five settings controls to the left of the "Run Simulation"-button.

As seen in Fig. 6 the users starts off by choosing which type of controller they want to use for their scheduling. The application supports having multiple different controllers using different simulation strategies to choose from. Then the users need to set the time span that they want to run the simulation over, by setting a start and end date. Finally it is possible for the users to select which power and weather data they want to use for the simulation. Here it is possible to select the earlier created input models (Weather and Power), such that they can simulate a user created setup. To run the simulation the user presses the "Run Simulation"-button and then waits for the program to notify when the simulation is finished. While waiting, the user can see the progress of the simulation in a window that opens when the simulation starts.

The simulation divides the modeled components into two main groups: Components which consumption can be automatically controlled and components which are turned on and off manually by users. Components that can be controlled will be scheduled based on the price model chosen for the simulation such that the total consumption of the component is as small as possible while still respecting defined rules for the component such as temperature limits for refrigerators or total on-time required for washing machines. Components that cannot be scheduled will consume energy based on the pattern defined by the user when modeling the scenario. The algorithms used for calculating the consumption of components are very general and the intention is that the prototype is extended with models of real components containing more specific calculations for the given component. The consumption of the components is calculated independently of each other.

F. Reporting

Based on every simulation performed, one or more reports can be made. A report consists of an overview of the simulation results along with a number of charts. The reporting window lets the user see and edit the reports of a simulation. The reporting window can be found either by clicking on the *"Show Report"* button when a simulation is finished or by choosing an older simulation in a drop down menu in the ribbon and clicking on the *"Inspect Selected Simulation"* button. The reporting window can be seen in Fig. 8.

At the top left is a drop down menu saying *"Apartment"* in Fig. 8. This menu allows the user to choose for which home to show the simulation result (in this screenshot the name of the chosen home is *"Apartment"*). The chart called *"Overview"* shows the consumption pattern of each of the components of that home in the simulated period.
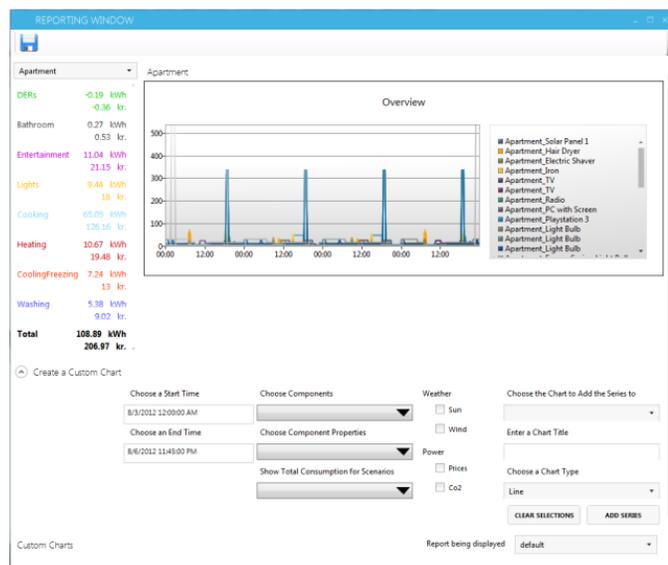
The x-axis shows the time and the y-axis shows the consumption at the given time in watt-hours. To the left of the overview chart is a view very similar to the summary. This area works similarly, but instead of showing the estimated yearly consumption it shows the consumption in the simulated period for each category as well as the total price for the power used by each category. These values are from the actual simulation rather than estimations and should therefore be more accurate than the ones found in the summary in the main window. These categories are also shown in the specific color of the category as in the summary. Upon cursor hovering over each category an estimation of the yearly consumption, price and CO_2 emission can be seen, calculated by scaling up the simulated results to match a year. Therefore the longer the simulation period is, the more accurate these results become (see Fig. 9).

Below the overview is a region called *"Create a Custom Chart"*. It provides the user with controls to create charts based on the simulation results. The charts can be generated from a wide range of parameters. The first column of controls allows the user to choose the start and end time of the data for the new chart. By default this is the first and last data entry of the simulation result. The second column allows the user to select which data series to add from the model. In the first drop down the user can select what components to include in the chart. The second drop down holds the properties to be displayed for the selected components such as price, CO_2 emission, consumption or on-time intervals. The last drop down allows the user to add data series to the chart that shows the total consumption of the homes selected in this drop down. All three drop down menus allow for multiple elements to be chosen at the same time.

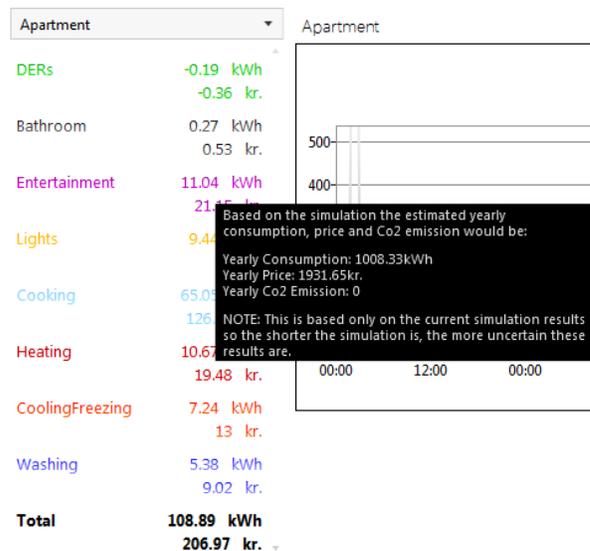Figure 8.   The reporting window used to analyze the results of a simulation.

Figure 9.   A screenshot showing the tooltip displayed when hovering the cursor over a category in the overview.

In the third column the user can choose which input data to display in the chart such as wind speeds, power prices etc. By adding these it becomes easy to see how the components have been scheduled compared to the fluctuating prices and weather. The final column specifies information about the chart that should hold the series specified in the other columns. The first drop down allows the user to add series to an existing chart or to choose to add them to a new chart. If "*New Chart*" is chosen (i.e. leaving the value empty) the user can specify a title and a chart type (line, bar, pie etc.) for the chart. Clicking the "*Add Series*" button creates the series and/or chart specified and adds them to the "*Custom Charts*" region.

Fig. 10 shows two examples of charts that have been created. The first one contains series showing the prices over time for a refrigerator and a washing machine in the home. The second chart shows the fluctuating power prices in the same period. It is important to note that all series (consumption, prices, wind speeds etc.) can be combined in the same charts. This provides a very high flexibility when creating charts, but also means that the y-axis of the custom charts is without unit as it may at the same time represent currency, watt-hours, m/s and more. Additionally, some series, such as consumption of washing machines, will show very large values whereas some series, such as power prices, will show very low values, making these series impractical to have in the same chart. Therefore it is advisable to separate them into two charts and compare these as shown in Fig. 10.

In the upper right corner of each chart are three buttons: Edit, save and delete. The delete button removes the chart from the report. The edit button opens an *Edit Chart* window (see Fig. 11) in which the user can change the name and type of the chart as well as remove series from the chart by clicking on the trash bin next to the series.
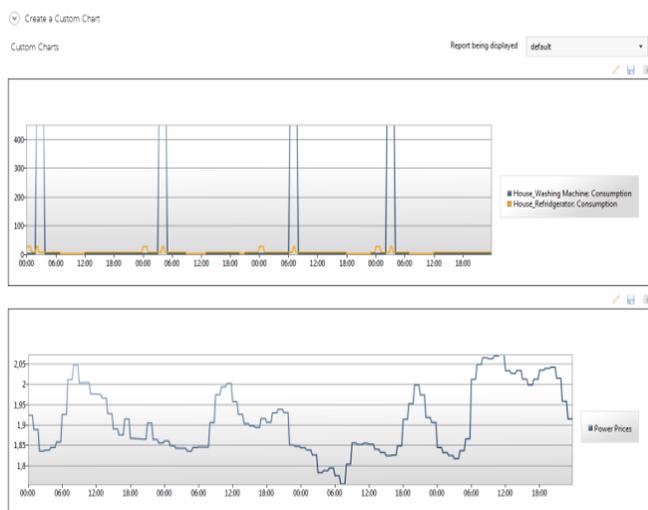
Figure 10   A selection of the reporting window showing the Custom Charts section in which the charts created by the user are displayed.
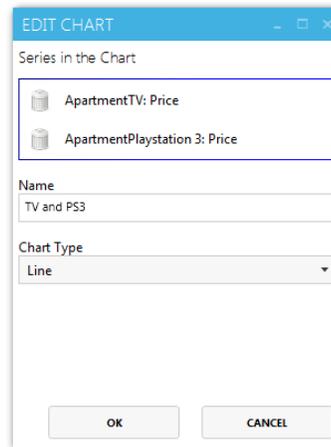
Figure 11   The view in which the user can edit an existing chart.

When a report has been created with the desired charts it can be saved using the save icon in the upper left corner of the reporting window. This allows the user to choose a name for the report and when it has been saved the user can toggle between which report to show, using the drop down menu just above the region with the custom charts. Since reports are connected to a single simulation result only reports connected to the same simulation result can be displayed this way. If reports connected to another simulation result needs to be displayed the reporting window must be closed and the desired simulation result should be found in the drop down menu in the ribbon using "*Inspect Selected Simulation*" as explained earlier in this section. It is, however, possible to have two reporting windows open at the same time showing different simulation results if comparison across simulations is required

IV.   CONCLUSION

As mentioned in the beginning of the paper the Danish workgroup, Smart Grid Netværket, recommended that the knowledge of the smart grid should be increased at consumer level in order to increase the acceptance of the smart grid. The presented work contributes to reaching this goal by allowing consumers to model their own home and see how their energy consumption would change if the appliances were smart grid enabled and the house used fluctuating energy prices rather than flat rate prices.

An analysis of the state-of-the-art simulation tools showed that very little work has been done in the area and that the existing tools are mostly non-graphical and therefore not very easy to use. The current work offers a low entry barrier for people to model their own home energy consumption. Additionally, previous work does not provide much opportunity to model specific homes and the incorporation of real energy and weather data is non-existent. Lastly the graphical analysis and comparison of different simulations were virtually absent in all previous work. These above mentioned facts provided the incentive for developing this simulation tool.

Additionally, a number of other areas have been identified in which the application may prove valuable. This includes salesmen that wish to illustrate the economic benefits of their product (see Fig. 1) and utility companies that wish to measure the flexibility of the homes connected to the grid, as identified during a meeting with a leading manufacturer of smart home controllers in Denmark.

The application has multiple uses. One of them is to answer the question: "How will the smart grid affect my energy bill?". This question can be answered by creating a model of the home and running a simulation using no smart grid enabled components and a flat rate power model and running a simulation using smart grid enabled components and fluctuating energy prices. A case study has been performed that models a typical four person house with dryer, washing machine, dishwasher, freezer, refrigerator etc. The case study showed that changing from flat rate to fluctuating energy prices the consumer can save €78. If the consumer also invests in smart grid enabling all the appliances the savings increase to €148 for the house [12]. This should be enough motivation for most users to participate in the smart grid [13]. Obtaining such a saving requires investments in new equipment and the price of such investments is difficult to estimate due to the limited range of products and the unsettled prices. Therefore the return of investment of the modeled setup is not easily predicted.

V. FUTURE DEVELOPMENT

The goal was to create an application that would be simple and easy to develop further. Therefore, one of the main focus areas of the application was extensibility and the application was designed with the best practice software principles of *loose coupling*. A list of the most significant suggested extensions is shown below.

- Integration with other simulation tools, such as GridLab-D, to include other aspects of consumptions for e.g. heating of homes.
- Include floor plans of the modeled household to provide more precise heating calculations.
- Provide a list of components based on the simulations that would provide lower energy consumptions than the components in the home.
- Support dependencies between components, such that the scheduling between them can be synchronized when simulating, such as TVs and DVD players.
- Historic consumption data from the user could be integrated to provide knowledge of usage patterns and thereby provide even more accurate results.
- Web based deployment such that the application can use more computational power and the users can get access to the application through a web browser.

The application's installation files can be found at [14].

REFERENCES

[1] Klima- Energi- og Bygningsministeriet, "Smart Grid Netværkets Arbejde - Sammenfatning og anbefalinger," 2011, p. 12.

[2] K. Haas, "Peaking power plants," 2009 [cited 2012 13/2]; Available from: http://www.energyandcapital.com/articles/peaker-plants-bridge-the-gap/991.

[3] Illinois Environmental Protection Agency. "Peaker power plant fact sheet," 2011 [cited 2012 13/2]; Available from: http://www.epa.state.il.us/air/fact-sheets/peaker-power-plant.html.

[4] K. Nozik, "Eastern power outage unfortunate but entirely predictable," n.d. [cited 2012 25/8]; Available from: http://www.frugalmarketing.com/dtb/electricity.shtml.

[5] Control4. "Introducing the Control4 EMS 100," 2012 [cited 2012 24/2]; Available from: http://control4.com/energy/products/.

[6] Tendril. "Tendril Connect," 2012 [cited 2012 24/2]; Available from: http://www.tendrilinc.com/platform/connect/.

[7] California Energy Commission and Charles River Associates, "Statewide pricing pilot: Summer 2003 Impact Analysis," Charles River Associates, 2004.

[8] GridLAB-D. "GridLab-D simulation software," 2011 [cited 2012 22/2]; Available from: http://www.gridlabd.org/.

[9] EnergyPlus, "Getting started with EnergyPlus, in basic concepts manual - essential information you need about running EnergyPlus," US Department of Energy, 2011.

[10] PowerMatcher Smartgrid Technology, "Basic structure and agent roles," n.d. [cited 2012 6/6]; Available from: http://www.powermatcher.net/get-technical/powermatcher-technology/basic-structure-and-agent-roles/.

[11] PowerMatcher Smartgrid Technology, "PowerMatcher simulation tool," Energy Research Centre of the Netherlands, 2012.

[12] M. Nysteen and H. Mynderup, "A smart home simulation tool – the development of a simulation tool for measuring the impact of a smart grid on a private home," The Technical University of Denmark, 2012, pp. 1-148.

[13] Konkurrencestyrelsen, "Konkurrencen på detailmarkedet for el," 2009, pp. 9-14.

[14] M. Nysteen and H. Mynderup, "Program Installation," https://www.dropbox.com/s/8eiuhsmyfusvuwk/Program%20Installation.rar

Analysis of Energy Saving Effects on Korean Style Residential Buildings using Energy Information Device for Smart Metering

Yoon-Sik Yoo and Il-Woo Lee

Smart Grid Technology Research Team
Electronics and Telecommunications Research Institute
218 Gajeongno, Yuseong-gu, Daejeon, KOREA
{midasyoo, ilwoo}@etri.re.kr

Abstract— **The paper focuses on monitoring and analyzing the energy consumption load profile patterns of consumers and implementation of integrated operation monitoring system using distributed energy information device to attract active energy conservation. This paper also analyzes the electric power consumption patterns in energy consumption data gathered from the monitoring system according to Korean style residential building type and discuss the feasibility of considering of energy saving effectiveness factor.**

Keywords-Monitoring; Analysis; Load Profile; Energy Savings; IHD; Smart Metering

I. INTRODUCTION

Reducing energy demand in the residential buildings is an important problem worldwide. Recently, the trends toward case of using energy information device such as heterogeneous In-Home Displays (IHDs) by consumers are gradually growing. In addition, studies for analyzing energy saving effect depending on the use of IHD are ongoing actively. Consumers using IHD mainly receive feedback from energy information of that for energy saving. Related researches are as follows. Faruqui et al. covers 12 direct feedback trials using an IHD device between 1989 and 2009, including time of use and prepayment trials. Ehrhardt-Martinez et al. [2] covers 57 studies from 1974 to 2010. This is the most comprehensive review to date and splits the conservation effects from feedback by study size, era, type of feedback and region. In [3], companies with specific, proven expertise in making attractive consumer products are quickly changing the IHD industry and competition is growing. For example, General Electric (GE) recently announced its Brillion suite of home energy management products, including an IHD. Other researches discuss about direct feedback [4][5] and the inexpensive IHD device [6]. IHD is also provided with important information using various communication interfaces (Ethernet, ZigBee [7][8], and WLAN [9]) and security policy [10]-[12].

So, in aspect of usage effectiveness of heterogeneous IHD, we developed Integrated Operation Monitoring System (IOMS) in order to monitor meter data and analyze the effect of the energy saving of heterogeneous IHD as an extension study of [13].

We completed the prototype level development of the IOMS, and gather and manage the metering data from the IHD server. Therefore, the main purpose of this paper is to develop the IOMS that can be monitored by the integrated management of metering data of the distributed IHD server, and the second objective is to analyze the pattern of metering data using the IOMS. Finally, based on the results of the pattern analysis, we are about to proceed to reflect to the government policy or promote to the energy savings policy.

Eventually, this IOMS is useful for integrated management and analysis of metering data of the customers, and has the usefulness for the taking advantage of the energy savings policy and promotional material.

The remainder of this article is organized as follows: In Section II, we introduce IOMS implementing architecture. In Section III, we analyze energy consumption patterns of experimental result. Finally, we present the conclusion in section IV.

II. IOMS IMPLEMENTING ARCHITECTURE

In this section, we implement the IOMS and collect metering data. After that we examine the empirical results.

A. IOMS Interoperable Networking

IOMS is connected with each heterogeneous IHD server through 70Mbps Ethernet (IEEE 802.3). The heterogeneous IHD server and AMR server are also connected with 100Mbps Ethernet. The IHD server is connected to several AMR servers, and accesses directly many IHDs through the internet. Two types of IHD can be divided into IPTV-based IHD and Dedicated Terminal-based IHD. Electricity meters to measure the power usage are linked to the AMR server through 0.2Mbps RS-485 communication line. The AMR server gathers power usage information from each electricity meter and transmits the aggregated data to the IOMS. Fig. 1 describes interoperable networking with each heterogeneous IHD server and IOMS.

B. Implementation Requirement and Function

The IOMS sends TCP/IP protocol-based control to IHD server and receives communication messages from IHD server for collecting and saving the meter data. The IOMS collects meter data from heterogeneous IHD server periodically. The IOMS checks the status of heterogeneous IHD server. If the heterogeneous IHD server is off, the IOMS saves the status of the heterogeneous IHD server to the DB of IOMS and stops the collecting of meter data. After
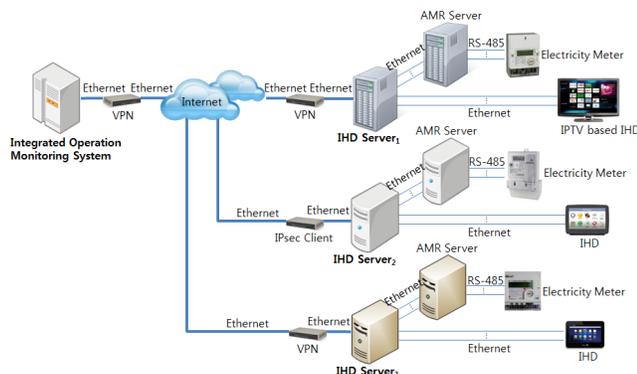
Figure 1. Interoperable Networking Configuration

that, if the status of heterogeneous IHD server is on, the IOMS collects and saves meter data again. In order to these functionalities of the IOMS, the following requirements must be met.

- Ethernet-based communication interface with IHD server must be satisfied.
- Meter data received from IHD server should be encrypted.
- Meter data of IOMS must be synchronized with that of IHD server.
- Scheduler must read meter data of IHD server periodically.
- In order to user-friendly meter data lookup, web-based user interface should be implemented.

To satisfy IOMS requirement, functional blocks can be divided into Network Layer, Data Processing/Management Layer and Data Presentation Layer. The Network Layer consists of TCP/IP-based data communication block for meter data collection and monitoring. The Data Processing/Management Layer consists of management engine block, analysis engine block and web service application server block. The Data Presentation Layer consists of operator web portal block for web-based user retrieval. Fig. 2 shows IOMS layered architecture.
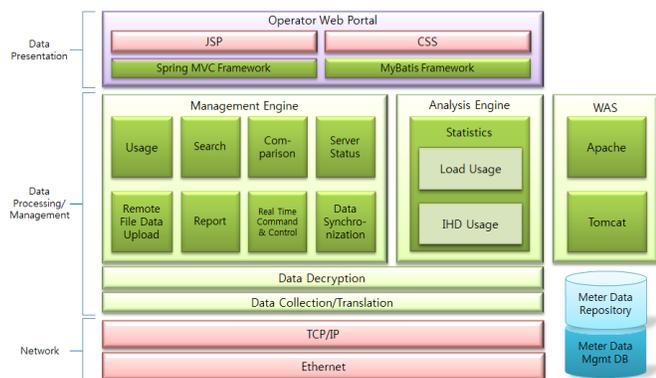
Figure 2. IOMS Layered Architecture

TABLE I. SEARCH CRITERIA ITEMS IN IOMS

Event Item	Event Time Zone	Load Usage	Energy Items	Search Period	Customer Number	Search Area
Selection	Hourly, Daily, Monthly, Yearly	Average, Maximum, Total	Electricity, Gas, Water	Hour, Day, Month, Year	Unique number of IOMS	The whole country
IHD Service	Current Month Charges, Monthly Charges, Hourly Statistics, Daily Statistics, Monthly Statistics, Daily-Monthly Usage, Annual Usage, Usage Comparison					

At first, the management engine block includes following meter data management functions: energy usage view and retrieval, which is decrypting the collected metering data and then stored in the meter data repository, comparison, business IHD servers status identification, file-based remote meter data upload, documenting for energy usage result and extracting to its report, and IHD usage. Secondly, the analysis engine block includes energy consumption statistics and IHD usage analysis. And finally, the web service application server block provides the meter data to web services. The Data Presentation Layer consists of user interface block for intuitively understanding graphs, provided by a java-based GUI in the local and remote meter data query, search, and comparison through the web.

Looking at the main features offered by the IOMS, the IHD server status identification function checks that all the IHD severs are basically connected on the network for collecting the meter data. If the IHD server makes an error during collecting the meter data of the IOMS, the IOMS attempts to resolve the status of the IHD server for three times at 10-second intervals. The IOMS recognizes the normal state of the IHD server if the IHD server responds to the IOMS during the checking of the IHD server status. However, otherwise, the IOMS recognizes abnormal state of the IHD server and should retry to check the status of the IHD server on every hour.

Energy usage search and query function retrieves related information, applied with meter data which were collected from the IHD server through various search criteria, and confirms the results of search query. Search criteria could be divided into search item, energy type item, period (hourly, daily and monthly), business and search region. In this point, we can select the item of "date", "time zone", "weekday" or "weekend + holiday" if want to look up the period as a time zone average usage. And we do not need the item of time zone if we want to look up the period as an average daily usage or an average monthly usage, but only can select the item of "date", "weekday" or "weekend + holiday". The items of search and lookup are shown in Table I.

The energy usage comparison function can compare the experimental group and the control group by setting the search criteria and selecting the item that we want to compare on the time zone, daily, monthly. The file-based remote meter data upload function provides meter data of a

TABLE II. INTERACTIVE OPERATION ENVIRONMENT

Type	IOMS	IHD Server₁	IHD Server₂	IHD Server₃
OS	Windows Server	Linux Redhat	Windows Server	Linux CentOS
DB	MSSQL	Oracle	MSSQL	MySQL
VPN	IPSec VPN Server (HW)	IPSec VPN Client (HW)	IPSec VPN Client (SW)	IPSec VPN Client (HW)

Figure 4. Secure Internetworking Architecture for Metering Data

missed section to the IOMS through web GUI by administrator directly among meter data that were collected from IHD servers.

From the items of search criteria, the administrator can also upload file-based meter data to the IOMS after confirming the missed section with selecting energy item, business, search period, and search region. The reporting function for documenting and extracting the results of energy consumption supports easy documentation, analysis from remote client, and energy meter data lookup in the web with search criteria. According to these search criteria, the energy usage statistics function provides analytical data, which can be used for analysis of various results, with identifying the information about the energy average usage, the energy peak usage and the consumption total energy. The IHD usage statistics function supports the ability to analyze the IHD utilization level of customer from current month price lookup, monthly fee views, hourly statistics, daily statistics, monthly statistics, daily usage inquiry, monthly usage inquiry, annual usage inquiry, and usage comparison.

The Operator web portal block using JSP user interface also provides users with location-based graphic UI of AMR server and IHD server depicted as Fig. 3. Operation

Figure 3. Location-based UI of AMR server and IHD server

environments of each heterogeneous IHD server to interact with IOMS are depicted as Table II.

C. Secure Internetworking Architecture

The IOMS is connected to all the IHD server through VPN tunneling. The IPSec VPN server is installed on the external network side of the IOMS and The IPSec VPN client is installed on the external network side of the IHD server. In other words, the IPSec VPN hardware-based client is installed in front of some IHD server and the IPSec VPN software-based client is installed in the rest internal IHD server. Therefore, the VPN tunnel must be formed between the VPN server and the VPN client if we want to complete the VPN connection between the IOMS and the IHD server. The VPN tunnel is formed if the encryption key value is matched between the VPN server and the VPN client after the encryption key value is exchanged between VPN server and the VPN client. When the VPN tunnel is established, the authentication is completed through Secure Hash Algorithm 1 (SHA1) between the VPN server and the VPN client and the data is encrypted with 3 Data Encryption Standard (3DES) between the IOMS and the IHD server. We are considering the security aspects for collecting the metering data from the public network via the IHD server. This means that we need to collect the metering data from the public network via the IHD server in aspect of security. In addition, the metering data of all the IHD server is encrypted with Advanced Encryption Standard (AES 128) in order to protect more securely. Therefore, the metering data is encrypted with AES 128 in the IHD server and the encrypted data is re-encrypted to 3DES in VPN, and finally the re-encrypted data is delivered to the IOMS. The Fig. 4 shows the secure internetworking architecture for collecting the metering data.

D. Data Internetworking Procedure

The VPN-based data flow for collecting and saving the metering data between the IOMS and the IHD server is depicted as Fig. 5. At first, the VPN tunneling is negotiated and established between the VPN server and the VPN client. Then, the VPN tunnel is completed between the IOMS and
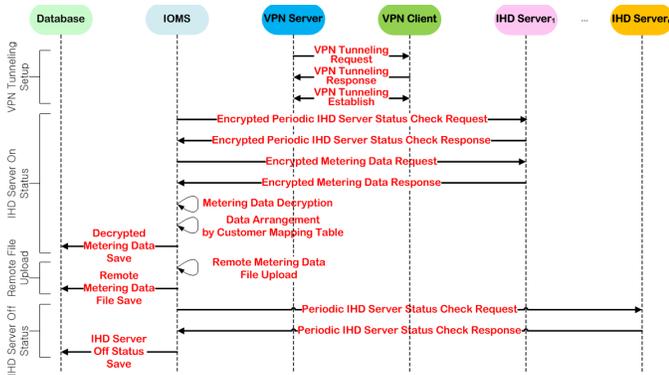
Figure 5. Data Flow of VPN-based Metering Data Gathering and Save

the IHD server. The IOMS also identifies the IHD server status periodically through VPN tunnel and collects and saves the metering data to the DB. The metering data collected from the IHD server decrypts the encoded data due to encryption of the metering data, and because subscriber numbers are different from each IHD servers, the metering data is stored in the DB after converting into the unique subscriber numbers. Among the metering data collected from the IHD server, we upload the missed meter data with file format through the web GUI of the IOMS, and manage and save the missed metering data. Additionally, when the IHD server is off state, the IOMS saves its state in the DB and stops collecting metering data.

III. ENERGY CONSUMPTION ANALYSIS

In this part, we monitor meter data through IOMS and analyze the power consumption patterns in according to usage of heterogeneous IHD. The consumer uses the IHD for identifying the energy usage information. So the IHD function is inducing energy saving actively. From this perspective, we expected different energy saving rate in according to these energy information display infrastructure. So we distinguished IPTV-based IHD (IIHD) from Dedicated Terminal-based IHD (DIHD) by information display infrastructure. Mandatory information is the same in all IHD, but optional information may be different depending on the type of terminal.

In case of Apartment on a weekday in 2012, we analyzed the hourly average power consumption. But in case of Commercial Building and Apartment-type Factory, we only analyzed the hourly average power consumption pattern. The customers of the number of 19,720 were participated in the survey for the Apartment, and the customers of the number of 2,780 and 4,000 were involved in the survey respectively for the Individual Commercial Building and the APT-type Factory. Table III represents the hourly power consumption in the Apartment between the experimental group of the IIHD and the control group of the DIHD.

As a result, the average daily power consumption per household is 10 kWh. Fig. 6 describes load profile of hourly average power consumption of apartment in case of IIHD

TABLE III. HOURLY POWER CONSUMPTION BETWEEN EXPERIMENTAL GROUP AND CONTROL GROUP

Type	Hourly Power Consumption (kWh)				
	Experimental Group				Control Group
Time (Hour)	$IIHD_1$	$IIHD_2$	$IIHD_3$	$IIHD_4$	$DIHD_5$
1	0.56	0.61	0.56	0.56	0.42
2	0.46	0.48	0.46	0.45	0.34
3	0.37	0.4	0.38	0.37	0.29
4	0.32	0.35	0.34	0.32	0.26
5	0.3	0.33	0.33	0.3	0.26
6	0.3	0.3	0.32	0.3	0.25
7	0.31	0.32	0.34	0.3	0.29
8	0.38	0.38	0.41	0.37	0.37
9	0.49	0.51	0.51	0.46	0.43
10	0.5	0.49	0.51	0.49	0.43
11	0.45	0.46	0.48	0.44	0.39
12	0.41	0.42	0.43	0.42	0.36
13	0.39	0.4	0.41	0.39	0.36
14	0.39	0.37	0.38	0.38	0.36
15	0.37	0.38	0.37	0.39	0.35
16	0.37	0.36	0.37	0.38	0.35
17	0.39	0.38	0.38	0.42	0.39
18	0.43	0.41	0.4	0.43	0.43
19	0.48	0.46	0.46	0.48	0.46
20	0.56	0.54	0.56	0.54	0.53
21	0.61	0.63	0.63	0.57	0.55
22	0.63	0.66	0.63	0.61	0.58
23	0.66	0.65	0.65	0.6	0.58
24	0.64	0.64	0.64	0.6	0.54
Mean	0.45	0.46	0.46	0.44	0.40
Total	10.77	10.93	10.95	10.57	9.57

and DIHD. As we can identify this graph, households using DIHD reduce more 0.05 kWh than that of households using IIHD. This result implies as follows. The more the customer is exposed to the energy usage information, the more the energy saving consciousness can be affected. In other words, DIHD provides energy usage information during 24-hour-a-day, but IIHD provides energy usage information during watching TV. Therefore, we can expect to provide the motivation to participate in active energy savings in case of using DIHD.

And energy saving factors can be considered as the method of information and the time of information. In this paper, we focused on the comparative analysis of the heterogeneous IHD users and Korean Style Residential Buildings. However, for the qualitative analysis, we will study the research topic of comparative analysis issues between metering data and questionnaire survey. In other words, in order to gain some qualitative insight from collected metering data, we will study the energy saving index through conducting a survey of participants.

As shown in Fig. 7, for the seasonal hourly average power consumption among the weekday, we can see that the apartment residents who use the IHD consume more 1.7 kWh in the summer than compared to other seasons with the average daily power consumption. Due to using the air
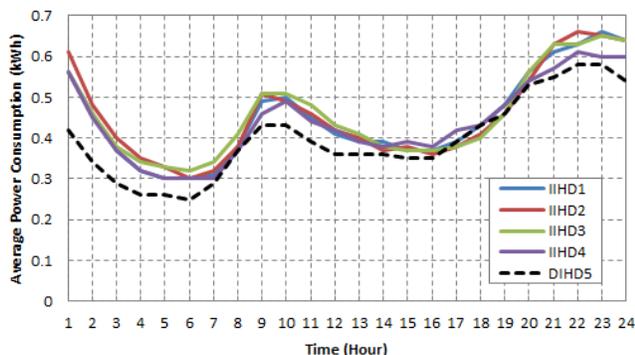
Figure 6. Hourly Power Consumption Load Profiles among Heterogeneous IHDs

Figure 8. Hourly Power Consumption Load Profiles among Korean Style Residential Buildings

conditioning in the summer, we can see that the power consumption in the summer compared to the other seasons will be used more.

Also, we compared the hourly average power consumption among the DIHD-based APT, the DIHD-based APT-type Factory and the DIHD-based Individual Commercial Building. Because the APT-type Factory and the Individual Commercial Building use only DIHD, we cannot compare directly the DIHD and the IIHD. Thus, depending on the type of each building, we are about to distinguish the intuitive difference from the hourly average power consumption. As shown in Fig. 8, in case of the APT-type Factory, the hourly average power consumption is approximately 5 times more than that of the Apartment. And in case of the Individual Commercial Building, we know that the hourly average power consumption is approximately 6 times more than that of the Apartment. In addition, the APT-type Factory consumes more 3.21 times than that of the Apartment in terms of the average daily power consumption and the Individual Commercial Building consumes more 3.42 times than that of the Apartment in terms of the average daily power consumption.

In future research, we will analyze the energy saving rate

using DIHD through the power statistical reports of the APT-type Factory and the Individual Commercial Building.

IV. CONCLUSION AND FUTURE WORK

In this paper, we designed the IOMS layered architecture, and developed IOMS to interact with Heterogeneous IHD server. The IOMS monitors energy usage information from the Korean style residential building which is installed with Heterogeneous IHD, and collects energy usage information. When we analyzed energy usage information patterns in IOMS, we found that the increasing frequency of information providing of the energy consumption is proportional to the energy savings. Therefore, we concluded that the method of information and the time of information as well as seasonal factors, residential building types are important factors in energy saving effectiveness through this study.

For future study, we will analyze a wide variety profile-based energy usage pattern of Korean style residential building type and will research the correlation of energy saving effectiveness through comparative analysis between non-IHD user and IHD user.

Figure 7. Hourly Power Consumption Load Profiles for Seasonal among Apartment Residents

REFERENCES

[1] A. Faruqui, S. Sergici and A. Sharif, "The Impact of Informational Feedback on Energy Consumption- A Survey of the Experimental Evidence," International Journal of Energy, vol. 35, no. 4, pp. 1598-1608, Apr. 2010.

[2] K. Ehrhardt-Martinez, K.A. Donnelly and J.A. Laitner, "Advanced Metering Initiatives and Residential Feedback Programs: a Meta-Review for Household Electricity-Saving Opportunities," Report E105, ACEEE 2010.

[3] GE, "Home Energy Manager, Energy Monitor: GE Nucleus," [online] available at <http://www.geappliances.com/home-energy-manager/> [retrieved: January 15, 2013].

[4] S. Darby, "The Effectiveness of Feedback on energy Consumption. A Review for DEFRA of the Literature on Metering, Billing and Direct Displays," Environmental Change Institute, University of Oxford, 2006.

[5] C. McKerracher and J. Torriti, "Energy Consumption Feedback in Perspective: Integrating Australian Data to Meta-Analyses on In Home Displays," International Journal of Energy Efficiency, Aug. 2012.

[6] S. Park, H. Kim, H. Moon, J. Heo, and S. Yoon, "Concurrent Simulation Platform for Energy-Aware Smart Metering Systems," IEEE Trans. Consumer Electron., vol. 56, no. 3, pp. 1918-1926, Aug. 2010.

[7] B. Park, J. Yoo, and H. Kim, "Modeling and Analysis of Queuing Effect of Two-Level Approach to Network Localization," ETRI Journal, vol. 34, no. 4, pp. 625-628, Aug. 2012.

[8] H. S. Kim and J. Yoon "Hybrid Distributed Stochastic Addressing Scheme for ZigBee/IEEE 802.15.4 Wireless Sensor Networks," ETRI Journal, vol. 33, no. 5, pp. 704-711, Oct. 2011.

[9] J.-H. Kim, H. Yu, and S.-K. Lee, "Channel Estimation Scheme for WLAN Systems with Backward Compatibility," ETRI Journal, vol. 34, no. 3, pp. 450-453, June 2012.

[10] L. Zhang, Q. Wu, and Y. Hu, "Hierarchical identity-based encryption with constant-size private keys," ETRI Journal, vol. 34, no. 1, pp. 142-145, Feb. 2012.

[11] K. Rajendiran, R. Sankararajan, and R. Palaniappan, "Secure Key Predistribution Scheme for WSN Using Elliptic Curve Cryptography," ETRI Journal, Vol. 33, no. 5, pp. 791-801, Oct. 2011.

[12] T.-H. Chen and W.-K. Shih, "A Robust Mutual Authentication Protocol for Wireless Sensor Networks," ETRI Journal, Vol. 32, no. 5, pp. 704-712, Oct. 2010.

[13] Y.-S. Yoo, W.-K., Park and I.-W. Lee, "Integrated Energy Management System Framework for the Distributed Smart Meters in Smart Grid Environment," in Proc. of the International Conference on IT Convergence and Security, pp. 567-572, Dec. 2011.

Communication Network Architectures based on EPON for Offshore Wind Power Farm

Mohamed A. Ahmed
Department of Computer Engineering
Chonbuk National University
Jeonju, Korea
mohamed@jbnu.ac.kr

Young-Chon Kim
Smart Grid Research Center
Chonbuk National University
Jeonju, Korea
yckim@jbnu.ac.kr

Abstract—**Wind power gains a great attention among renewable energies because of its relative cost competitiveness and the maturity of the technology. This paper proposes the EPON-based communication network architectures as a promising candidate to replace the current switched-based networks for wind power farm (WPF). EPON-based networks combine the advantages of both the optical network technology and Ethernet technology. The proposed WPF network architecture is based on the electric transmission topology and consists of a central optical line terminal (OLT) placed in the control center and optical network units (ONUs) deployed on the wind turbines (WTs) side. There are no active electronic devices used between the OLT and the WT-ONUs to reduce the cost and complexity of maintenance and deployment. Two different EPON-based architectures for an offshore WPF have been proposed and compared in terms of optical power budget, path loss and network cost. The power budget is analyzed to ensure that the received signal power is enough to maintain acceptable performances even though the lengths of communication link between the control center and WTs are different. The total path loss includes the power reduction from fiber, connectors and splitters. The results show that the received power of each WT in EPON-based architecture satisfies the power budget requirements of IEEE standard. Also, the analysis of network elements shows that EPON-based architectures have a cost effective solution.**

Keywords-smart grid; SCADA; EPON; power budget; wind power farm.

I. INTRODUCTION

Nowadays, large-scale wind power farm become a reality. Many offshore projects were scheduled to be constructed in the coming years, to overcome the fossil fuel dependence, and reduce CO2 emissions. These platforms represent many engineering challenges related to electrical connectivity and communication infrastructures [1]. Current communication infrastructures were designed several decades ago, based on old technologies that have a low bandwidth such as telephone lines which need to be upgraded. To meet the demand of control operations and smart grid applications, communication infrastructures need fundamental changes to be reliable, scalable and extendable to future services and applications [2]. Supervisory control and data acquisition (SCADA) is the basis for monitoring, control and data gathering of wind power farm (WPF). It allows for remote control, monitoring of individual turbines, and monitoring the wind farm as a whole from central control station or from a remote control center. Furthermore, it enables full communication with the operator, manufacturer and maintenance staff [3].

Ethernet passive optical network (EPON) technology is an optical access network, designed to serve customers with voice, video and data services. EPON is broadly deployed in Korea and Japan while GPON (Gigabit PON) is widely deployed in parts of the U.S. and Europe [4]. EPON-based networks were developed to solve the bandwidth bottleneck in the access network. Recently, many researchers have begun to address the applications of EPON technology in smart grid and power system.

In China, there are many successful installations of communication network based on EPON technology in electric power system such as Yangjiaping power distribution [5]. Also, the authors in [6] present an integrated communication platform of a photovoltaic (PV) power station, which integrates EPON, power line carrier and wireless communication. The authors of [7] propose transmission lines monitoring system based on long distance EPON and wireless technology of Wi-Fi.

Modern large-scale WPFs are located at offshore, with growing in wind farm size (number of WTs) and wind turbine capacity. These platforms represent great challenges in view of accessibility, reliability, cost and maintenance. The conventional communication infrastructures are switched-based, where each wind tower comprises an Ethernet switch (ESW) to interconnect communication devices inside the wind tower and the ESW requires electric power supply in the WPF field [8]. The prices of these industrial switches are expensive compared with the normal Ethernet switch [9]. The EPON is one of the promising candidates for next generation WPF because it provides high performance communication network with low cost, low power consumption and high reliability. The EPON doesn't use active components between OLT and ONUs. This may result in the limitation of growing wind farm size because optical power budget operates as a bottleneck for EPON-based WPF architectures.

In this paper, we proposed two different communication network architectures based on EPON technology for offshore WPF. The power budget is calculated and compared to ensure that the received signal power is enough to

maintain acceptable performance under all conditions, due to limitation of EPON optical power. The optical power budget calculation in this paper includes losses in fiber, connectors and splitters. Also, we analyze the network cost to ensure the advantages of proposed architectures.

The rest of this paper is organized as follows. Section 2 presents the related work. In Section 3, the WPF layout is described. Section 4 details the proposed EPON-based network architecture with two different configurations; architecture (A) and (B). Section 5 shows the optical power budget, path loss and network cost calculation and discussion. Finally, Section 6 presents the conclusion and the direction of future work.

II. RELATED WORK

A. WPF Architecture

First, The WPF consists of individual WTs connected together, and tied to the collector system. From the electrical topological point of view, the WTs are connected to the utility system through transformers and distribution lines, including one or more meteorological mast [10]. From the communication network point of view, the WPF includes the integration of SCADA systems, wind power management systems, protection and control systems, and various intelligent electronic devices (IEDs) [11].

The information gathered from the WPF is processed for control and protection functions. Logically, according to IEC 61850, the WPF comprising of three levels; process level, bay level and station level. In the process level, sensors and IEDs transmits/receives small raw/control data packets [12]. At the station level, different devices are existing including for example; WTC (wind turbine controller), condition monitoring system (CMS), HMI (human machine interface), circuit breaker (CB) control, metering server, meteorological server and WPF server. Fig. 1 shows the WPF information flow, where each server provides information of, or controls to its devices.
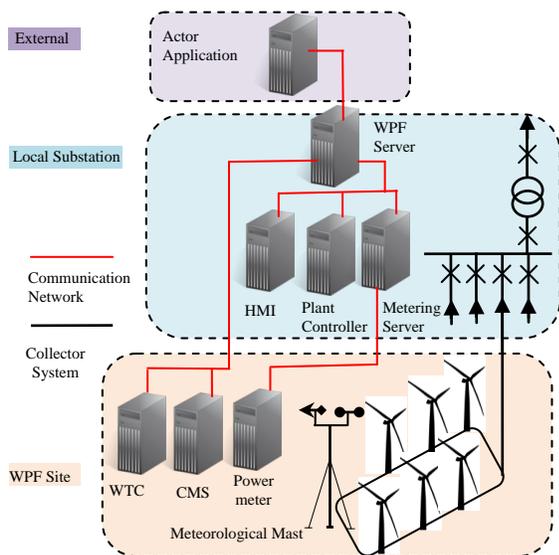
Figure 1. Logical architecture of WPP
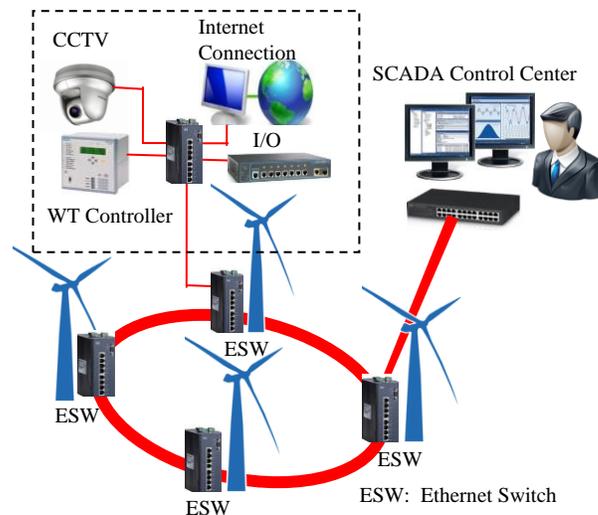
Figure 2. Schematic view of conventional WPF network [14]

B. Swiched-based Communication Networks for WPF

The communication systems for WTs can be classified into two different categories [13][14]:

a) **Inside the turbine tower**, including the WT controller, sensors and IEDs. This network supports different traffic such as CCTV system and internet connection to enable monitoring and controlling of the WT tower, as shown in Fig. 2;

b) **Between wind turbines** (within the WPF), including the connection between wind towers. This part constitutes the SCADA system. Due to the high cost of laying cables, optical fibers are integrated with the medium voltage and may follow the electrical topologies. The most widely types of communication networks are radial, ring and star.

III. WIND POWER FARM LAYOUT

A. Offshore WPF-Collector System Layout

We designed the EPON network architectures based on the electric system layout of medium size offshore WPF, with four radials, 48 WTs and different cable lengths as shown in Fig. 3. All electric radial cables are connected to the offshore platform where the voltage is stepped up and the power is transmitted to shore. The circuit breakers and the main protection instruments are configured at different regions of the system [15][16]. The WPF is divided into four groups with 2Km space between each other. Each group is served by an electric power cable connected from the offshore platform to the nearest WT. Inside each group; the spacing between individual WTs is equivalent to 1Km along rows and between rows.

The longest cable length between the offshore platform and WTs is 4.47Km, connected to WT-B01 and WT-G01, while the short cable length is 2.23Km connected to turbines WT-D01 and WT-E01. The submarine cable between the point of common coupling (PCC) from grid side and the offshore platform is 5Km. Due to the high cost of laying

cables; the logical option is to use the same power distribution cable routes for the SCADA communication network. In this work, the optical fibers are integrated within the electric submarine cables, therefore the communication network follow the wind farm electrical topology.

B. Design of WPF Communication Network Based on EPON Technology

EPON technology is an optical access network, consists of optical line terminal (OLT), optical network units (ONU), and passive optical splitter (POS). The physical layer uses PON technology while MAC layer uses the Ethernet technology which combines the advantage of low flexibility, low cost and high bandwidth [17]. Fig. 4 shows the proposed structure of the WPFs based on EPON technology. There are three main architectures; architectures (A&B) represent the design of new planned WPFs, while architecture (C) illustrates modifying existing WPFs to support EPON technology.

The proposed network model consists of ONU deployed on the WT side, for collecting data from different internal networks. All ONUs from different WTs are connected to a central OLT, placed in the control center. There are no active electronic elements used between the ONUs and the OLT, which reduce the costs and complexity of maintenance and deployment.

PON network architectures include different configuration such as star, ring and bus which can be applied for different WPF configurations. Also, it can be deployed in a redundant configuration as a double ring or a double tree. In Fig. 4, in control center side, there are two OLT connected to different WPF servers through Ethernet networks.

Figure 3. Layout of wind power farm with four radials and 48 WTs
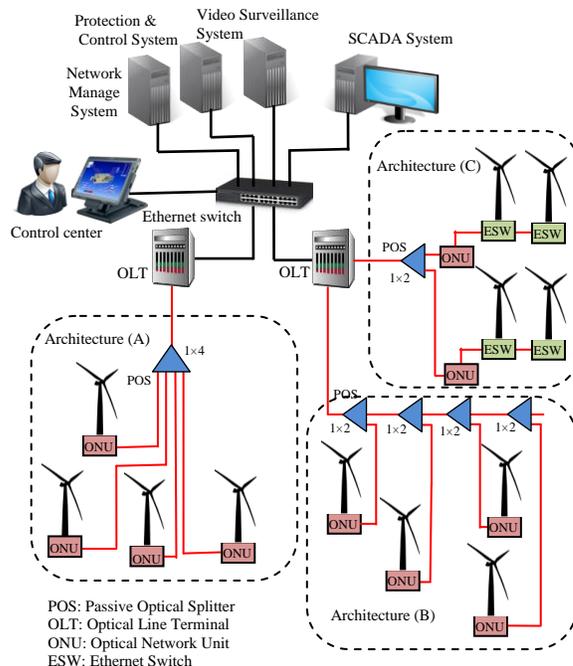
Figure 4. Schematic view of proposed EPON-based architectures for WPF

Architecture (A) represents star configuration, there are four ONUs connected to one POS using distributed fibers, where one ONU supports one standalone wind turbine traffic. In architecture (B), there are four POSs connected in cascade, each ONU supports standalone WT traffic.

Architecture (C) shows the radial design, where all WTs are connected in series to one feeder. This hybrid architecture comprises both Ethernet switches (ESW) and ONUs. The traffic from each feeder is supported by one ONU, transmitted to control center through POS. All transmissions in a PON are performed between OLT and ONUs. Therefore, in the downstream direction (from OLT to ONUs), PON is a point to multipoint network, while in the upstream direction; it is a multipoint to point network.

IV. EPON-BASED NETWORK ARCHITECTURE

This section considers two different network architectures for WPF. All WTs are connected together in a radial topology from WTs (A01→A06) till WTs (H01→H06) as shown in Fig. 3. We assumed that:

- An OLT unit is installed in the control center;
- Each WT has only one ONU device;
- Different network architectures are designed, using feeder fiber (FF), distributed fiber (DF) and passive optical splitter (POS).

For simplicity, the detail description of communication network for one group is given is more details, which is applicable for the whole WPF. The primary POS (1×8) is located on the offshore platform. Part of its output ports are connected to the four WTs groups, while the remaining ports are connected to IEDs and protection devices located on the offshore platform.

POS: Passive Optical Splitter
OLT: Optical Line Terminal
ONU: Optical Network Unit
FF: Feeder Fiber
DF: Distribution Fiber

Figure 5.   Schematic view of a medium size WPF

TABLE I.   NETWORK ELEMENTES OF WPF

	# OLT	FF [Km]	# Prim. POS	Sec. FF [Km]	# Sec. POS	# Other POS	Other DF [Km]
A	1	5	1 (1×8)	13.4	4 (1×2)	48 (1×2)	44
B	1	5	1 (1×8)	13.4	4 (1×2)	8 (1×6)	124

Architectures A: This begins with an OLT unit located in the control center, connected with a 5Km feeder fiber to the primary POS (1×8) located on the offshore platform. Cascade splitters are used to reduce the amount of deployed fiber in the network. The primary POS is connect to secondary POS in WT-B01, while the other output ports are connected to other secondary POS located at WT-D01, WT-E01, WT-G01 respectively. The WT-B01 has the two identical POS (primary and secondary); both are (1×2). For all other wind turbines (WT-A01→WT-A06 and WT-B02→WT-B06), each WT has only one POS (1×2); one port is connected using DF to the next WT, while the other port is connected to the WT-ONU unit.

Architectures B: It differs from configuration (A) as it has less number of POS. Secondary POS is (1×2), the same like configuration (A). The first output port from secondary POS is connected to another POS (1×6) exist in WT-D01 with six output ports. Each WT from (WT-D01→WT-D06) has dedicated optical fiber cable from WT-D01. Also, the POS exist in WT-C01 is (1×6) connects a dedicated path to (WT-C01→WT-C06). Group (2) in Fig. 5 shows the network elements of configuration (B).

V.   PERFORMANCE ANALYSIS

In this section we will study the performance of proposed WPF architectures with respect to power budget, path loss

and network cost. In our network model, the OLT located in control center use a single port to serve the 48-WTs through different POS. There is no splices exist in the fiber link in our calculation. Passive optical splitters with split ratio of (1×2) and (1×8) are used for WPF configuration. The OLT transmits its downstream data to WTs over single-mode fiber (1000BASE‑PX20‑D) which supports at least 20 Km with 1490nm, while each WT transmitted its upstream data to control center at 1310nm. Both upstream and downstream operate at 1.25Gbps [17][18].

A.   Optical Power Budget

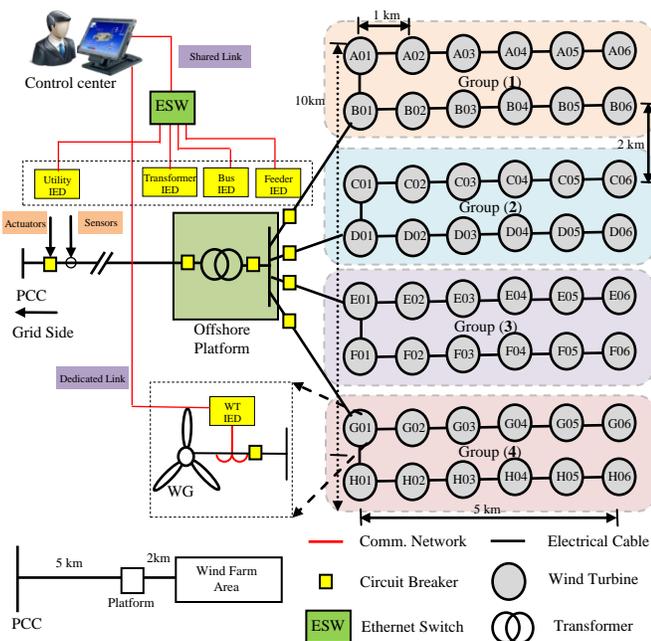In case of 1000Base-PX20, the power budget for EPON specified in IEEE 802.3ah is 26 dB, including channel insertion loss, dispersion and noise. Table 2 shows the EPON specification according to IEEE 802.3ah standard.

The optical budget [dB] in Equation (1) is the difference between the minimum transmitter launch power (P_{tx},dBm) at the input of the optical link, and the minimum sensitivity of the receiver (P_{rx},dBm) at the output of optical links.

$$\text{Power Budget} = P_{tx} - P_{rx} \qquad (1)$$

Along the communication link, there are many sources of attenuation including, connections, splitters and the fiber cable itself. The total insertion loss must be less than the value of power budget [19]. The total power loss (P_s) of passive components of optical link given in [dB] is calculated as shown in Equation (2). Note that, a safety margin due to aging of the Tx/Rx elements, and the effect of temperature should be considered in total optical power loss calculation.

$$P_s = IL_{cable}\ \text{Length} + IL_{conn}\ N_{conn} + IL_{split} \qquad (2)$$

Where:

IL_{cable} : Insertion loss for cable [dB/Km]
Length : length of cable from OLT to ONU
IL_{conn} : Insertion loss per connector [dB]
N_{conn} : number of connectors (optical connection pairs)
IL_{split} : Insertion loss for splitters [dB]

TABLE II.   IEEE STD 802.3 EPON SPECIFICATIONS [18][20]

Parameter	Data rate	Power budget
EPON	1.25Gb/s (D &U) 1000Base-PX20	PX-20U 26 dB PX-20D 26 dB
ExEPON	1.25Gb/s (D &U) 1000Base-PX30 1000Base-PX40	PX-30U 29dB PX-30D 29dB PX-40U 33dB PX-40D 33dB

TABLE III.   COMPONENT INSERTION LOSS[7]

	Fiber	Connector	Splitter		
Attenu-ation	0.4 dB/Km	0.2 dB	1×2 (5%:95%) 0.4 dB	1×2 (50%:50%) 3 dB	1×8 9 dB

We considered the fiber attenuation is 0.4dB/Km, and the connector loss is 0.2dB. Passive optical splitter (1×2) insertion loss is 0.4dB for split ratio of 5%:95%, while the insertion loss is 3dB for split ratio of 50%:50%, as shown in Table 3.

Figures (6, 7) show the total optical path loss calculation for 24 WTs of architectures (A&B) from WT-A01to WT-D06. The remaining WTs from WT25 to WT48 are not shown as the WPF is symmetrical with same dimensions and configurations. In Fig. 6, group (1) represents 12 WTs connected in radial topology. With 1Km between WTs, fiber attenuation is 0.4 dB, connector loss is 0.4 dB and POS has the insertion loss of 0.4 dB. This is reason why the total loss appears to be growing linearly. The highest optical path loss value represents the farthest turbine (A06), about 24.18 dB, while the lowest value represents the nearest turbine (D01), about 16.89 dB. Considering the IEEE 802.3 standard, architecture (A) satisfies the requirements. In Fig. 7, the highest value of optical path loss for (A06) is about 28.78 dB, while the lowest value for WT19 is about 25.49 dB.
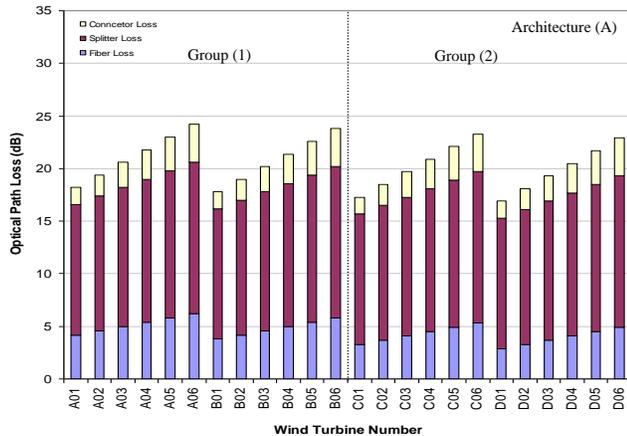
Figure 6. Total optical Path Loss of WPF- Architecture (A)
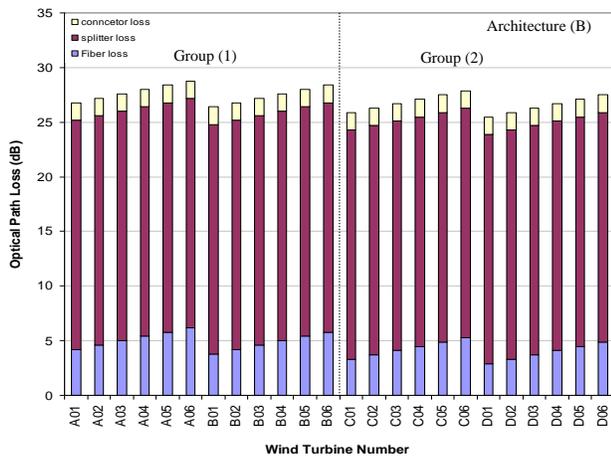
Figure 7. Total optical Path Loss of WPF- Architecture (B)

In architecture (B), the connector loss and POSs insertion loss are not changing according to distance: about 1.6 dB and 21 dB, respectively. Each WT has a dedicated fiber to the secondary POS with fiber attenuation loss of 0.4 dB/Km. The dominant part of path loss is related to primary POS with 1:8 split ratio. Thus, architecture (B) does not fulfill the required of power budget. If we consider the new power budget classes (PX-30, PX-40) for symmetric (1/1G) as shown in Table 2, both network architectures will fulfill the required power budget [20]. If we re-configured the network architecture by placing the OLT unit at offshore platform, and connect each WPF group using dedicated optical fibers (DF1→DF4). In this case, we could eliminate the path loss of the POS (1×8), which has a significant contribution to the path loss. For architecture (B), the highest value of path loss for (A06) will be 17.38 dB, while the lowest value for (D01) will be 14.1 dB. We believe that the industry will provide optical components with lower insertion loss which increases the launch power and improve the receiver sensitivity.

B. Network Cost

Capital expenditure (CAPEX) consists of both initial network equipments, and network installation costs [15]. In this work, the submarine cable comprises the electrical collection cables and optical fiber cables. Because the communication network part is integrated with the electrical system, we consider only the network equipment cost to evaluate and compare different network architectures. The network cost is calculated as follows:

$$C_{EPON} = C_{OLT} + C_{FF} + C_{POS} + C_{DF} + N_{ONU}.C_{ONU} \qquad (3)$$

where C_{OLT}, C_{POS} and C_{ONU} represent the equipment of OLT, POS and ONU, respectively. C_{FF} and C_{DF} represent the costs of optical fiber cable of feeder fiber (FF) and distributed fiber (DF), respectively. N_{ONU} represents the number of WTs-ONUs.

Fig. 8 shows the network cost for different WPF layouts and sizes: 12WTs, 24WTs and 48WTs. EPON-based architectures have the lowest cost compared with switched architectures. Architecture (A) is about 42,284 US $ for 48WTs, and about 59,084 US $ for architecture (B). Compared with switched-based architecture with consists of optical fiber cables and Ethernet switches, the network cost is 99,984 US $. The comparative cost analysis shows that EPON-based architectures are a cost effective solution for wind power farm.

TABLE IV. COMPONENT COST [21]

Components	Cost (US $)
OLT	12,100
ONU	350
Ethernet Switch	1,800
Splitter 1 × 2	50
Splitter 1 × 16	800
Fiber (/Km)	160
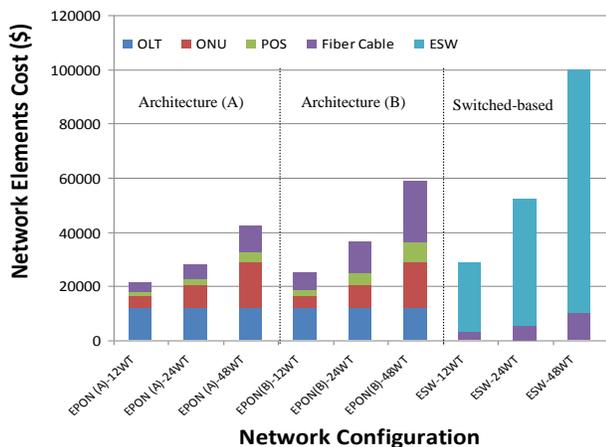
Figure 8.   Comparison of Network cost

VI.   CONCLUSION

In this paper, we proposed two different communication network architectures based on EPON technology for offshore WPF. In order to evaluate the proposed communication architectures, we consider offshore WPF with 48 WTs which is located 5Km from the shore. The total path loss is analyzed for each wind turbine to assure that optical power budget is sufficient for maintaining an acceptable performance under any conditions. Considering the new power budget classes (PX-30, PX-40), our proposed architectures are valid for the offshore WPF, and the power budget can be also satisfied. This analysis results showed the applicability and robustness of EPON-based communication network architectures that can be applied for next generation WPF. There are two directions of our future works. First is building a simulation model of EPON-based architectures to evaluate network performance such as reliability, packet delay and packet loss in both upstream and downstream direction, considering turbine monitoring and controlling data and a real WPF dimensions. Second direction is considering hybrid network architectures of EPON and IEEE 802.16 (WiMAX) for large-scale WPF. The wireless link of WiMAX can be incorporated into the WPF design as backup link to increase the network reliability.

ACKNOWLEDGMENT

REFERENCES

[1]   A. L. Pettener, "SCADA and communication network for large scale offshore wind power systems," IET Conf. on Renewable Power Generation, Edinburgh, pp. 1-6, 2011.

[2]   F. Bouhafs, M. Mackagy, and M. Merabti, "Links to the future: communication requirements and challenges in the smart grid," IEEE Power & Energy Magazine, Jan./Feb. 2012, vol. 10, no. 1, pp. 24-32.

[3]   O. Anaya-Lara, N. Jenkins, and J. R. McDonald, "Communications requirements and technology for wind farm operation and maintenance," Int. Conf. on Industrial and Information System, ICIIS, pp. 173-178, Aug. 2006.

[4]   K. Grobe and J.-P. Elbers, "PON in adolescence: from TDMA to WDM-PON," IEEE Communication Magazine, Jan. 2008, vol. 46, no. 1, pp. 26-34.

[5]   L. Zhihong, L. Zhiyong, X. Tao, and S. Liwei, "Exploration and practice of the smart distribution network by Yangiaping power supply Bureau," China Int. Conf. on Electricity Distribution, CICED, Nanjing, pp. 1-6, Sept. 2010.

[6]   H. Jiang, J. Yao, and G. Zhangand, "The construction scheme analysis of distant communication systems for subscriber power consumption information acquisition systems," Journal of Int. Council on Electrical Engineering, vol. 1, no. 4, pp. 482-485, 2011.

[7]   H. Huang and H. Zhang, "Application and analysis of long distance EPON in transmission lines monitoring system," Advanced Materials Research, vol. 317-319, pp. 1583-1589, 2011.

[8]   R. Carolsfeld, "Practical experience from design and implementation of IEC 61850 based communications network in large offshore wind installation," Tech. Meeting, CIGRE-AORC, Thailand, pp. 1-8, Oct. 2011.

[9]   W. Yong, W.-G. Xu, and Y.-L. Zhang, "Application analysis on EPON technology applied in communication system of smart substation," Lecture Notes in Electrical Engineering, vol. 154, pp. 1373-1379, 2012.

[10]   A. Ellis and E. Muljadi, "Wind power plant representation in large scale power flow simulation in WECC," IEEE Power & Energy Society Metting, Pittsburgh, pp. 1-6, Jul. 2008.

[11]   B. Badrzadeh et al., "Wind power plant SCADA and controls," IEEE Power and Energy Society General Metting, San Diego, pp. 1-7, Jul. 2011.

[12]   I. Lofgren, "Reference Architecture for Wind Power Integration, A Wind Power Plant System Structure based on Analysis of Wind Power Information Flow," Master Thesis, KTH Electrical Engineering, 2009.

[13]   M. Wei and Z. Chen, "An Ethernet LAN based distributed generation system load shedding strategy," IEEE Conf. on Industrial Electronics Society, Melbourne, pp. 819-824, Nov. 2011.

[14]   M. A. Ahmed and Y.-C. Kim, "Network modeling and simulation of wind power farm with switched gigabit Ethernet," Int. Symposium on Comm. and Information Technologies, Australia, pp. 1009-1014, Oct. 2012.

[15]   E. Eriksson, "Wind Farm Layout, A Reliability and Investment Analysis," Master Thesis in Energy Systems Engineering, Uppsala University, Sweden, 2008.

[16]   R. Maclean, "Electrical System Design for the Proposed One Gigawatt Beatrice Offshore Wind Farm," Master Thesis, University of Strathclyde, Sep. 2004.

[17]   G. Kramer, Ethernet Passive Optical Networks, McGraw-Hill, 2005.

[18]   IEEE Std 802.3ah, IEEE-SA Standard Board, Jun. 2004.

[19]   B. Ng, M. Syuhaimi, A. Premadi, and K. Jumari, "Optical power budget and cost Analysis in PON-based i-FTTH," Research Journal of Information Technology, vol. 2, pp. 127-138, 2010.

[20]   X. Zhang, Z. Fu, J. Xu, and M. Hajduczenia, "Extended EPON PMD specification," Extended EPON Task Force, Interim Meeting Material, Minneapolis, USA, May 2012.

[21]   L. Wosinska, J. Chen, and C. P. Larsen, "Fiber access networks: reliability analysis and Swedish broadband market," IEICE Transactions on Communications, vol. E92-B, no.10, pp. 3006-3014, Oct. 2009.

Planning of Sustainable Urban Districts based on Smart Micro-Grids Concept: A Case Study in Brazil

Cesare Pica, Tania Leites, Franco Tumelero

Sustainable Energy Center

CERTI Foundation

Florianópolis/SC, Brazil

e-mail: cqp@certi.org.br; tnl@certi.org.br, ftu@certi.org.br

Ricardo Trentin

Telecom Department

CELESC

Florianópolis/SC, Brazil

e-mail: ricardogt@celesc.com.br

Abstract — **The work presented in this paper is based on two converging trends worldwide: an increasing interest in sustainable solutions for urban districts and an increasing integration of distributed energy resource in the electric power system. The aim of this work is to analyze, adapt and combine existing methods used for power system planning in one easy-of-implement methodology suitable for rapid-planning of the deployment of distributed energy generation in urban districts. This paper also presents preliminary results of the application of the methodology for the case of a sustainable urban district in Brazil. The work is based on the concept of smart micro-grids, which are considered as building blocks of the "macro" smart grid and a potential solution for massive integration of distributed generation in the electric power system.**

Keywords – Micro-Grids; Sustainable Energy; Urban District

I. INTRODUCTION

As an important part of the Smart Grid deployment, Distributed Generation (DG) based on alternative energy resources and installed near the consumption has become increasingly interesting, from both technical and economic points of view. The benefits expected from DG applications include, for example, increase of power quality and energy supply reliability, reduction of power losses in transmission and distribution networks, reduction of greenhouse gas emissions, flexibility to choose between different energy sources (according with local environmental and commercial conditions), and creation of business opportunities for small and medium entrepreneurs. Due to these benefits, a massive insertion of DG connected to the power distribution network is expected in the near future.

On the one hand, this picture is very motivating; but, at the same time, it generates concerns with the possibility of a large amount of DG units, may cause negative impacts to the electrical system, as stability problems, protection, and technical and commercial relationship with a new kind of consumer /producer.

A promising solution to this concern arises with the concept of micro-grids. A micro-grid is defined as an integrated system, consisting of an interconnected set of distributed generators, different types of loads and smart grid technologies. In general, a micro-grid has one (or few) point of connection to the grid, and it is controlled by strategies of energy management that makes micro-grid introduce itself to a utility system as a single customer, consumer and producer, "well behaved".

Micro-grids are regarded as building blocks of the "Macro" Smart grid, enabling the deployment of smart grid gradually / modular (blocks), and promoting the coordinated integration of distributed generation without affecting the system and propose new commercial arrangements interest of the distributor and its customers. In Figure 1, the concept of micro-grid is represented differently, highlighting some potential applications, such as, commercial and industrial districts, campus and technology parks, urban districts and rural locations that require better security and energy efficiency, and quality sustainability in energy supply.
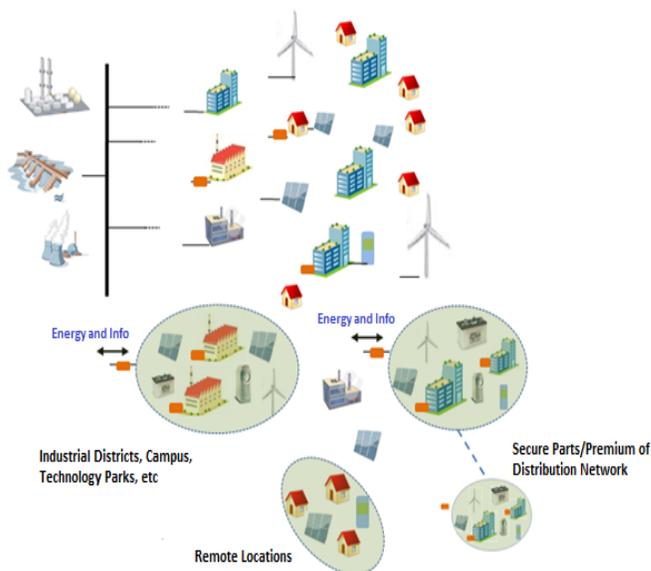

Figure 1. Examples of micro-grids applications.

Some technical characteristics of micro-grids:

- They have two main modes of operation: connected mode and islanded mode. In case of the island mode, the DG inside micro-grid should be able to maintain the local loads, mainly the critical load

- Enable better management of the integration of distributed generators (including those dispatchables, non-dispatchable, intermittent) and energy storage.
- Provide customers (consumer / producer) inside the micro-grid with improved energy service: better power quality, energy efficiency, optimizing the use of local energy resources, supply security.
- They enable the provision of ancillary services to the distribution network (main grid), while micro-grid is operating in connected mode.

It is estimated that by 2017, more than 4,7 GW of micro-grid will be operating in the world, with a total cumulative investment of over $ 17 billion [1].

There are already some known cases of worldwide applications of micro-grids, as: a campus in the city of Sendai (Japan); a pilot system installed in a laboratory of the Institute of Solar Energy Technology – ISET (Germany); a remote system placed in the north of British Columbia (Canada) and a neighborhood of vacation homes near to Amsterdam, in which homes are equipped with roof mounted photovoltaic systems (Holland). Since the results of these earlier application cases are showing the feasibility of micro-grids, the next step shall be the scale-up and increasing number of micro-grid installations as part of the electric power system.

A potential application of micro-grids should arise with the development of Sustainable Urban Districts (or Smart Cities/Communities), such as the solutions planned to be implemented in *Masdar City* (Arab Emirates) and *Tianjin Eco City* (China) [2]-[5]. In general the objective of a micro-grid in a Smart City/Community should be to locally provide energy with high level of power quality, energy efficiency and reliability, taking advantage of local energy resources (including solar, wind and urban residues) and, at the same time, promoting environmental and socio-economic benefits.

The work presented in this paper has been realized in this context of micro-grids deployment in Sustainable Urban Districts. Specifically, the urban district considered in this work, is an Innovation and Tech Park (named Sapiens Park) which nowadays is under construction in Florianópolis, southern Brazil. This district/park is located in an urban environment that requires more intelligent, practical and sustainable solutions, whether from the point of view of energy generation, power consumption, as well as the rational use of water, treatment and reuse of waste and local socioeconomic development.

The main objectives of this work are: (i) study of micro-grid solutions suitable for urban districts; (ii) analyze, adapt and combine existing methods and tools used for power system planning in one easy-of-implement methodology suitable for rapid-planning of the deployment of distributed energy generation in urban districts and (iii) pilot application of the methodologies, for the planning of a real micro-grid, to be implemented in Sapiens Park (Figure 2) as a model of sustainable urban district and living laboratory/showroom for demonstration and dissemination of innovative technologies.

The sections presented in this paper are organized according to the steps of the methodology used for planning micro-grids for sustainable urban districts. In Section II, methods and information about the characterization and growth projection of electrical loads are presented. Section III shows content regarding the identification and projection of energy resources available locally. In Section IV, a roadmap micro-grid technologies is briefly presented. In Section V, the paper presents preliminary results about the project of the micro-grid to be implemented in Sapiens Park. These results include simulations, analysis for the optimization of renewable energy generation and additional information about the urban district. Finally, Section VI presents conclusions and comments about future works.

Figure 2.   Sapiens Park Overview

II.   ELECTRICAL LOADS CHARACTERIZATION AND DEMAND GROWTH PROJECTION

The *Step 1* adopted for the planning of the micro-grid has been the characterization of local loads, presented in the urban district, and the projection of energy demand growth. The study of the demand growth should be realized for both systems: the urban district (where the micro-grid is intended to be implemented), but also for the region in which the district is inserted, in order to allow analysis about the impact that the micro-grid should have in the region and vice-versa. The following subsections present the methods for the study of both systems: the region (Method of Trends Extrapolation) and the urban district (Method of Land Use)

A.   Method of Trends Extrapolation

The method of trends extrapolation, selected in this work because its practical application, is based on a study of demand forecasting, conducted by Brazilian researchers [5]. Initially, it is necessary to collect a set of information, comprehending:

- Population growth in the region.
- Gross Domestic Product (GDP) in the region.
- Price index for consumer durables.
- Data regarding electricity rates.
- Per capita consumption of electricity in the region.

From this set of data (historical), it is possible then to use an econometric analysis tool. The result should be determining consumption and demand for electricity in a considerably large area, in which the urban district is inserted. In the case of the project, this region refers to the northen of the island of Florianópolis.

B. Method of Land Use

The Land Use Method is designed to regions of smaller geographical area (as the Sapiens Park – Sustainable Urban District considered as model of urban district in this work), allowing a more detailed energy consumption characterization and projection.

The method takes into account the rate of occupation of the land in relation to parameters such as, example, activity type of building, architectural and constructions characteristic, socioeconomic parameters, kind of installed loads, climatic characteristic, etc. A simplistic way of applying the Method of Land Use is based solely on energy intensity (kWh/m²), ie, the average consumption (kWh) per unit area (m²), which is dependent on the type of final energy use/building (ie. industrial plants, commercial buildings, restaurants, schools, hospitals, etc.).

In this work, the Energy Plus software has been used as platform for the study of load characterization and growth projection at Sapiens Park [6]. The study resulted in very detailed information, providing for example forecasts of energy demand by economic sector/activity and load curves during the period analyzed.

In Figure 3, daily power demand curves are show, considering the status of Sapiens Park in 2015 and 2030. In Figure 4, the share of different types of loads in Sapiens Park is presented.

Figure 4.   Share of different types of loads in Sapiens Park.

In Figure 5, it is shown the projected demand (peak demand) of both systems: the region (at the substation that supply energy to the region where Sapiens Park is located) and the urban district (Sapiens Park). Among other important conclusions resulted from this step of the methodology, it was possible to verify that in 2018, the demand of Sapiens Park added to the demand of the region will exceed the capacity of the existent substation that supply all the region. Therefore, it is worth to consider the application of a micro-grid with local DG in the Sapiens Park. The micro-grid would guarantee energy supply to the district, while at the same time it would allow utility postpone investments with additional infrastructure to increase the supply capacity.
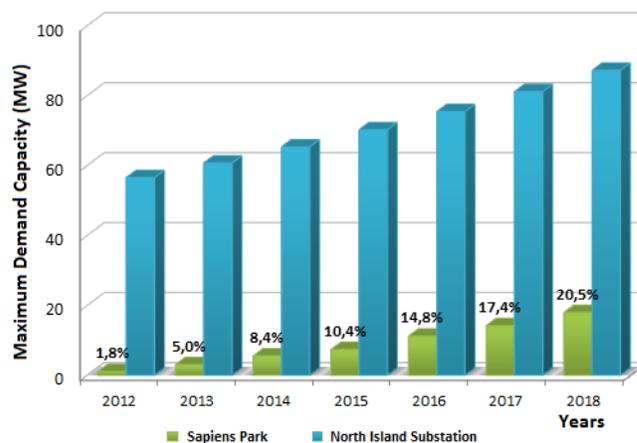
Figure 5.   Predicted Demand for North Island Substation and Sapiens Park over the years.

III.   IDENTIFICATION AND PROJECTION OF AVAILABLE ENERGY RESOURCES

The *Step 2* of the methodology for micro-grid planning aims to identify the available resources that could be used for energy generation. The identification is firstly realized by collecting information about the energy sources available locally, for example: wind (speed, direction, etc.), solar

Figure 3.   Predicted demand for Sapiens Park on 2015 and 2030.

(radiation, daily curve, climate, etc.), urban residues (amount of waste and sewage, forecasts, population, etc.).

In order to quantify the energy resources available in the urban districts, it is necessary to make use of local weather stations (for solar and wind resources), and a survey of water quantity and waste production (for questions related to sewage and solid waste, respectively).

Once the potential of the energy resources were calculated (in terms of radiation/m2, wind speed and direction, etc.) the next task is to determine the potential in terms of power and energy production (watts and watts per hour, respectively). This task can be accomplished by calculating power/energy production, considering the utilization of commercially available technologies for electricity generation [2]-[10]. Besides the type of technologies, it is important also to consider other particular features of the local. For instance, the micro-grid at Sapiens Park should respect the park constructive issues, such as the height limit of 25 meters to the buildings in the region, which prevents the use of large wind turbines.

For the conditions of Sapiens Park the following technologies have been chosen for the determination of the potential of local generation: polycrystalline silicon modules, vertical axis wind turbines and generation systems for urban solid waste and sewage, composed by digesters and generator group. Table 1 presents the potential for the electricity generation of Sapiens Park as well as indicators to compare the technical feasibility of using different sources of renewable energy in the region.

TABLE I.   RESUME OF ANNUAL ENERGY RESOURCES POTENCIAL IN SAPIENS PARK

Generation Type		Maximum Capacity [MW]	Amount of Energy Generated [MWh]		Capacity Factor[a]	Productivity[b]
Wind		19.98	26.838		0,09	1344
Solar	Roofs	0.31	432	0,29%	0,16	1395
	Lakes	7.71	10.749	7,13%		
	Parking	5.24	7.303	4,84%		
	Common Area	94.83	132.298	87,74%		
Urban Solid Waste		0.02	125		0,66	5808
Urban Sewage		0.33	1.897		0,66	5808

a. The Capacity Factor (FC) reflects the annual capacity of a system to produce energy has operated at their rated power for 24 hours per day.
b. The Productivity reflects how many energy can be produce for each kW installed.

From Table 1 it is possible to realize that Sapiens Park has a great potential of local energy generation, highlighting the numbers found for the potential of photovoltaic systems.

IV. TECHNOLOGIES ROADMAP

Based on the information of the *Step 1* (characterization of energy demand) and *Step 2* (identification of energy resources), presented respectively in Sections II and III, the aim of the *Step 3* described in this Section is to analyze different types of technologies that could be used for the implementation of the micro-grid in the urban district. This Section evaluate existing solutions and the solutions to come, based on trends of cost reduction, efficiency improvement, market acceptance, life cycle, etc. The consolidation of the knowledge gained in this Step can be summarized into three themes:

- Distributed Generation based on Renewable Energy
- Solutions for Smart Grids
- Storage Devices and New Technologies.
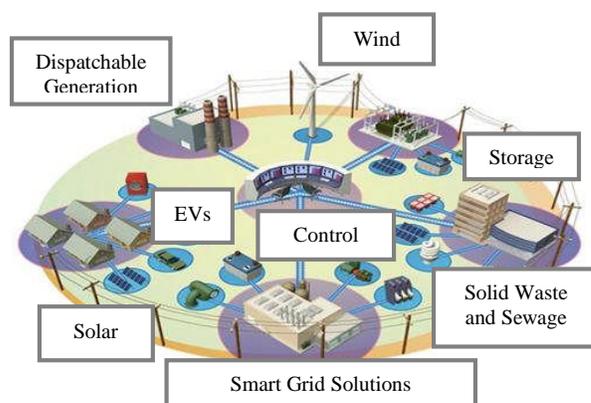
Figure 6.   Overview of important themes presents in micro-grids.

TABLE II. TRENDS IN DISTYRIBUTED GENERATION BASED ON RENEWABLE ENERGT.

Photovoltaic Energy	• Cost savings of up to 60% in 2030 compared to the current scenario. • Efficiency increasing of thin film technologies, CIS / CIGS, and monocrystalline silicon.
Wind Energy	• Improvement in the manufacturing process of wind turbines with increased lifespan and capacity factor. • New solutions for harnessing wind power through mini and micro wind turbines, with emphasis on the vertical axis wind systems.
Energy harvesting from Solid Urban Waste and Sewage	• Use of decentralized biodigesters in urban districts for the available volumes of sewage and organic part of municipal solid waste. • More feasible solutions for gasification and other advanced technologies.

TABLE III. TRENDS IN SOLUTIONS FOR SMART GRIDS

Electric System Topologies	• Improved reliability of energy supply should be achieved through more flexible topologies, including mesh networks, and new approaches to energy distribution, as in the case of Microgrids.
AC and CC Power Systems	• Increase the use of energy directly into direct current, which points to a trend of DC power distribution.
Aerial and Underground Distributions Systems	• Increased adoption of underground networks, due to the context of urban forestry and visual pollution. Note to reduced costs of these systems, and already existent technologies allow easy control and maintenance.
Protection	• More autonomy and intelligence with application of so-called "adaptive protections," which fit the real-time operation of electrical systems.
Communication Infrastructure	• Implementation of an effective communication and advanced metering infrastructure. • Complete replacement of all legacy infrastructure for measuring smart meters, multi meters in the same residence given greater "energy education", and demand response features. • Adoption of media and protocols that ensure interoperability between systems, including IEC61850 LV networks.
Control and Automation	• Implementation of solutions for demand response and demand side management • Distributed control and dispatch centers distributed generation. • Trend of use of Multi-agents systems for distributed control • Dissemination of advanced distribution automation.

TABLE IV. TRENDS IN STORAGE DEVICES.

Batteries	• Greater use of sodium-sulfur batteries (NaS) batteries and lithium in solutions integrated with the electrical grid, as well as in stationary electric vehicles.
Flywheels	• It is expected that the market begin to see the advantages of flywheels and provide greater feasibility for the technology, which provides high initial cost but a low cost of maintenance and ability to operate in very adverse conditions.
Supercapacitors	• Reduced costs will facilitate the application of supercapacitors in solutions in which they want high power density with fast dynamics

V. ANALYSIS AND DESIGN OF THE MICRO-GRID FOR THE SUSTAINABLE URBAN DISTRICT

This Section refers to *Step 4* of the methodology for micro-grid planning. It starts with the creation of possible scenarios for micro-grid implementation, defining, for example, the amount of renewable energy it is desired to generate locally at the urban district, the price of energy negotiated with the main utility etc. In the pilot application in Sapiens Park, the scenarios are related to the installed capacity of each source, and are established for short, medium and long terms (respectively 2015, 2020 and 2030). For 2015, the scenarios consider to attend 10%, 20% or 30% of energy consumption. But in 2020 the initial scenario becomes with the optimal solution in 2015 and the others are kept and so on (Figure 7).
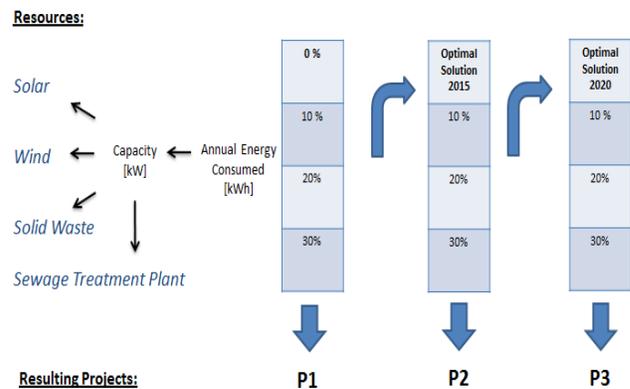
Figure 7.   Solution Problem Description.

Based on the proposed scenarios, the next task is to specify the technologies to be used (supported by the roadmap created in Step 3) and other technical and economic parameters, and then simulate the system in order to find the best solution for the micro-grid deployment (ie, mix of energy resources, technologies to be adopted, amount of local generation, commercial relationship with the main utility, etc.). In this work, the software Homer Energy has been used. The software Homer Energy [17] seeks to assist in determining the more viable technically and economically solution to attend the local loads in the years 2015 to 2030.

With the results of the consumption and demand for electricity over the years compared with the technology roadmap in terms of efficiency and cost of systems for electricity generation, it is possible to create enhanced simulation scenarios. One of the results obtained with Homer Energy can be seen in Figure 8, where the optimal solution is identified in 2015. It is possible observe the presence of main grid and photovoltaic systems in the optimal solution. The system in question has an installed capacity of 170kW and has a share of 10% of all energy produced in that year.

Figure 8.   Homer Energy window with the viable solutions based on Total NPC

In Figure 6, the first line show the optimal solution based on the less Total Present Cost of Micro-grid (Total NPC) that is R$53.871,788 for proposed problem. Besides the costs related to the power grid and renewable sources, this cost are composed of annual real interest, project life, cost of the system and another economics variable. The best way to represent the distribution cost of this project is shown in Figure 9 that related Capital, Replacement, Savage and Operating costs over the years.

It is important to emphasize that Homer Energy has a lot of ways to present the results, including reports generation, comparing the investing in micro-grids with the benefits to stakeholders.

As other years are simulated, other generation sources can be found most viable than those chosen for 2015 simulation. This work is currently in time to implement scenarios adjustment in order to allow further studies and the creation of a computational platform for investment

evaluation and quantification of benefits (to the local distributor, urban district and investors), so that it can be replicated to any urban district to submit interest in making use of renewable sources for electricity generation.
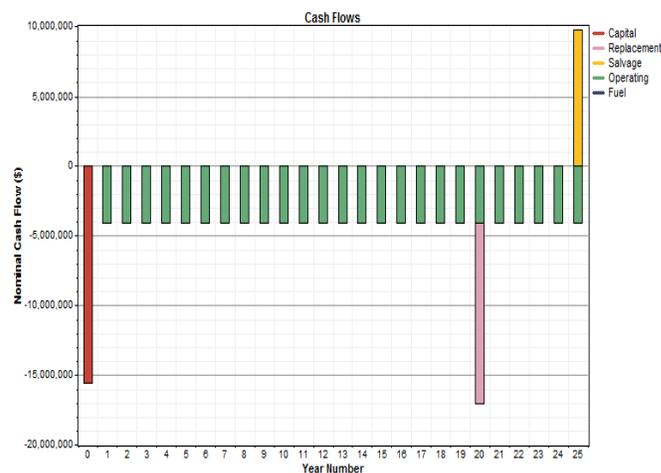
Figure 9.   Homer Energy window with the Cash Flow based on optimal solution considering the Total NPC

VI.   CONCLUSION AND FUTURE WORK

This paper presented a methodology for planning the deployment of micro-grids in sustainable urban districts. Also, the work includes a pilot application of the methodology, considering an urban district located at Florianópolis, southern Brazil.

The methodology has been subdivided in four main Steps: (1) characterization and growth projection of electrical loads, (2) identification and projection of available energy resources, (3) technology roadmap and (4) design and simulations of technical and economic aspects of the chosen solutions for the micro-grid. The *Step 4* is currently under development, but earlier results of the pilot application have shown that the methodology and the chosen computational tools are worthy for the objective of sustainable urban district planning, and could be applied in the context of other applications of micro-grid (industrial and commercial, districts, institutional campus, others) .

Parallel to this work, a project to implementing a pilot plant composed by renewable energy resources is currently under development. The pilot plant will include photovoltaic panels, small wind system, microturbine and batteries, totalizing approximately 100kW. The main objective of this new project is to develop, implement and evaluate control strategies for the optimal use of renewable resources and supply of priority loads in Sapiens Park.

REFERENCES

[1] Pike Research, 2012 [retrieved: March, 2013]. Available: http://www.pikeresearch.com/newsroom/worldwide-revenue-from-microgrids-will-reach-17-3-billion-by-2017

[2] M. Barnes, J. Kondoh, H. Asano, J. Oyarzabal, G. Ventakaramanan, R. Lasseter, N. Hatziargyriou, and T. Green, "Real-World MicroGrids-An Overview", in System of Systems Engineering, SoSE 2007. IEEE International Conference on, pp. 1-8.

[3] N. Hatziargyriou, H. Asano, R. Iravani, and C. Marnay, "Microgrids: An Overview of Ongoing Research, Development, and Demonstration Projects", Power and Energy Magazine, IEEE DOI - 10.1109/MPAE.2007.376583, vol. 5, no. 4, pp. 78-94, 2007.

[4] Masdar City, 2011; Abu Dhabi [retrieved: November, 2012]. Available: http://www.masdar.ae/en/home/index.aspx

[5] Eco-City, 2011; China. [retrieved: November, 2012]. Available: http://www.tianjinecocity.gov.sg/

[6] M. L. Siqueira and H. Holanda. "A Demanda Setorial por Energia Elétrica no Pós-Racionamento de 2001: Previsões de Longo Prazo para o Nordeste Brasileiro", X Encontro Regional de Economia, Fortaleza, Brazil, 2005.

[7] NREL, "Commercial Reference Building Models of the National Building Stock", United States, February, 2011.

[8] C. Ronaldo dos Santos. "Wind Energy to Produce Electricity", Eletrobrás, Rio de Janeiro, 2009.

[9] A. Ricardo, "Solar Energy" Published by: Artliber.São Paulo, 2002.

[10] Comcap, "Physical Characterization of Municipal Solid Waste in Florianópolis", Final Report . Florianópolis 2002.

[11] S. P. Chowdhury, P. Crossley, and S. Chowdhury, "Microgrids and Active Distribution Networks". Institution of Engineering and Technology, 2009.

[12] G. D. N Veslasco, A. M. L. P. Lima, and H.T. Z. do Couto . "Analysis of the comparative costs of different networks of power distribution in the context of urban forestry". Forestry Research Society, 2006.

[13] EWEA, 2011 [retrieved: November, 2012]. Available: http://www.ewea.org/fileadmin/ewea_documents/documents/publications/statistics/Stats_2011.pdf

[14] A. A. Bittencourt. Dissertation in electrical engineering entitled "Protection of feeders adaptive power distribution considering distributed generation". Federal University of Santa Catarina, 2011.

[15] Y. Yan, Y. Qian, H. Sharif, and D. Tipper. "A Survey on Smart Grid Communication Infrastructures: Motivations, Requirements and Challenges". *Communications Surveys & Tutorials*, IEEE PP, no. 99 (2012): pp. 1–16.

[16] D. Connolly, "A Review of Energy Storage Technologies for the integration of fluctuating renewable energy", University of Limerick, 2010.

[17] F. A. Farret and M. G. Simões, "Integration of Alternative Sources of Energy", Published by John Wiley & Sons, Inc., 2006.

Introducing Energy Efficiency into SQALE

Luca Ardito, Giuseppe Procaccianti, Antonio Vetro', Maurizio Morisio

Dipartimento di Automatica ed Informatica

Politecnico di Torino

Torino, Italy

E-mail: name.surname@polito.it

Abstract—Energy Efficiency is becoming a key factor in software development, given the sharp growth of IT systems and their impact on worldwide energy consumption. We do believe that a quality process infrastructure should be able to consider the Energy Efficiency of a system since its early development: for this reason we propose to introduce Energy Efficiency into the existing quality models. We selected the SQALE model and we tailored it inserting Energy Efficiency as a sub-characteristic of efficiency. We also propose a set of six source code specific requirements for the Java language starting from guidelines currently suggested in the literature. We experienced two major challenges: the identification of measurable, automatically detectable requirements, and the lack of empirical validation on the guidelines currently present in the literature and in the industrial state of the practice as well. We describe an experiment plan to validate the six requirements and evaluate the impact of their violation on Energy Efficiency, which has been partially proved by preliminary results on C code. Having Energy Efficiency in a quality model and well verified code requirements to measure it, will enable a quality process that precisely assesses and monitors the impact of software on energy consumption.

Keywords-Energy Efficiency; energy-aware software; SQALE

I. INTRODUCTION

The rapid growth and significant development of Information Technology (IT) systems has started to cause an increase of worldwide energy consumption [1]. This issue moved technology producers, information systems managers, and researchers to deal with energy consumption reduction [2]. For this reason, research has increasingly focused on improving the Energy Efficiency of hardware, but the literature still lacks in quantifying accurately the energy impact of software. Software does not consume energy directly, however it has a direct influence on the energy consumption of the hardware underneath. As a matter of fact applications and operating systems indicate how the information is processed and, consequently, drive the hardware behaviour. Considering each IT device, it has its own theoretical energy consumption, which can range from 0, when it is turned off, to x if all its internal components are used simultaneously. Through the management of each part there is a variation Δx of its consumption that is between 0 and x. The management of system components can be done either in hardware or software. Previous work [3] suggested
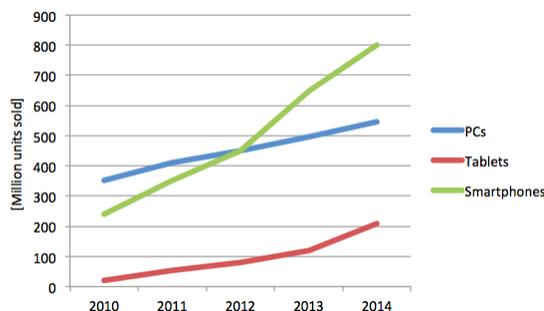
Figure 1. IT devices sale forecasts

that software can reach up to 10% of the total system power (measured as the difference between an idle activity, used as a baseline, and the most power-consuming scenario). This figures ought to be taken into account especially when considering mobile environments and data centers. Mobile handsets sales are increasing sharply [4] (see Fig. 1) and this class of devices have to deal with battery-related issues, so energy savings can impact significantly on the device autonomy. On the other hand, small energy reductions in data centers can result in big energy savings: for example, just in 2009, data centers consumed about 330 TWh [5].

Having regard to the influence of software in energy consumption, it is necessary to quantify the Energy Efficiency of source code. For this reason, we envision a software quality model that includes Energy Efficiency in order to take it into account as a key aspect during the software development and utilization. Having in mind this scenario, we suggest the usage of the Software Quality Assessment based on Lifecycle Expectations (SQALE) model [6] to include Energy Efficiency as a measurable quality attribute. This paper is organized as follows. Section II describes SQALE methodology. Section III discusses the adaptation of SQALE quality model to include Energy Efficiency. Section IV introduces a software framework for Requirements Empirical Validation. Section V presents our conclusions and future work.

II. SQALE

SQALE [6] is a methodology to support the evaluation of the software quality. It is applicable to any software artifact (such as code, UML models, documentation, and so on),
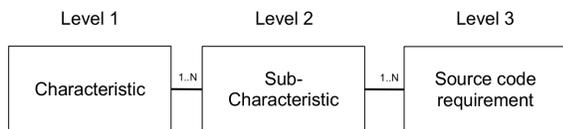
Figure 2.   Hierarchical quality model structure

Characteristic	Sub-characteristic	Generic Requirement Description
Maintainability	Understandability	File comment ratio > 35%
Maintainability	Readability	File size (LOC) < 1000
Maintainability	Readability	No commented-out code
Efficiency	RAM related efficiency	Class depth of inheritance (DIT) < 8
Efficiency	RAM related efficiency	No unused variables, parameter or constant in code
Changeability	Logic related changeability	If, else, for, while structures are bound by scope
Reliability	Fault tolerance	Switch statements must have a default condition
Reliability	Data related reliability	No use of unitialized variables
Testability	Integration level testability	Coupling between objects (CBO) < 7
Testability	Unit Testing testability	No duplicate part over 100 token
Testability	Unit Testing testability	Number of parameters in a module call (NOP) < 6

however the main focus is on source code, whose quality is perceived as a non functional requirement. The goal of using SQALE is to quantitatively assess the distance between the code current quality and its expected quality objective. To achieve that, the following main concepts are introduced:

1) A quality model
2) An analysis model
3) Indices and Indicators

A. Quality model

The quality model proposed by SQALE is an orthogonal quality model derived from the ISO/IEC 9126 [7] (revised by the ISO/IEC 25010 [8]) . It is organised in three hierarchical levels, which are represented in Fig. 2. The first level is composed of characteristics that are based on the theoretical lifecycle of a source file and are from the ISO 9126 standard. They depend on the code internal properties and directly impact the typical activities of a software application's lifecycle. Characteristics are listed in the order they appear in a typical software application lifecycle: Testability, Reliability, Changeability, Efficiency, Security, Maintainability, Portability, Reusability.

The second level is composed of sub-characteristics, based on sub activities and requirements domain. There are two types of sub-characteristics: those corresponding to lifecycle activities (e.g., unit test, integration test), and those resulting from taxonomies in terms of good and bad practices relating to the software's architecture and coding. A sub-characteristic is attached to only one characteristic, the first in the chronology of the characteristics (to preserve orthogonality). The third level is composed of requirements that relate to the source code's internal attributes. These requirements usually depend on the software's context and language, and they are also attached to the lowest possible level, i.e. in relation to the first quality characteristic to which it chronologically contributes. In this way orthogonality is preserved also at the bottom level. Requirements relate to the artifacts that compose the software's source code, e.g. software applications, components, files, classes, and so forth. TABLE I is excerpt from the SQALE standard [6] and it contains examples of how requirements in the Java language are inserted in the structure of characteristics and sub-characteristics. Fig. 3 represents graphically the hierarchy.

B. Analysis model

The SQALE Analysis Model contains the rules to normalize and control measures relating to the code. For each

Figure 3.   Hierarchical representation of the model described in TABLE I

violated source code requirement, a remediation cost (a work unit, a monetary unit, or a time unit) is associated to make the code conformant to the quality objective. For instance, looking at TABLE I and Fig. 3, the remediation cost for the requirement "Coupling between objects (CBO) < 7" is the cost to decrease the coupling from its current value X to 7. The remediation cost might not be constant but expressed by a remediation function. For example, reducing the coupling from 13 to 7 (-6) might cost more than twofold effort of reducing it from 10 to 7 (-3). The total refactoring cost for a sub-characteristic is the sum of the remediation cost of each requirement violation.

C. Indices and Indicators

All of the SQALE indices represent the costs related to a given characteristic, estimated by adding up all of the remediation costs of the requirement violations of the connected sub-characteristics. For instance, in the example of TABLE I, the SQALE Testability Index STI is the sum of the remediation costs of all violated requirements related to Integration level testability and Unit testing testability, i.e. "No duplicate part over 100 token", "Number of parameters in a module call < 6", "Coupling between objects < 7". The sum of all indexes is the Software Quality Index (SQI). It is also possible to obtain consolidated indices in the following way: the consolidated index of a given characteristic is equal to the sum of all the indices of the previous characteristics. For instance, the SQALE Consolidated Reliability Index (SCRI) is equal to STI + SRI, i.e. the sum of Testability and Reliability indexes. Moreover, a density index has to be associated with each absolute index dividing it by a measure representing the size of the artifact (lines of code, complexity, etc.). Finally the SQALE Method defines several synthesised indicators to summarize the overall quality status of the application: since indicators are out of the scope of the current work, we point the reader to the SQALE document to more detailed information.

III. TAILORING SQALE QUALITY MODEL TO INCLUDE ENERGY EFFICIENCY

We propose to tailor the SQALE model to include Energy Efficiency. As stated in the introduction, including Energy Efficiency in a quality model is an important step towards a measurable, repeatable and objective way to evaluate and improve the Energy Efficiency of a given application. We suggest to introduce Energy Efficiency as a sub-characteristic related to the main characteristic "Efficiency". It cannot be a characteristic itself, because it is not an activity in the typical software lifecycle, but it is a sub-characteristic of second type (i.e., "a recognised taxonomy in terms of good and bad practices relating to the software's architecture and coding"). The next step is to identify appropriate code requirements that can be applied to evaluate the Energy Efficiency of a software product. For this purpose, in this section we propose a list of guidelines, derived from the literature [9] [10] [11], provided as solutions to developers in order to produce energy-efficient software. From these guidelines, we extract a set of proper requirements to be included in the SQALE Quality Model. Most of the guidelines suggested in the literature are not strictly code-related, but rather recommending general programming techniques. As a consequence, we selected from the original list those requirements that can be traced to actual code structures. These are listed in TABLE II.

Starting from the selected guidelines, we express the set of requirements for evaluating software Energy Efficiency.

TABLE II
GUIDELINES THAT CAN BE TRANSLATED INTO SQALE REQUIREMENTS

Nr.	Guideline	Explanation
GD1	Decrease algorithm complexity	Despite different algorithms can complete the same task, the way the task is performed can be totally different. Reducing the algorithm complexity can lead to save energy.
GD2	Use Event-Based programming	Event based programming avoids a waste of resources involved in doing unnecessary operations. If polling cannot be avoided, it is advised to select a fair time interval.
GD3	Batch I/O	Buffering I/O operations increases Energy Efficiency; the OS can power down I/O devices when not used.
GD4	Reduce data redundancy	Storage and transportation of redundant data impacts Energy Efficiency
GD5	Reduce memory leaks	With memory leaks the application can stall or crash. This unpredictable behavior can alter the energy consumption and, more generally, they must always be avoided.

These requirements, as specified in the SQALE Model Definition Document, [6], must be:

- Atomic
- Unambiguous
- Non-redundant
- Justifiable
- Acceptable
- Implementable
- Not in contradiction with any other requirement
- Verifiable

Our approach, in coherence with the SQALE methodology, is based upon translating the guidelines into code patterns automatically detectable with static analysis tools. We propose an estimate, basing upon the presence of particular implementations that may cause energy waste. Since requirements are meant by SQALE to be language-dependant, we use the Java language as a reference in this paper. TABLE III contains the requirements identified and mapped to the guidelines they derive from. This is not to be intended as an exhaustive list but a first step towards source code Energy Efficiency quantification.

A. RQ1: Halstead's Effort < K

Halstead's Effort [12] is a technique for describing the structural properties of algorithms. This metric has been selected because it gives an estimation of algorithm complexity, which is language-dependant, but not implementation-dependant as other metrics commonly used in this field (such as McCabe's Complexity [13]). K is a parameter to be defined according to specific application domains and project characteristics.

TABLE III
REQUIREMENTS FOR ENERGY EFFICIENCY

Nr.	Guideline	Nr.	Requirement
GD1	Decrease algorithm complexity	RQ1	Halstead's Effort$< K$
GD2	Use Event-Based programming	RQ2	Nr. of polling cycles=0
GD3	Batch I/O	RQ3	Nr. of FileInput-Stream.read() method calls = 0 [8]
GD4	Reduce data redundancy	RQ4	Nr. of unused variables = 0
GD5	Reduce memory leaks	RQ5.1	Nr. of Dead Store issues per class = 0
		RQ5.2	Nr. of String|Boolean|Integer|Double constructor = 0

B. RQ2: Nr. of polling cycles = 0

To date and up to our knowledge, no static analysis tool is able to detect polling cycle, because polling structures can be implemented in various ways. However, we decided to keep this requirement and to devote further work to find a relevant metric to detect polling.

C. RQ3: Nr. of FileInputStream.read() method calls = 0

This requirement derives from a particular issue regarding the FileInputStream.read() method, that triggers a direct call to the underlying OS. If inserted into a cycle, it will realize an inefficient I/O policy. The use of a BufferedReader greatly improves performance and supposedly Energy Efficiency of the operation [14]. For example, the code shown in Listing 1 continuously calls the read() method of a FileInputStream object, thus triggering a large number of RPC calls to the Operating System.

```
FileInputStream fis = new FileInputStream(filename]);
int cnt = 0;
int b;
while ((b = fis.read()) != −1)
{
        if (b == '\n')
                cnt++;
}
fis.close();
```

Listing 1.   Example of inefficient I/O policy

The code shown in Listing 2 makes use of a Buffered-InputStream, which reads larger chunks of data than the FileInputStream. This greatly reduces Remote Procedure Calls, which improves Energy Efficiency by allowing the Operative System (OS) to turn off the I/O device when not needed.

D. RQ4: Nr. of unused variables = 0

The code in Listing 3 shows an example of Unused Field issue: the AClass contains a private field named "b", which

```
FileInputStream fis = new FileInputStream(filename);
BufferedInputStream bis = new BufferedInputStream(fis);
int cnt = 0;
int b;
while ((b = bis.read()) != −1)
{
        if (b == '\n')
                cnt++;
}
bis.close();
```

Listing 2.   Example of efficient I/O policy

is never used (the class does not provide a get() method for that field).

```
private class AClass
{
        int a;
        int b;

        public int getA(){return a;}
}
```

Listing 3.   Example of Unused Field

An optimization of this code would be providing a get() method for the "b" field, or removing the field if unnecessary.

E. RQ5.1: Nr. of Dead Store issues = 0

The code shown in Listing 4 contains a Dead Store issue, which means assigning a value to a local variable which is not read by any subsequent instruction.

```
public int DeadLocalStore(int x)
{
        int constant_a = x;
        constant_a = 3

        return constant_a + x;
}
```

Listing 4.   Example of Dead Local Store

In the code above, x is stored to *constant_a* but it is overwritten in the subsequent code line. A more efficient code is shown in Listing 5.

```
public int noDeadLocalStore(int x)
{
        int constant_a = 3;
        return constant_a + x;
}
```

Listing 5.   Example of refactored Dead Local Store

The value of x is no longer stored to *constant_a* and then replaced.

The requirements specified above are derived from the guidelines [11] [10] of good programming practices provided in the literature. However, it is worth mentioning that

such guidelines, despite being intuitive and acknowledged as effective by software industry specialists [1], did not receive any empirical validation. For this reason, and in order to make the choice of the above specified requirements *justifiable*, an empirical validation that quantitatively assess their impact on Energy Efficiency is needed.

The last steps in the introduction of Energy Efficiency into SQALE are: tailoring the analysis model, tailoring the indices and the indicators. These steps are more straightforward then the previous one. Regarding the analysis model, a remediation function for each requirement violation ought to be defined in order to obtain the remediation cost. An example of remediation function for the requirement

$$\text{number of dead stores} > 0$$

could be:

$$10 + 2x$$

where x is the number of deadstores, 10 is the cost (in time units) of running a static analysis tool to detect them and 2 is the estimated time cost to review and refactor each dead store. We plan to estimate the remediation cost of each requirement violation through controlled experiments (e.g. observing the time required by subjects for a refactoring action) and questionnaires (i.e. asking directly to practicioners for estimations of refactoring actions). Being Energy Efficiency a sub-characteristic of efficiency, the sum of the remediation costs of all its source requirements will be added to the total cost of the other efficiency sub-characteristics, obtaining the SQALE Efficiency Index. Finally, the indicators do not need tailoring because they are at the highest level of the quality model.

IV. REQUIREMENTS EMPIRICAL VALIDATION: EXPERIMENT PLANNING AND PRELIMINARY RESULTS

As said in the previous section, the guidelines we propose are not supported by an empirical evidence regarding their impact on Energy Efficiency. We plan an experiment to test whether the requirements identified in TABLE II have a measurable impact on energy consumption. We use as example the RQ 5.1 (dead stores) to explain the experiment framework. In our experimentation, we set up two source code fragments: one containing a dead store to an Integer variable and the corresponding refactored version, functionally identical but without the dead store. The instrumentation required for our experiment is a system to execute the code, meanwhile logging power consumption data through a Data Acquisition Board. On the software level, the infrastructure implementing the experiment is inspired by the JUnit2 [15] framework for automated software It consists of an abstract class, Experiment, that can be extended by concrete experimental classes. Each experimental class must provide two methods performWithViolation() and performWithoutViolation() that contain respectively the code including the

violation and with the violation refactored out. In addition the method setUp() may be optionally redefined to prepare for the execution. For instance, the experimental class for the integer dead store explained in the previous section is shown in Listing 6.

```java
public class DLS_DEAD_LOCAL_STORE extends Experiment
{
        public int performWithViolation()
        {
                int constant_a = x;
                constant_a = 3;
                return constant_a + x;
        }

        public int performWithoutViolation()
        {
                int constant_a = 3;
                return constant_a + x;
        }
}
```

Listing 6.   Dead Local Store experimental class

The power consumption of the methods performWithViolation() and performWithoutViolation() are expected to be very small. Unfortunately the standard measurement methods are not able to record precisely power at order of magnitude below microWatts. For this reason, the execution of each method is repeated consecutively a very high number of times (e.g. 1 million) to consume a measurable amount of power. We assume that each execution of the measured methods is independent on each other. This is true if no attribute is used except those initialized in the setUp() method.

Moreover, the framework provides the method: perform(int nSamples , long nIter). It takes as parameters two integers: the number of measurement samples to be generated (nSamples, set to 100 by default), and the number of iterations of the perform methods (nIter, set to 1 million by default). At the end of the experiment we will have nSamples samples, each of them representing the execution times of nIter iterations of both perform methods. We also plan to have a batch of different runs of the basic experiment carried on at different random times during the day to compensate the possible confounding effect of periodical tasks performed by the operating system. Finally, having the power consumption data of the two methods, it is possible to compare them and assess the possible impact of the requirement violation on Energy Efficiency.

To date, we have finished the instrumentation of a Desktop machine where to run the experiment, by means of an electrical power meter [16], through which we will log power consumption data during the execution of software specifically written for our experimental purposes. We previously conducted a similar experiment on an embedded system with a integer dead store implemented in the C

language. Our results show the impact related to the dead store was in the order of 20-40 picowatts per instruction. Such a small power saving could be higher if code patterns are executed thousands or millions of time, as it might happen in loops. Moreover, considering devices running on batteries, dead store have a negative impact on battery consumption.

V. CONCLUSIONS AND FUTURE WORK

Energy efficiency is becoming a key factor in software development, given the ubiquity of software in everyday life and its hardware-related power consumption. Moreover, in devices running on batteries, efficient energy consumption is a key aspect. For this reason we propose to introduce Energy Efficiency into the existing quality models. We selected SQALE, whose quality model is derived from the ISO/IEC 9126 and it is strictly related to the software lifecycle activities. We tailor SQALE inserting Energy Efficiency as a sub-characteristics of efficiency, and we propose a set of specific requirements for the Java language starting from guidelines currently developed in the literature. The requirements identified are:

- Halstead's Effort $< K$
- Nr. of polling cycles = 0
- Nr. of FileInputStream.read() method calls = 0
- Nr. of dead store issues per class = 0
- Nr. of unread variables = 0
- Nr. of String|Boolean|Integer|Double constructor = 0

We identified two major challenges in requirements elicitation:

1) the translation of the guidelines in measurable requirements, whose violations are automatically identifiable by tools;
2) the validation of the negative impact of the requirements violation on energy consumption.

We are planning an experiment to empirically verify the impact of requirements on Energy Efficiency and we presented the results of a preliminary work for an integer dead store implemented in the C language, where we verified that it actually causes an increase of power consumption per instruction. Future work will be devoted to execute the experiment to empirically validate the requirements, estimating both their negative impact on power consumption and the related remediation costs. We will also investigate whether other requirements are eligible to be included in the quality model under Energy Efficiency sub-characteristic.

REFERENCES

[1] The Climate Group, "Smart 2020: Enabling the low carbon economy in the information age," GeSi, Tech. Rep., 2008.

[2] A. Berl et al., "Energy-efficient cloud computing," The Computer Journal, vol. 53, no. 7, pp. 1045–1051, 2010. [Online]. Available: http://comjnl.oxfordjournals.org/content/53/7/1045.abstract

[3] G. Procaccianti, A. Vetro, L. Ardito, and M. Morisio, "Profiling power consumption on desktop computer systems," in Information and Communication on Technology for the Fight against Global Warming, ser. Lecture Notes in Computer Science, D. Kranzlmller and A. Toja, Eds. Springer Berlin / Heidelberg, 2011, vol. 6868, pp. 110–123.

[4] "Global PC, Tablet, Smartphone Sales," 2010, last accessed on January 8, 2013. [Online]. Available: http://en.wiser.org/article/61670dc8dd5c5a3adba285f39bbf8145

[5] J. Scaramella, "Worldwide Server Energy Expense 2009–2013 Forecast," International Data Corporation (IDC), Tech. Rep., 2009, Document No. 221346. [Online]. Available: www.idc.com

[6] J.-L. Letouzey and T. Coq, "The sqale analysis model: An analysis model compliant with the representation condition for assessing the quality of software source code," in Advances in System Testing and Validation Lifecycle (VALID), 2010 Second International Conference on, aug. 2010, pp. 43 –48.

[7] ISO/IEC, 9126. Software engineering – Product quality. ISO/IEC, 2001.

[8] ISO/IEC, 25010. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models. ISO/IEC, 2011.

[9] S. Gude and P. Lago, "Best Practices for Energy efficient software," last accessed on February 17, 2012. [Online]. Available: http://wiki.cs.vu.nl/green_software/index.php/Best_practices_for_energy_efficient_software

[10] P. Larsson, "Energy-Efficient Software Guidelines," Intel Software Solutions Group, Tech. Rep., 2011.

[11] G. Kaefer, "Green SW Engineering: Ideas for Including Energy Efficiency into your Software Projects," Online Presentation, Siemens AG, Munich, Germany, 2009. [Online]. Available: http://www.cs.uoregon.edu/events/icse2009/images/postConf/TB-Energy-Efficient-SW-ICSE09.pdf

[12] M. Halstead and R. Bayer, "Algorithm dynamics," in Proceedings of the ACM annual conference, ser. ACM '73. New York, NY, USA: ACM, 1973, pp. 126–135. [Online]. Available: http://doi.acm.org/10.1145/800192.805693

[13] T. McCabe, "A complexity measure," IEEE Trans. Soft. Eng., vol. SE-2, no. 4, pp. 308 – 320, Dec. 1976.

[14] G. McCluskey, "Tuning Java I/O Performance," 1999, last accessed on February 17, 2012. [Online]. Available: http://java.sun.com/developer/technicalArticles/Programming/PerfTuning/

[15] "Junit – a programmer-oriented testing framework for java," last accessed on Feb, 17, 2012. [Online]. Available: http://www.junit.org/

[16] "WattsUp Pro ES," last accessed on Feb, 17, 2012. [Online]. Available: https://www.wattsupmeters.com/secure/products.php?pn=0&wai=0&spec=2

Definition, Implementation and Validation of Energy Code Smells: an Exploratory Study on an Embedded System

Antonio Vetro', Luca Ardito, Giuseppe Procaccianti, Maurizio Morisio

Dipartimento di Automatica ed Informatica

Politecnico di Torino

Torino, Italy

E-mail: name.surname@polito.it

Abstract—**Optimizing software in terms of energy efficiency is one of the challenges that both research and industry will have to face in the next few years. We consider energy efficiency as a software product quality characteristic, to be improved through the refactoring of appropriate code pattern: the aim of this work is identifying those code patterns, hereby defined as Energy Code Smells, that might increase the impact of software over power consumption. For our purposes, we perform an experiment consisting in the execution of several code patterns on an embedded system. These code patterns are executed in two versions: the first one contains a code issue that could negatively impact power consumption, the other one is refactored removing the issue. We measure the power consumption of the embedded device during the execution of each code pattern. We also track the execution time to investigate whether Energy Code Smells are also Performance Smells. Our results show that some Energy Code Smells actually have an impact over power consumption in the magnitude order of micro Watts. Moreover, those Smells did not introduce a performance decrease.**

Keywords-Code Smells; Energy Code Smells; Green Software; Software Metrics

I. INTRODUCTION

The issue of sustainability is starting to be addressed among the software engineering community. Although, it is still unclear how to design sustainable software. While for common quality characteristics (reliability, performance, security, etc.) processes and metrics have been proposed and widely investigated by the Software Engineering (SE) community, as regards sustainability the discussion is still in its initial phase.

Among the kaleidoscope of aspects related to software sustainability, one of the most visible is the energy (or, alternatively, power) consumption of software systems. Indeed, software does not consume energy directly, however it has a direct influence on the energy consumption of the hardware underneath. In fact, applications and operating systems indicate how the information is processed and, consequently, drive the hardware behaviour: previous work [1] suggested that software can increase the total power consumption of a computer system up to 10%. This and other initial findings [2] open investigation spaces on the optimization of energy and power consumption of IT devices

acting on the software instead of the hardware. Moreover, nowadays the same software runs on multiple devices, thus it might be more productive and feasible for software houses to green the single software rather then relying on the greening of all the hardware implementations underneath (that could require competences commonly not owned by software houses). Optimizing a software product in terms of energy efficiency has also some issues. The absence of a standard procedure, or a benchmark, to compare systems is the most prominent one. This is because software is intangible and it is deployed on devices with their own specifications and features. This makes really difficult to standardize a transparent, platform-independent measuring system for every software system.

Another consideration must be done regarding software architectures. During the last years, software engineers always tried to increase the number of software layers - that is, for improving interoperability, abstraction, decoupling, etc. However, the steep increase of software layers directed the optimization efforts only on each layer ("horizontal" optimization) and not across them ("vertical" optimization). Since energy efficiency directly relates with hardware technologies, a more intense communication flow between hardware and software is needed to achieve significant optimizations. In this sense, embedded systems make a perfect case study, because their architecture is simplified by design, and also because power consumption issues acquire a peculiar importance, for operational reasons (most embedded systems are battery-powered). For this reason our work uses an embedded system as the testbed to validate a new approach for the design and implementation of sustainable software. We investigate, and here we also introduce the goal and main contribution of this study, how software can be optimized by identifying code patterns that use in a sub-optimal way the hardware resources. These code patterns ought to be refactored in order to improve the energy efficiency of the software at run time. We define and name the code patterns Energy Code Smells, inspired by the well-known book of Fowler and Beck [3].

This study empirically validates the impact of Energy Code Smells over power consumption. We provide back-

ground in Section II, then in Section III we describe the used approach for the validation of the concept of Energy Code Smell. In Section IV we describe the experimental setup of our analysis: results (V), discussion (VI) and threats to the validity (VII) follow. Finally we mention the related works (VIII) and we expose our conclusions and future research (IX).

II. ENERGY CODE SMELLS: BACKGROUND AND DEFINITION

The term "code smells" was coined by Fowler and Beck [3] referring to poor implementation choices that make the software difficult to maintain. These bad implementation practices can be characterized as patterns in source code. For instance, the smell "*Long Method*" refers to a method that has grown too large: typically, the longer is the method the more difficult is to maintain it. One or more refactoring actions are associated to code smells: for example all you have to do to refactor a Long Method is to extract parts of the method that seem to go nicely together and make a new method. As a result the original method is shorter and easier to maintain. Refactoring code smells might have an effect not onaly on maintainability but also on other properties of the software, such as portability, testability or, as in the case of this work, the energy efficiency. As a consequence, we take inspiration by the original work of Fowler and Beck and we introduce the concept of smells into the Green IT community, introducing the Energy Smells:

A Energy Smell is an implementation choice that makes the software execution less energy efficient.

Since software has different levels of abstractions and organizations, Energy Smells can be located at code, design or architectural level. Therefore, Energy Code Smells are implementation choices *at source code level* (code patterns) that make a sub-optimal usage of the hardware resources underneath. As a consequence, they provoke a higher energy (or alternatively, power) consumption.

III. VALIDATION OF ENERGY CODE SMELLS

The aim of our research is to identify Energy Code Smells. In addition to that, we are also interested in understanding whether the Energy Code Smells also degrade the performances of the application in terms of execution time. We set up two research questions for our investigation:

RQ1. Which code patterns have an effect on power consumption (i.e. which code patterns are Energy Code Smells)?

RQ2. Code smells that have an effect on execution time do also have effect on energy consumption (i.e. are Energy Code Smells also Performance Smells) ?

The epistemological approach adopted for this research is the empirical one. We set up an experiment observing two

dependent variables: power consumption (W) for RQ1 and execution time (ms) for RQ2. The two dependent variables are measured on the execution of C++ functions running on an embedded device. The choice of the embedded device has several advantages, the main two being:

- it has no operating system and thus confounding factors in the experiment are minimized;
- it runs on a battery and it really needs energy efficient code.

In other terms, refactoring Energy Code Smells in such an environment might lengthen the life of the battery.

The potential Energy Code Smells selected for the experiment are code patterns used by two popular static analysis tools. For each code pattern selected for the experiment, we set up a C++ function with two implementations, one that violates the code pattern (thus contains a potential Energy Code Smell) and the refactored one without the violation. Therefore the treatment is the refactoring of the smell and it is possible to observe an effect on the two variables by comparing the measurements on the two versions of the code.

A. Potential Energy Code Smells selection

As introduced above, the software that runs on the selected device is C++ code. In order to identify Energy Code Smells on C++ code we look at already existing code patterns. In particular, we examined patterns implemented by Automatic Static Analysis (ASA) tools. ASA tools examine source and compiled code and check it against good programming practices and possible bug patterns. The advantage of using ASA tools is the speed of the verification and the applicability before testing or production phase.

The two tools selected for this study are Cpp-Check and Findbugs. CppCheck is a well-known static analysis tool for C/C++ which contains many patterns regarding a variety of desired software properties: safety, portability, performance, etc . An example of C/C++ pattern on portability is "*64 bits portability*", i.e. assign address to int or long. An example of checked pattern on performance is instead "*Address not taken*" of the category "Memory leaks", which detects when the address to allocated memory is not taken. In order to identify which patterns can be considered relevant for energy efficiency, two of the authors carefully read all patterns and selected independently which ones could cause a higher power consumption of the Waspmote. All conflicts (a pattern selected by only one expert) were resolved in a reconciliation meeting, where patterns were discussed and a final decision taken. In addition to the Cpp-Check patterns, we also reviewed the patterns of another static analysis tool, Findbugs. It is similar to Cpp-Check, but it analyzes Java code. The same two authors reviewed all FindBugs patterns and decided firstly if they can be applied to C++ code, then whether they might be related to energy efficiency.

The selection process ended up with the patterns shown in TABLE I.

TABLE I
POTENTIAL ENERGY CODE SMELLS SELECTED FOR VALIDATION.

Pattern Name	Pattern Description	Tool
Parameter By Value	Passing a parameter by value to a function	CppCheck
Self Assignment	Assignment of a variable to itself. (e.g., x=x).	CppCheck
Mutual Exclusion OR	OR operator between two mutually exclusive conditions (thus always evaluating to true).	CppCheck
Switch Redundant Assignment	Redundant assignment in a switch statement: for example, assigning a value to a variable in a case block without a following break instruction, then re-assigning another value to the same variable in the subsequent case block.	CppCheck
Dead Local Store	A statement assigning a value to a local variable, which is not read or used in any subsequent instruction.	FindBugs
Dead Local Store Return	A return statement assigning a value to a local variable, which is not read or used in any subsequent instruction. (i.e. return(x=1);)	FindBugs
Repeated Conditionals	A condition evaluated twice (e.g., x==0 —— x==0).	FindBugs
Non Short Circuit	Code using non-short-circuit logic boolean operators (e.g., & or ‖) rather than short-circuit logic ones (&& or ‖‖). Non-short-circuit logic causes both sides of the expression to be evaluated even when the result can be inferred from knowing the left-hand side.	FindBugs
Useless Control Flow	Control flow constructs which do not modify the flow of the program, regardless of whether or not the branch is taken (e.g., an if statement with an empty body).	FindBugs

Subsequently, we wrote for each of the patterns a pair of C++ functions, one containing a potential smell and another one refactored without that smell. For example, the "Non-Short Circuit Logic" pattern, shown in Listing 1 has the following two functions:

```
void NonShortCircuit_With(){
        int count = 0;
        int total = 345;
        if ( count > 0 & total / count > 80 )
                count=0;
}

void NonShortCircuit_Without() {
    int count = 0;
        int total = 345;
        if ( count > 0 && total / count > 80 )
                count=0;
}
```

Listing 1.   Non Short Circuit Code Pattern

The function NonShortCircuit With() is the one with the potential smell "*Non-short circuit logic*". The smell is in

the line $if(count > 0 \& total/count > 80)$ because the AND operator is single & and so both predicates in the expressions will be evaluated at run-time. In the function, NonShortCircuit Without() the code is refactored replacing & with &&. All functions are available online [4] for the sake of replication.

IV. EXPERIMENT

A. Context: the WASP

The device used for the experiment is the Waspmote V1.1 (Libelium Comunicaciones Distribuidas S.L. Esso). The hardware architecture is based on a ATmega 1281 microcontroller with a CPU frequency of 8 MHz and 8KB of SRAM. It has no operating system: programs are directly loaded on a FLASH memory of 128 K. This architecture well suites our experiment because no other threads run in parallel with the chosen program, thus eliminating any software noise for the energy measurement. The device is basically a motherboard with connectors to plug in other elements such as sensors, wireless modules (ZigBee, XBee, Bluetooth), GSM/GPRS modules and a GPS (Global Positioning System) module. For this reason it is used in different fields, such as Smart Metering, Building Automation, Agriculture etc. It runs on a lithium battery (3.7V and 1150mAh), so the energy consumption of software has a key role here. To compile and load the C++ programs it is sufficient to use the IDE provided by the manufacturer and connect it to a computer via USB cable.

B. Experiment setup

The objective of the experiment is to measure power consumption and execution time on each function pair, in order to evaluate if the potential smell affects the two dependent variables. We divide the experiment in two parts: one for measuring power consumption, and another one for the execution time.

Measuring power consumption and execution time for a single function is a challenging task because usually execution is too fast to get reliable data. We control this threat repeating each function 1 million of times, that makes one sample. We collect 50 sample in order to reach statistical significance. Each function pair is loaded on the Waspmote and evaluated two times: the first one for the execution time, the latter one for the power consumption.

No specific instrumentation was needed to obtain the execution time, because the Waspmote embeds a Real-Time Clock (RTC) with a millisecond accuracy. We measure the execution time of every loop (i.e. 50 measurements).

On the other side, analyzing power consumption is more complicated. The only way to obtain a precise measure of the power consumption is using a power meter. The RTC is powered by an auxiliary battery, which makes it completely independent from the main power supply. Therefore it is possible to power the Waspmote with a constant voltage

(V_G = 3.7 V) by means of a generator and use a shunt resistor to measure the current intensity. An analog to digital converter (ADC) connected to the PC reads the voltage drop across a resistor R of 1 Ω. The current flowing in the circuit can be computed by measuring the voltage drop on the resistor ($I = V_{ADC}/R$). The instant power consumption value can be computed as:

$$P = V_L \cdot I = (V_G - V_{ADC})\frac{V_{ADC}}{R} = \frac{V_G V_{ADC} - V_{ADC}^2}{R} \quad (1)$$

Fig. 1 represents the circuit described.
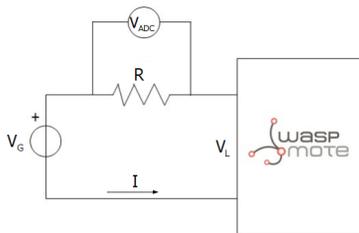
Figure 1.   Circuit built to measure the power consumption.

The device used to measure the power consumption has a frequency of 49KHz, i.e. it gets 49000 measurements each second. In order to precisely measure the power consumption relative to the execution of the function pairs, we inserted a sleep interval at the beginning of the data acquisition to exclude the peak of device power on, and we filtered out, through a threshold, all the measurements corresponding to the idle consumption between the iterations of the function execution. As shown in Fig. 2, the threshold filters out the transient and includes only the peaks corresponding to the actual execution of the function.
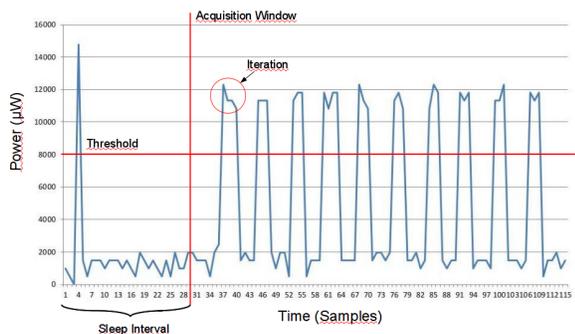
Figure 2.   Sampling current intensity: an example.

C. Analysis methodology

For each research question we derived a pair of null and alternative hypotheses to test.

RQ1:

$$H1_0 \quad : P_{with} \leq P_{without}$$

TABLE II
RESULTS OF POWER CONSUMPTION.

Smell name	Mean with smell (μW)	Mean w/o smell (μW)	Diff. Means (μW)	P-val	Impact%
Dead Local Store Return	41241	41278	-37	1	-0.09
Dead Local Store	40249	40205	44	**≤ 0.01**	0.11
Mutual Exclusion OR	40758	40772	-14	1	-0.03
Non Short Circuit	41113	41043	70	**≤ 0.01**	0.17
Parameter By Value	40967	40723	244	**0**	0.60
Repeated Conditionals	41155	41126	29	**≤ 0.01**	0.07
Self Assignment	40952	40879	73	**≤ 0.01**	0.18
Switch Redundant Assignment	40724	40756	-32	1	-0.08
Useless Control Flow	41051	41142	-91	1	-0.22

$$H1_a \quad : P_{with} > P_{without}$$

where P is the power consumption of the function, with and without the potential smell. If the refactored version of the function consumes less than the function with the smell, the null hypothesis is rejected in favor of the alternative one. As a consequence we consider the pattern a Energy Code Smell. The hypothesis is tested with the Mann-Whitney test, given $\alpha = 0.05$.

RQ2:

$$H2_0 \quad : T_{with} \leq T_{without}$$
$$H2_a \quad : T_{with} > T_{without}$$

where T is the execution time of the two functions. If the smell has a negative impact on performance, the refactored function will be faster and the null hypothesis is rejected. In that case, we consider the pattern a Performance Smell. In order to answer RQ2, we compare which Energy Code Smells are also Performance Smells. We also use Mann-Whitney and $\alpha = 0.05$ to test the hypotheses.

At the end of the experiment each function has 50 measurements of execution time and about 25 millions of power measures. Then, after filtering out values below the idle threshold (8mW), we obtained about 8 millions values for power measurement, on which we ran the analysis.

V.   RESULTS

We report results on the power consumption and execution time respectively in TABLE II and III. The two tables report the name of the smell, the means and their difference for both the dependent variables, the p-value of the Mann Whitney test [5] and the difference in percentage of the power consumption (or execution time) between the execution of the code with the smell and the execution with the refactored code.

TABLE III
RESULTS OF EXECUTION TIME.

Smell name	Mean with smell (ms)	Mean w/o smell (ms)	Diff. Means (ms)	P-val	Impact%
Dead Local Store Return	3288.76	3288.74	0.02	0.41	6.08e-04
Dead Local Store	17707.34	17707.38	-0.04	0.66	-2.26e-04
Mutual Exclusion OR	3540.76	3540.60	0.16	**0.04**	4.52e-03
Non Short Circuit	3288.74	3288.80	-0.06	0.76	-1.82e-03
Parameter By Value	3288.76	3288.74	0.02	0.41	6.08e-04
Repeated Conditionals	3288.80	3288.74	0.06	0.24	1.82e-03
Self Assignment	3288.66	3288.78	-0.12	0.90	-3.64e-03
Switch Redundant Assignment	3540.58	3540.62	-0.04	0.65	-1.13e-03
Useless Control Flow	3288.80	3288.74	0.06	0.24	1.82e-03

We observe from TABLE II that all power consumptions ranged from 40mW to about 42mW. Five code patterns over nine have a p-value ¡ 0.05 (in bold) and therefore the null hypothesis is rejected for them. The code patterns are:

- DeadLocalStore
- NonShortCircuit
- ParameterByValue
- RepeatedConditionals
- SelfAssignment

Overall the saved power consumption is less than 1%.

The answer to RQ1 is: *five code patterns (DeadLocalStore, NonShortCircuit, ParameterByValue, RepeatedConditionals, SelfAssignment) are Energy Code Smells, and their impact is in the order of μW*.

Focusing on performance, from TABLE III becomes evident that there is no difference in execution time. The null hypothesis is rejected only for MutualExclusionOr, however the magnitude order is μ seconds. We also notice that DeadLocalStores are about 5 times slower.

Thus, our answer to RQ2 is: *Energy Code Smells are not Performance Smells*.

VI. DISCUSSION

We identified five smells which provoked a higher power consumption of the Waspmote in the use cases prepared for the experimentation. However, we observe that the saved power is less than 1 %. A first motivation resides in the implementation choices: the function pairs executed only differ in a single instruction, and the operations are done with primitive types (e.g., integer). The motivation of such implementation was the exclusion of any possible confounding factor in the analysis, but the drawback of such a choice is a very small achievement in energy efficiency improvements. Let us take dead stores as example: the smell *DeadLocalStore* is implemented with an integer (we

save a value on a variable and immediately overwrites it with another integer). Using a struct with several members is totally different and might lead to a higher impact, because the resulting compiled code requires the CPU to produce more instructions and interact more intensively with the memory. If increasing the complexity of the data structure will result in still negligible power consumption saving, the next step is to increase the logical complexity of the function, i.e. comparing complete algorithms that are functionally equivalent but differ in the implementation. A further step is to move the focus towards the comparison of functionally equivalent design choices. Understanding the impact of Energy Code Smells over real power consumption could also contribute to build more precise models of the power consumption of software. As a matter of fact, it may be possible to categorize software instructions beforehand in terms of energy efficiency, then subsequently use this information in order to predict the resulting energy efficiency of a complete software product.

Yet another research direction that is suggested by this first leap is: can the impact of Energy Code Smells be higher in code that drives an hardware resource with higher energy needs? For instance the impact on the code that handles the GPS transmitter is expected to be very different from the one used in this experiment, where the small functions use only CPU and RAM, besides in a not intensive way. The same investigation approaches can be applied to the domain of execution time. As can be noticed from the results, all the execution times are equal, exception given for the *DeadLocalStore* function pairs. We have observed that Energy Code Smells do not degrade the performances, but we cannot generalize the findings for more complex code structures and usage scenarios, with different hardware resources involved (e.g, sensors).

VII. THREATS TO VALIDITY

In this section, we expose the threats to validity that might affect our study.

As regards construct validity, our main threat regards instrumentation. We carefully evaluated the precision of our measures, comparing them with the specifications from Waspmote manufacturers. During our experimentation, the difference between actual and expected values was negligible and inside the specified ranges. As far as conclusion threats are concerned, in order to increase the statistical reliability of the results, we collected a relevant amount of values (e.g., every function is looped 1 million times for power consumption measurement resulting in 25 millions of samples). Internal validity is represented by confounding factors such as other processes running during execution. However, the Waspmote does not have an operating system and the only thread in execution during the tests is the code loaded. As regards external validity, we do not aim at generalizing our results to a family of embedded devices.

This study aims at assessing the existence of the Energy Code Smells in a single context: other empirical validations are necessary for other environments or devices.

VIII. RELATED WORK

We did not find in the literature similar approaches for energy efficiency optmization. However, we found techniques that rely on algorithmic and data optimizations. The algorithmic optimization has a high potential, but it is also a hard and time-consuming task, with no guaranteed results. Data optimization is based upon the efficient use of the system architecture. For example, as regards embedded systems, often software libraries are used for emulating floating-point hardware components. Those libraries do not take into account the architecture of a specific system, thus their usage often leads to a high power consumption and low performance. Šimunić et al. [6] show that by removing those libraries and optimizing the source code, it is possible to significantly reduce power consumption (up to 77%).

In terms of benchmarks, SPECpower [7] is an initiative to extend existing SPEC benchmarks to power and energy measurement. SPECpower ssj2008 reports the energy efficiency in terms of overall $ssj\_ops/watt$. This metric represents the sum of the performance measured at each target load level (in ssj_ops) divided by the sum of the average power (in watts) at each target load including active idle.

In battery-powered systems, it is not enough to analyze algorithms based only on time and space complexity. Several research proposed energy aware algorithms for specific functionalities, such as supporting randomness [8] or focusing on cryptographic [9].

Previous work by Bunse et al. [10] addresses the relationship between energy and performance optimizations, which is one of the research questions of the present work. Authors analyzed different implementations of several sorting algorithms, showing that implementations optimized for energy performed differently with respect to those optimized for performance. This findings holds in our work, since we found that Energy Code Smells are not Performance Smells.

IX. CONCLUSIONS

This is an exploratory study: we defined for the first time the concept of Energy Code Smells and we performed a first validation to understand not only the impact, but also the boundaries of the concept. We identified some Energy Code Smells starting from code patterns implemented by two common Automatic Static Analysis tools - namely, CppCheck and FindBugs. We performed an experiment, on an embedded system, in order to assess the energetic impact of those code patterns and determine whether Energy Code Smells are also performance smells. Our experimental results showed that some of the code patterns actually have an impact over power consumption. This impact, however, is in the magnitude order of μW. Our future research works

will be devoted to analyzing more complex data structures and using hardware resources which could increase this impact with respect to the overall power consumption. As regards time analysis, only one pattern had an actual impact over execution time (a few μ seconds), and it is not identified as a Energy Code Smell. Thus, we conclude that Energy Code Smells are not Performance Smells. Results suggest that the target and applicability of Energy Code Smells should be refined with further investigations. The lessons learned in this exploratory study let us identify several research threads that the research community might address, such as the identification of Energy Code Smells that are higher-level constructs with more complex data structures, the identification of Green Design Smells and the use of more complex systems as test beds. Finally, the experimental results that will be collected might be also used to build more precise models of the power consumption of software.

REFERENCES

[1] G. Procaccianti, A. Vetro', L. Ardito, and M. Morisio, "Profiling power consumption on desktop computer systems," in Information and Communication on Technology for the Fight against Global Warming, ser. Lecture Notes in Computer Science, D. Kranzlmüller and A. Toja, Eds. Springer Berlin / Heidelberg, 2011, vol. 6868, pp. 110–123.

[2] A. Vetro', L. Ardito, M. Morisio, and G. Procaccianti, "Monitoring it power consumption in a research center: Seven facts," in Proceedings of The First International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies, ser. ENERGY 2011, 2011, pp. 64–69.

[3] M. Fowler and K. Beck, Refactoring: improving the design of existing code. Addison-Wesley Professional, 1999.

[4] A. Vetro', L. Ardito, G. Procaccianti, and M. Morisio, "Energy code smells," January 2013, retrieved January 2013. [Online]. Available: http://softeng.polito.it/greensmells/

[5] L. Sachs, Applied Statistics–A Handbook of Techniques, S. S. in Statistics, Ed. Springer-Verlag, 1984.

[6] T. Šimunić, L. Benini, G. De Micheli, and M. Hans, "Source code optimization and profiling of energy consumption in embedded systems," in Proceedings of the 13th international symposium on System synthesis. IEEE Computer Society, 2000, pp. 193–198.

[7] K.-D. Lange, "Identifying shades of green: The specpower benchmarks," Computer, vol. 42, no. 3, pp. 95–97, Mar. 2009. [Online]. Available: http://dx.doi.org/10.1109/MC.2009.84

[8] R. Jain, D. Molnar, and Z. Ramzan, "Towards understanding algorithmic factors affecting energy consumption: switching complexity, randomness, and preliminary experiments," in Proceedings of the 2005 joint workshop on Foundations of mobile computing. ACM, 2005, pp. 70–79.

[9] N. Potlapally, S. Ravi, A. Raghunathan, and N. Jha, "A study of the energy consumption characteristics of cryptographic algorithms and security protocols," Mobile Computing, IEEE Transactions on, vol. 5, no. 2, pp. 128–143, 2006.

[10] C. Bunse, H. Hopfner, E. Mansour, and S. Roychoudhury, "Exploring the energy consumption of data sorting algorithms in embedded and mobile environments," in Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on. IEEE, 2009, pp. 600–607.

Resource-efficient Management Scheme of Sensor Data

Hyunjeong Lee, Jaedoo Huh, Il-Woo Lee
Energy IT Technology Research Section
Electronics and Telecommunications Research Institute
Daejeon, South Korea
{hjlee294, jdhuh, ilwoo}@etri.re.kr

Sang Ho Lee
Dept. of Software Engineering
Chungbuk National University
Cheongju, Chungbuk, South Korea
shlee@cbnu.ac.kr

Abstract- In this paper, we propose a resource-efficient management scheme of sensor data for building energy saving. To manage buildings, we collect data from the variety of sensors in buildings able to recognize illumination, motion, temperature, humidity and so on. We analyze this data and derive energy saving measures from it. In case there is a large amount of data from the sensors, most of it is redundant if the data period is shorter than needed for services. We need a method to manage the data and to save the data in a resource-efficient manner, in terms of energy, time and space to save the data. Therefore, we propose a method to select the adequate period of sensor data to manage and use some meaningful data for the energy saving service. Using this method, we can save the resources for sensor data and derive the energy saving measures more quickly and efficiently without all sensor data.

Keywords-Resource; Sensor Data; Compression; Energy.

I.   INTRODUCTION

These days, environment pollution and climate changes have become world-wide issues [1], and international community has made progress towards the reduction of carbon emissions. To support these efforts, many kinds of green technologies are studied in many areas including buildings. Since residential and commercial buildings consume over 21% of total energy use in South Korea, many energy management solutions in buildings are studied and developed such as Building Automation System (BAS) and Building Energy Management System (BEMS) [2]. BEMS manages and saves energy consumption in buildings by collecting many kinds of data from building automation system and sensors in buildings. Some BEMS technologies have more than 8,000 sensors, so the volume of sensor data is too large to save and manage. Many works have studied the compressing of data in order to save resources [3][4], such as energy, time, storage and so on. Shannon-Nyquist sampling theorem has been widely used as a base theory for digital systems [5][6]. But, the theorem is assessed again in academia by Compressed Sensing (CS) theory [7][8]. Donoho et al. introduce CS and show that fewer number of samples are required than the conventional approach. They consider that raw samples are redundant and can be compressed before being transmitted.

In this paper, we propose a method of efficient management of sensor data in terms of energy, time and space saving. We collect data from sensors and save it into the database according to the service interval. Using our

scheme, we can save energy, time and space for the sensor data for building energy saving.

The rest of this paper is organized as follows: We introduce background and related works in Section II. We propose resource-efficient management scheme of sensor data in Section III. Finally, we draw conclusions in Section IV.

II.   RELATED WORKS

A.   Conventional Compression and Compressive Sensing Mechanism

In a conventional sensor system, each sensor node performs Analog-to-Digital Conversion (ADC) at the Nyquist rate that should be set at least twice the maximum frequency of the measured signal [9]. Uniformly spaced samples are obtained during a fixed time period. These raw samples are redundant and they can be compressed before being transferred. Because sending the redundant data is inefficient, compression can be done in order to reduce the number of bits transmitted and save transmittal power. However, two things are considered for this approach. First, is that all the sensed samples are required to be gathered and stored at a collaboration location, and thus inter sensor correlation cannot be performed unless all the sensed samples are at a single location. Second, compression operations can be performed at the central location.
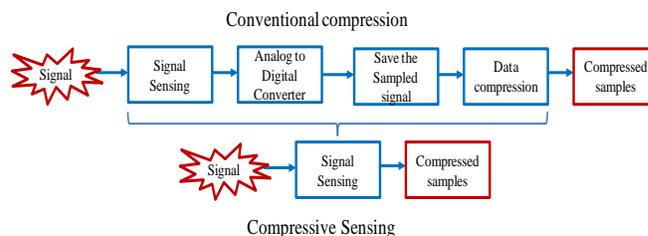
Figure 1.   Conventional Compression and Compressive sensing

Compressive sensing is a sampling method that converts high dimension input signal into another signal that lies in a space of smaller dimension [10][11]. As described in Figure 1, Donoho et al. suggested the compressive sensing theory, which obtains compressed signal from complex signal processing using simple matrix operations, under the assumption that the observed signal is sparse [7][9]. Using

CS, obtaining signal from sensors can be very simple and efficient without complex computation [6].

B. Building Energy Management Technology

As time goes on, the performance of various equipments in buildings is degraded, and energy usage increases because of faults and errors of the old equipments. So, building energy management is essential for maintaining initial performance [2]. However, equipments in most buildings are not managed systematically and professionally, because monitoring and inspection of the facilities are mostly based on the facility managers' experience. To solve these problems, many technologies, such as Energy Management Systems (EMS), Facility Management Systems (FMS) and BEMS, have been developed and used for building management. For these technologies, the main mechanism for efficient management of building is BEMS. The technology includes the software, hardware, and services associated with the Information and Communication Technology (ICT)-based monitoring, management, and control of energy. The main function of BEMS is to reduce the whole energy cost by decreasing energy consumption. The BEMS is similar to a traditional BMS, but it differs in that it places more emphasis on energy [3]. The BEMS has many functions for building energy management, and its main features are as follows:

- Monitor, track, control, and manage buildings to reduce energy use and maintain comfort of residents
- Provide optimized energy control measures by reducing unnecessary energy use
- Monitor, track, and manage equipment and energy usage via sensors, meters and systems

These days, most of BAS vendors have their own building management solutions. As shown in Figure 2, the Enterprise Buildings Integrator (EBI) system of Honeywell is for integrated facility management, application areas in buildings to provide control and manage functions [12]. The Architecture of Honeywell EBI system consists of various databases and application programs, which enable building managers to monitor, control and display the equipments in buildings. Siemens has building automation software, named APOGEE [13], which offers many functions for building management. First, it integrates multiple building systems no matter the manufacturer or installer achieving greater flexibility through open protocol interoperability. Also, it expands, upgrades, and optimizes facility systems in buildings.
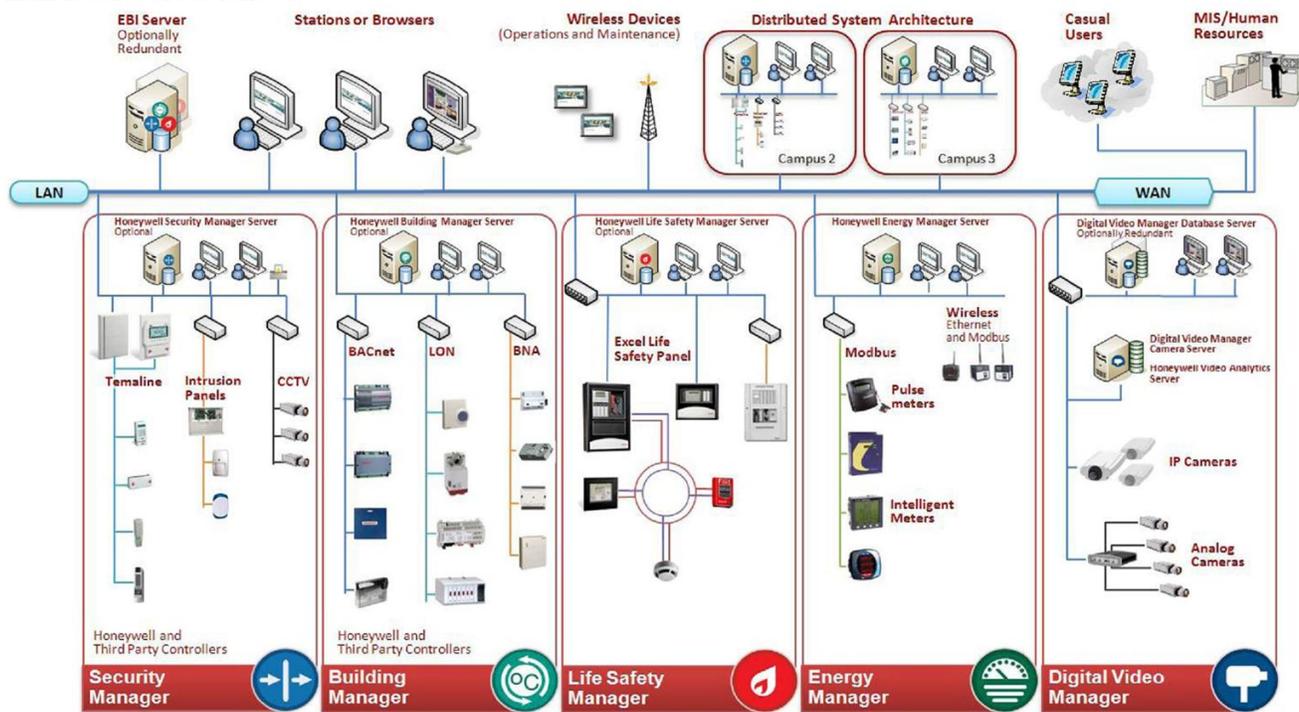
Figure 2.   An Architecture of Honeywell EBI system [12]

III.   RESOURCE-EFFICIENT MANAGEMENT SCHEME OF SENSOR DATA

We implemented a resource-efficient management scheme of sensor data for building energy saving. There is a lot of data from various sensors in buildings that can be used

for energy saving. From all this data, we periodically select only some data to use and manage, considering the service interval. For example, if the service interval is 15 minutes, then we choose 1 piece of data per 1 minute. This way, the amount of data is smaller than that of original data. Also, data processing time can be reduced. As illustrated in Figure

3, the Remote Building Management Center (RBMC) manages a lot of building energy. Each building has many sensors and meters to collect data such as temperature, humidity, illumination, motion, power and gas, and so on. For meaningful data, the number of sensors in a large building is over 8,000, and the interval of some sensor data is very short, such as 1 second or less. When data is collected from sensors in the unit time of 1 second, some data has different values at the same time due to the sensing range and the collecting time interval. Then, the data processing time, the size of the database and data consistency can be the problem in the RBCC.
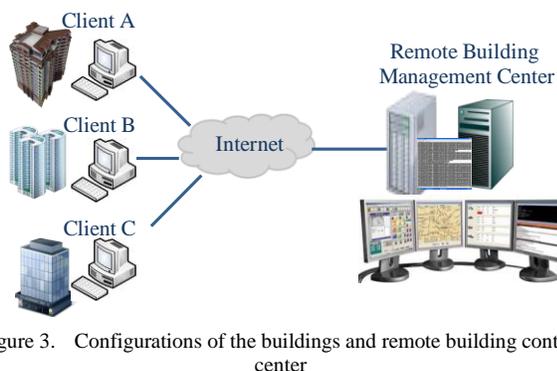
Figure 3.   Configurations of the buildings and remote building control center

We installed four kinds of sensors, such as illumination, temperature, humidity and motion, in our office as shown in Table I, where the temperature and humidity sensors exist in one module. We collected sensor data for 192 days in our office. Although the data is for only part of the building, the total amount of data is huge.

TABLE I.   DESCRIPTION OF FOUR KINDS OF SENSORS

Category	Number of Sensors	Total Number of Data	Number of Data for one day
Illumination	13	22,279,198	116,011
Motion	4	8,836,260	26,159
Temperature & Humidity	9	19,058,079	97,453

The interval of each sensor data and the number of data per minute is as illustrated in Table II. The data was collected and saved every 6-9 or 20 seconds, but energy saving services in buildings can be performed every 10 and 15 minutes. The energy saving service for lamps in restrooms of our building is performed every 10 minutes, and the duration of 15 minutes is frequently used as time interval for delivering power meter information [12]. Although the duration is shorter, the amount of energy saving may be similar because of the cost of the power on and off. Furthermore, the comfort of the residences in a building can be reduced. If the duration is longer, then the amount of energy saving may be reduced. In case that the time interval is 15 minutes, the saving scenario is as follows: if there is no-one in a zone for 15 minutes, then the lights in

the zone are automatically turned off. So, the number of data per minute is unnecessarily frequent for our service.

TABLE II.   INTERVAL OF EACH SENSOR DATA AND THE NUMBER OF DATA PER MINUTE

| Category	Sensor ID	Duration (second)	Number of data per one Minute	Average Number of Data per one Minute
Illumination	130f	20	3	7.3
	1310	20	3	
	1312	6~8	8.57	
	1313	6~8	8.57	
	1314	6~8	8.57	
	1315	6~8	8.57	
	1316	6~8	8.57	
	1317	6~8	8.57	
	1318	6~8	8.57	
	1319	20	3	
	131a	6~8	8.57	
	131b	6~8	8.57	
	131c	6~8	8.57	
Motion	1405	6~8	8.57	7.2
	1406	6~8	8.57	
	1407	20	3	
	1408	6~8	8.57	
Temperature & Humidity	120a	6~8	8.57	7.3
	120b	6~8	8.57	
	120c	20	3	
	120d	20	3	
	120e	6~9	8.57	
	120f	6~9	8.57	
	1210	6~8	8.57	
	1211	6~9	8.57	
	1212	6~8	8.57	

We propose a method to save and manage the sensor data considering the service interval. We supposed that the interval of data from illumination, temperature and humidity is every 1, 2, 5, 10 and 15 minutes. The duration of 15 minute is the unit of metering system for power usage statistics used in Korea [14]. Also, it is used for demand controller because the sensor values of temperature and humidity do not change much, and the value of illumination is needed only when the service is performed. However, the values of motion sensor are needed every minute to detect the duration of absence. So, we use the sensor data every minute among the collected data.

Table III shows the result of the number of data for one day to be saved in the database. The volume of data is very small and the time to save is shorter compared with the original data.

TABLE III.   THE NUMBER OF DATA BY THE PROPOSED MECHANISM

Category	Number of data (per minutes)				
	1	2	5	10	15
Illumination	15,819	7,916	3,167	1,584	1,056
Temperature & Humidity	5,760	2,880	1,152	576	384
Motion	12,951	6,479	2,592	1,296	864

As illustrated in Figure 4, we get the pattern of one illumination sensor data for 1 hour from 7 to 8 AM in the morning to view detailed data. As seen in the graph, if the interval is over 10-minutes, then the effect of the energy

saving service can be reduced. In these cases, we have to pick the reasonable duration for the service.
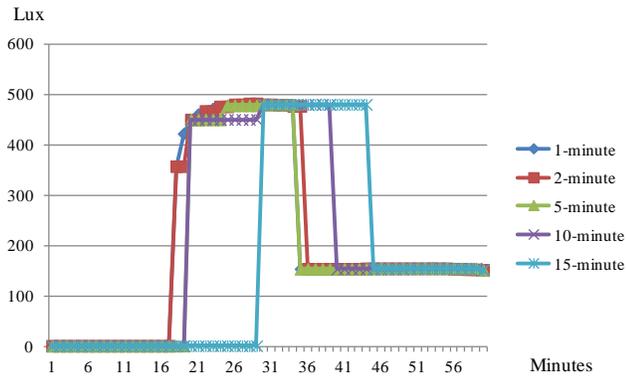
Figure 4.   Data of Illumination Sensor for 1 hour

When the lux (lx) value for illumination is over 300, we know that the lamp is on. So, we can get the accuracy rate for the energy saving service using the following model:

$$Accuracy^{s,i}= D_{300}/D_{Total} \qquad (1)$$

s: service, i: interval
D_{300} : Number of lux data over 300
D_{Total}: Number of total data

In the 2-minute interval, the accuracy rate is 98.3 because the number of accurate data is 59 out of the total of 60. The number of accurate data for the durations of 5, 10, 15 minutes are 58, 53 and 38, respectively. If the building policy of energy saving needs the accuracy rate of 90%, then the interval 10 and 15 have to be discarded. Table IV shows the comparison of the accuracy between the original data and the proposed mechanism for the data of 1 hour. In the table, we show the number of original data and the revised number of accurate data. Here, we assume that the data value is same as the previous data for the duration in the revised number of accurate data.

TABLE IV.   COMPARISON OF THE ACCURACY BETWEEN THE ORIGINAL DATA AND THE PROPOSED MECHANISM FOR THE DATA OF 1 HOUR

Category	Accuracy rate (%)					
	Original Data	Period (minute)				
		1	2	5	10	15
Original No. of Data	543	60	30	12	6	4
Revised No. of Accurate Data	543	60	59	58	53	38
Accuracy rate (per 1 minute)	100	100	98.3	96.7	88.3	63.3

We get the illumination pattern for 1 day with the same sensor above as described in Figure 5. The patterns are

almost the same even if the duration is longer. But, if we select sensor data for long duration, then the energy saving service also can't be provided in time and the amount of saving energy is reduced. For example, if we have data for each 5 minutes, then the light may be turned on for 4 minutes and 59 seconds after users' absence. So, we have to select the adequate period considering the energy saving rate and data management time and storage space as following:

$$Period^{s,i}= \alpha \times Saving^{energy} + \beta \times Saving^{db, time} - \gamma \times Error^{probability} \qquad (2)$$

s: service, i: interval
α : The weight of the energy model
β : The weight of the database and processing time
γ : The weight of the error probability

In (2), the weight for the period of data selection is determined by the building energy manager. The error probability is calculated as:

$$Error^{probability} =Interval^{data} /Interval^{service} \qquad (3)$$
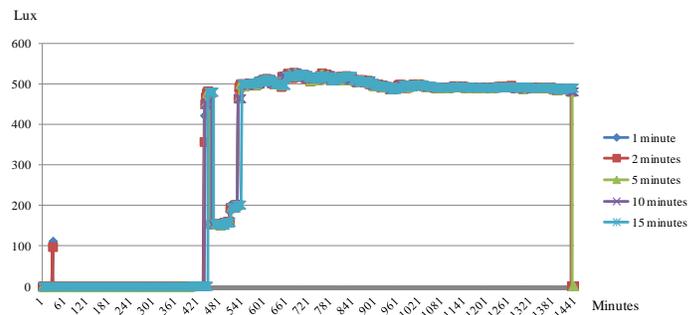
Figure 5.   Data of Illumination Sensor for 1day

TABLE V.   RESULT OF THE PROPOSED MECHANISM FOR THE DATA OF 1 DAY

Category	Accuracy rate (%)					
	Origin-al Data	Period (minute)				
		1	2	5	10	15
Original No. of Data	12,989	1,440	720	288	144	96
No. of Accurate Data	12,989	1,440	1,439	1,435	1,430	1,400
Accuracy rate (per 1 minute)	100	100	99.9	99.6	99.3	97.2
Reduction rate of No. of Data	0	88.9	94.5	97.8	98.9	99.3
Probability of Error (per 1 minute for 15-minute interval)	0	0	13.3	33.3	66.7	100
Result (α = β = γ =1)	100	188.9	181.1	164.1	131.5	96.5

Table V shows the result when we assume that all values of α, β, and γ are 1. We consider the accuracy rate for energy saving, then we select the data period as 1 minute. We can select best period as 1 minute. Also, 2 or 5 minutes of intervals can be chosen according to the building's energy policy. 10 or 15-minute interval is difficult to be selected because the probability of service error is high and the amount of energy saving is reduced.

IV. CONCLUSION

Because the environment pollution and climate changes have become world-wide concerns, energy saving is essential in all industrial areas. For this reason, several technologies are explored for saving energy in buildings. Among these technologies, BEMS manages and saves energy consumption in buildings by collecting several kinds of data. A challenge is that the large volume of sensor data constitutes a problem to save and manage.

Many works have studied data compressing in order to save resources, such as energy, time, storage, and so on. These works consider that redundant raw samples can be compressed before being transmitted.

In this paper, we proposed an efficient management of sensor data in terms of energy, space and time saving. Using our proposed mechanism, the time and space to save the sensor data was reduced compared to the case that did not use our mechanism. We plan to measure and analyze the amount of the energy saving and its effects as further study. We are also planning to apply the compressive sensing regarding the relations among sensors.

REFERENCES

[1] H. Lee, J. Han, Y.-K. Jeong, and I.-W. Lee, "Profile-based Building Energy Saving Service in Green Computing Environment," Proceedings of the 2011 Internation Conference on ICT Convergence, vol. 1, no. 1, Seoul, South Korea, Sept. 2011, pp. 768-769.

[2] C.-S. Choi, W.-K. Park, Y.-K. Jeong, and I.-W. Lee, "Remote Building Energy Management System Architecture and Integration Network Protocol for Heterogeneous Building Automation System," Proceedings of the 4th Internation Conference on Experiments/Process/System Modeling/Simulation/Optimization, vol. 4, no. 1, Athens, Greece, July 2011, pp. 548-555.

[3] M. Sharifi and M. Okhovvat "Scate: A Scalable Time and Energy Aware Actor Task Allocation Algorithm in Wireless Sensor and Actor Networks," ETRI Journal, vol. 34, no. 3, June 2012, pp. 330-340.

[4] J. Zhang, C. Wu, Y. Zhang, and P. Ji "Energy-Efficient Adaptive Dynamic Sensor Scheduling for Target Monitoring in Wireless Sensor Networks," ETRI Journal, vol. 33, no. 6, Dec. 2011, pp. 857-863.

[5] C.Shannon, "A Mathematical Theory of Information, " Bell System Technical J., vol. 27, July 1948, pp. 379-423.

[6] H.-N. Lee, S. Park, and S. Park, "Introduction to Compressive Sensing," The Magazine of the IEEK, vol. 38, no. 1, Jan. 2011, pp. 19-30.

[7] D. L. Donoho, "Compressed sensing," IEEE Trans. On Information Theory, vol. 52, no. 4, April 2006, pp. 1289-1306.

[8] E. J. Candes and T. Tao, " Near optimal signal recovery from random projections: Universal encoding strategies?," IEEE Trans. On Information Theory, vol. 52, no. 12, Dec. 2006, pp.5406-5425.

[9] J.-G. Choi, S. Park, and H.-N. Lee, "Comressive Sensing and its Applications in Wireless Sensor Network & Recovery Performance Analysis of Correlated Signals," The Summer Conference 2012 of KICS, vol. 48, Jeju, Korea, June 2012, pp. 1046-1047.

[10] R. Baraniuk, "Compressive sensing," IEEE Signal Processing Magazine, vol. 24, no. 4, July 2007, pp. 118-121.

[11] E. J. Candes and M. Wakin, "An Introduction To Compressive Sampling," IEEE Signal Processing Magazine, vol. 25, no. 2, March 2008, pp. 21-30.

[12] Honeywell Enterprise Building Integrator,
http://honeywell.com/sites/hbs/produkte/Systeme/Pages/systemintegration.aspx [retrieved: Jan, 2013].

[13] Siemens APOGEE, http://w3.usa.siemens.com/buildingtechnologies/us/en/building-automation-and-energy-management/apogee/pages/apogee.aspx [retrieved: Jan, 2013].

[14]   Korea Energy Management Corporation, http://www.kemco.or.kr/ [retrieved: Jan, 2013].

Next Generation Power Distribution System Architecture: The Future Renewable Electric Energy Delivery and Management (FREEDM) System

Alex Huang, Xu She, Xunwei Yu, Fei Wang, and Gangyao Wang

Future Renewable Electrical Energy Delivery and Management Systems Center
North Carolina State University
Raleigh, NC, US
{aqhuang, xshe, xyu8, fwang5, gwang3} @ncsu.edu

Abstract—**This paper presents the envisioned next generation power distribution system architecture, the so called FREEDM system. The presented system enables the plug-and-play of distributed renewable energy resources and distributed energy storage devices, therefore functions as the "energy internet". The recent technical developments are summarized in this paper, with emphasis on the distributed grid intelligence, solid state transformer, DC microgrid, and fault isolation device. It is demonstrated that the FREEDM system is a highly attractive candidate for the future power distribution system.**

Keywords-Smart grid; Microgrid; distributed generation; energy storage; DNP 3; Solid State Transformer.

I.   INTRODUCTION

The future renewable electric energy delivery and management (FREEDM) system center is a US National Science Foundation (NSF) generation-III Engineering Research Center (ERC) established in 2008 headquarters at North Carolina State University, US. The mission of the center is to develop the fundamental and enabling technologies for future power distribution system, which is called the FREEDM system [1].

Figure 1 shows the one-line diagram of the proposed FREEDM system, which integrates the advanced power semiconductor devices technology, power electronics technology, and information technology. The FREEDM system achieves seamless integration of distributed renewable energy generation and storage, therefore enables highly intelligent and distributed power grid architecture. Some key features of this innovative power grid are:

1. Plug and play of distributed renewable energy resources (DRER) and distributed energy storage devices (DESD).
2. Intelligent energy management (IEM) system through the revolutionary solid state transformer (SST) concept with direct interface to the distribution system.
3. Intelligent fault management (IFM) system with fast fault isolation capability by adopting the solid state fault isolation device (FID).
4. Distributed grid intelligence software embedded into the IEM and IFM.
5. Reliable and secured communication infrastructure to enable real time monitoring and control of all the nodes in the system.
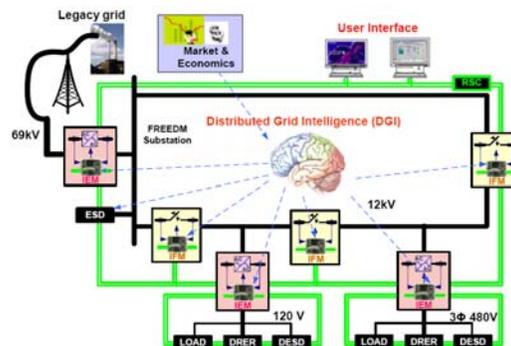
Figure 1. Electric grid diagram of the FREEDM system

In order to demonstrate the envisioned FREEDM system, recent efforts are made towards the living demonstration of a future home in the FREEDM system [2], as depicted in Figure 2. In this diagram, the DGI and SST are integrated to function as IEM. SST delivers 380V DC to the plug-and-play DC microgrid as well as 120/240V AC to the traditional AC load. FID is installed in the 12KV distribution line to isolate the fault current. Digital communication infrastructure based on DNP3 is also designed to enable the power management and condition monitoring of the whole system. Some breakthroughs have been achieved for the key components of such a system, including DGI, SST, DC microgrid, FID, and IEM.
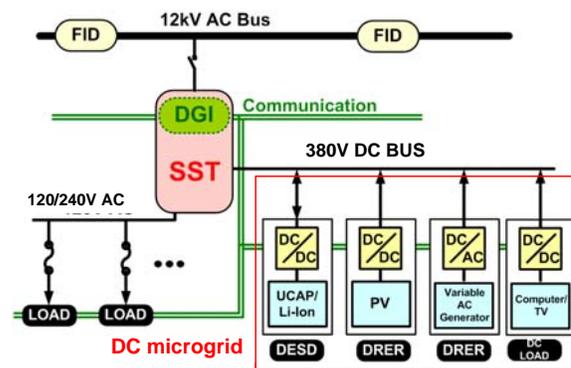
Figure 2. Electric grid diagram showing the interface of a future home in the FREEDM system

This paper summarizes and reports our newest progress in these areas. The hardware and software development of distributed grid intelligence was introduced in section II. In section III, two generations of solid state transformer are

presented with their characteristics highlighted. In section IV, the plug-and-play DC microgrid platform is illustrated. In section V, the development of Si based generation-I FID is presented. Lastly, the paper is concluded in section VI.

II. DISTRIBUTED GRID INTELLIGENCE

The DGI acts as the brain of the proposed FREEDM system. The diagram of DGI is shown in Figure.2, where green line represents the communication link, while black lines the electricity path. The main function of DGI is to collect information and manage power flow within a FREEDM hub or among multiple FREEDM systems over a wide area. Therefore, the hub is inherently distributed with components of the DGI algorithm executed on independent hardware within IEMs and IFMs. The DGI hardware is then required not only for localized control within each IEM and IFM, but also to implement control functions that must be handled collectively by groups of IEMs and IFMs. For DGI software, the main intelligence implemented is based on the multiple levels of control and time scales. The key challenge is to partition and distribute these controls as cooperating agents over multiple microprocessors within each IEM [3].

The distributed controller consists of a customized local controller that directly interacts with power converters and a commercial ARM board for communication implementation [4]. The division of hard real time and soft real time are therefore realized based on this hardware partitioning. A web based human machine interface (HMI) is also developed for system monitoring over the communication links. System operator can intervene the operation of SST by dispatching command through this HMI.

A. DGI Hardware Platform

The DGI hardware platform consists of three major components, namely, interface board, local controller board and ARM communication board, as shown in Figure 3. These three components can be stacked on top of each other to form a compact and reliable structure [4].
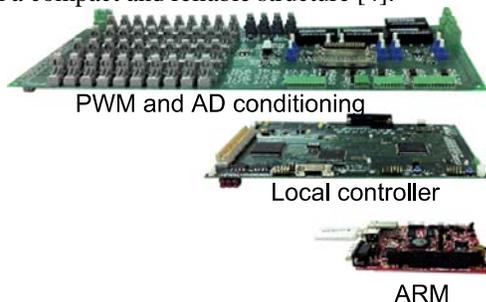
Figure 3. DGI hardware platform

The interface board (the top one in Figure. 3) is equipped with PWM output, analog conditioning circuit and fault protection. On the left side, there are 40 channels of optical fiber output to transmit switching signal to power devices.

These PWM signals are tied with fault protection signals in order to quickly turn off the power stage during fault situation. There are also 16 channels of analog conditioning circuit that has an output range of $-10V$ to $+10V$. Fault signals are generated using voltage comparators to compare the outputs of the conditioning circuit with tuned threshold values. The number of these channels is adequately designed for most of our converter applications in FREEDM, which means this hardware platform is universal and little modification is needed for new deployment.

Converter control is executed in local controller board (the middle one in Figure.3), which is able to work under both standalone and network modes. The circuit structure is based on combination of DSP and two FPGAs, as shown in Figure 4. With development of the IC technology, FPGAs have become the main stream in complex logic circuit design due to its flexibility, ease of use, and short time to market. In this structure, the major function of FPGA #1 is to implements PWM generation and interrupt the DSP periodically to evoke control loop. Once the DSP receives this external interrupt, an interrupt service route program is called and the control algorithm is executed. First step of the control algorithm is to sample converter voltage and current, this is done using FPGA #2 to generate AD control logic and pass the sampled value to DSP on data bus.
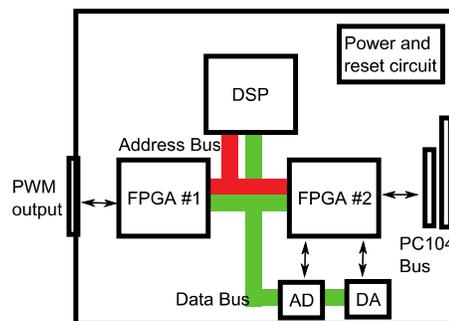
Figure 4. Functional diagram of Controller board

In order to transmit data to the ARM communication board, FPGA #2 is also implemented with a PC104 bus interface that can achieve fast and reliable data transmission. With the high performance bus access, the local controller is able to deliver measurement to and receive reference from the ARM communication board TS-7800 (bottom in Figure.3). The TS-7800 board features a Marvell 500MHz ARM9 SBC, which provides high end peripherals including gigabit Ethernet. The ARM board runs Linux 2.6 kernel with a full Debian distribution in on-board flash. PC104 bus is one of its external memory/program peripherals.

B. Design and Implementation of SST Communication

The objective of the communication system is two-fold: (1) the SST can send acquired measurement to the control center; (2) the control center can also send control commands and reference to the SST. To achieve such bidirectional communications between the SST and the

control center, we designed a systematic architecture as shown in Figure 5 [5]-[6]. Based on the hardware structure, the SST communication system is composed of two domains: network domain and device domain. In the network domain, a control center is connected to the distributed controller via a Local Area Network (LAN) over a DNP3 protocol stack. Thereby, the collected data and dispatched control commands can be delivered via DNP3 packets between devices. In terms of the device domain, it features aforementioned PC104 bus as data path to exchange data between the local controller and the ARM board.
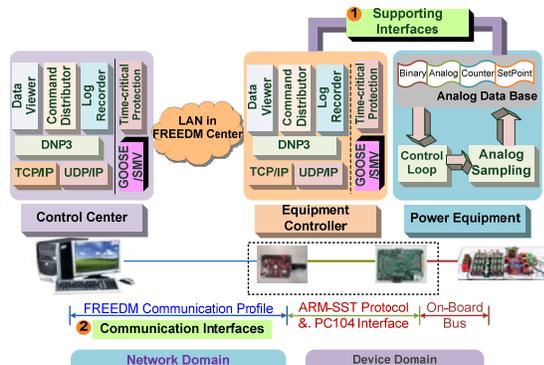
Figure 5. Communication schemes implemented at local DGI node

Since DNP3 over TCP/IP is employed as the underlying communication protocol, the design principal of network applications is to leverage basic DNP3 functions for the SST monitoring. We then introduce the network applications setup in a top-down manner. Left block in Figure. 5 depicts the application design in the network domain. Since DNP3 commonly works in a master/slave mode, the applications setup is correspondingly divided into two parts for the master and the slave. We envision three basic tasks on the master, including data query, command distribution and log collection, which are mapped into three modules, data viewer, command distributor and log recorder respectively. To support data query and modification in Data Viewer and Command Distributor, a database is programmed to store data from the device or issued by the Command Distributor. Two data channels are furnished under the database to retrieve data updates from the DNP3 function blocks, or dispatch commands to function blocks. As DNP3 is overlaid over TCP/IP in the communication layer model, the generated DNP3 packet will be fed to the slave via a TCP/IP channel. On the slave side, the application design is the same with that on the master.

III.   SOLID STATE TRANSFORMER

The SST is a power electronic device that replaces the traditional 50/60 Hz power transformer by means of high frequency transformer isolated AC-AC conversion technique, which is represented in Figure. 1 [7]. The basic operation of the SST is firstly to change the 50/60 Hz AC voltage to a high frequency one (normally in the range of

several to tens of kilohertz), then this high frequency voltage is stepped up/down by a high frequency transformer with significantly decreased volume and weight, and finally shaped back into the desired 50/60 Hz voltage to feed the load. In this sense, the first advantage that the SST may offer is its reduced volume and weight compared with traditional transformers.
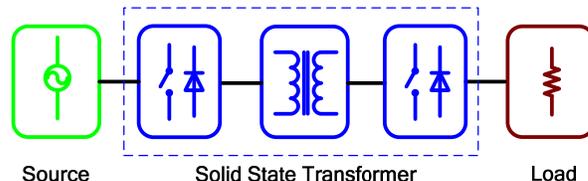
Figure 6. Solid state transformer configuration

It is further seen from the configuration of the SST that some other potential functionalities that are not owned by the traditional transformer may be obtained. First, the use of solid state semiconductor devices and circuits makes the voltage and current regulation a possibility; similarly to FACTS devices. This brings promising advantages such as power flow control, voltage sag compensation, fault current limitation, and others, which are not possible for traditional transformers. Therefore, the SST can enhance the stability and controllability of the power distribution grid. Second, voltage source converters connected from the secondary terminal of the SST could readily support a regulated DC bus, which could be connected to DC microgrids enabling the new microgrid architecture.

As a benchmark, a laboratory prototype of a 7.2 kV, 20 kVA single-phase, three-stage SST was built, which is named Generation-I (Gen-I) SST [8]. The topology of the Gen-I SST is illustrated in Figure 7.
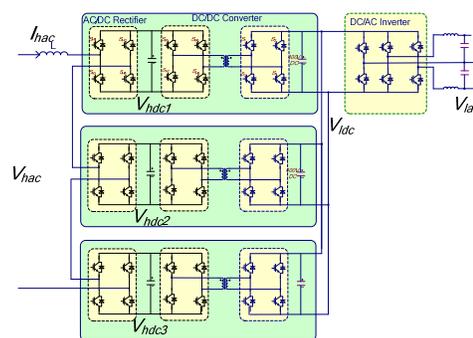
Figure 7. Topology of Gen-I SST based in Si IGBT

Cascaded multilevel rectifier with three H-bridges is adopted as the front-end stage. Three dual active bridge (DAB) converters are connected to each DC link with secondary side parallel as the DC/DC step down stage. In the last stage, the split phase inverter is adopted to provide both 120 and 240V AC.

The hardware prototype is shown in Figure .8. 6.5kV, 25A silicon based dual insulated gate bipolar transistor (IGBT) were customized and adopted for the rectifier and primary side of the DAB stages. It was packaged by POWEREX with chips supplied by ABB. 600V

commercial intelligent power modules (IPM) were adopted for the secondary side of the DAB and inverter stage. Based on the chosen power device, the switching frequency for the high voltage side rectifier, medium voltage DC/DC stage, and low voltage inverter stage was 1.2 kHz, 3.6 KHz, and 10.8 kHz respectively. A Metglas AMCC 1000 core was paralleled and used for the high voltage and high frequency transformer at an operating frequency of 3 kHz. The transformer was naturally cooled with the maximum temperature around 45 ^{0}C. Based on the detailed test results of the power device used in the power stage, PLECS simulations were used to estimate the efficiency of the converter, and the efficiency for this prototype at full power rating was close to 88 %.
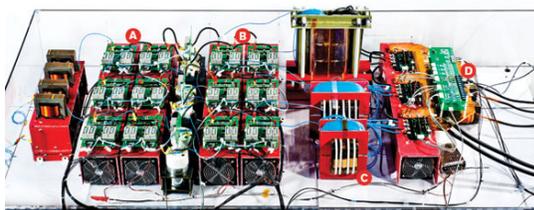
Figure 8. 20KVA three-stage SST based on 6.5KV Si IGBT

In order to realize the remote control and monitoring of the SST, the designed DGI hardware shown in Figure .3 is adopted. In this condition, this unit can be utilized as the IEM in the FREEDM systems. A HMI is developed for the whole FREEDM system, and Figure 9 shows the page for the SST control and monitoring. In this condition, all the data, including the command and the operating status information, can be transmitted between the DGI imbedded in the SST and the central controller server.
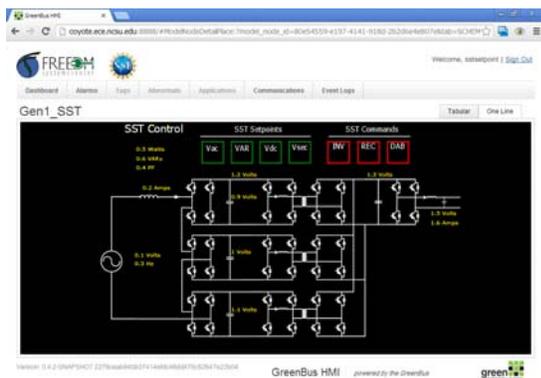
Figure 9. Control center webpage

Test result of the Gen-I SST is shown in Figure .10. The input AC voltage is 1.2KV, high voltage DC is 2.28KV, low voltage DC is 380V, and low voltage AC is 240V. The power rating is about 4KVA. The delivered 380V DC bus is readily for use to the plug-and-play DC microgrid.
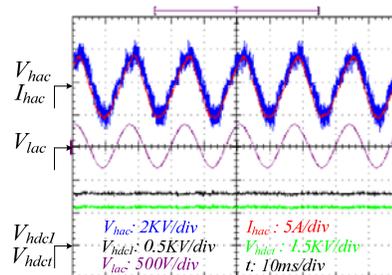
Figure 10. Test waveforms of Gen-I SST

The efficiency achieved was low mainly because of the relatively low power rating [9]. For MVA level design, the expected efficiency is higher than 95 %. In addition, the volume and weight of the SST was not effectively reduced compared with traditional transformers due to the relatively low operating frequency and large volume of power devices, heatsinks, and capacitors.

The performance improvement of the SST could be carried out in the following aspects:
1. Use of the SiC devices to increase the switching frequency as well as control the losses.
2. Use of nanocrystalline core to operate the transformer at a higher switching frequency.

Based on these considerations, another three-stage solid state transformer based on 15 kV SiC MOSFET was built, which was named Generation-II SST, as shown in Figure. 11.

(a) Front side: High voltage side; (b) Back side: Low voltage side
Figure 11. 20KVA SST based on 15KV SiC MOSFET

The newly developed 15 kV SiC MOSFET was adopted in the high voltage side and several low R_{dson} MOSFETs were paralleled to be used in the low voltage side [9]-[10]. A nanocrystalline core was used to assemble the 20 kVA high voltage transformer, which operates at 20 kHz. A certain soft-switching technique for the high voltage rectifier stage was adopted to solve the problem of high dv/dt introduced by the high switching speed. By using the measured loss data of the power devices, the estimated efficiency for the new Generation-II SST is close to 97 % [10]. In addition, its volume is about one third of the Generation-I SST. Therefore, the performance can be potential improved.

IV.   PLUG-AND-PLAY DC MICROGRID

The plug-and-play DC microgrid is one of the most attractive solution for future residential application due to its good scalability. As shown in Figure 12, the envisioned DC microgrid system is constructed. It includes: two DRERs (PVs) and DESDs (batteries), which are used to simulate multiple DRERs and DESDs, local loads and a SST, where the arrows show the modules power flow directions. Two unidirectional DC/DC converters are used for PVs to transfer power generated by PVs into the system, and two bidirectional DC/DC converters are used for batteries to balance the power differences between loads and the power supplied by PVs in islanding mode. The local load and the SST share a common 380V dc bus. The DC bus voltage can be enabled by the SST's output or be regulated by the battery's DC/DC converters. Correspondingly, when the SST disconnects from the microgrid system, the dc microgrid operates in the islanding mode; while the SST connects to the system, the microgrid enters into the SST-enabled mode. In SST-enabled mode, the batteries can be charged or discharged through their converters automatically based on their control algorithms.
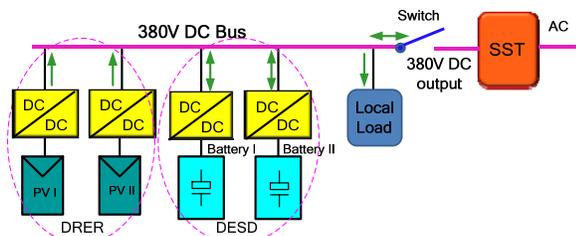
Figure 12. Diagram of plug-and-play DC microgrid

In the following discussion, the system's power management strategy is presented based on the Figure .11, including PV module control algorithm, battery control algorithm and plug and play function [11].

A. PV module control algorithm

The Maximum Power Point Track (MPPT) [12] control algorithm is shown in Figure 13.
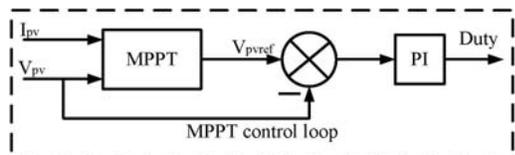
Figure 13. PV control algorithm.

$V_{pv}$is the PV panel's actual output voltage and V_{pvref} is the voltage reference, which is based on the MPPT curve. The difference of the two values as input to a PI controller is used to control the PV DC/DC converter.

B. Battery module control algorithm

To control each module distributed, integrate more DESDs into the system conveniently and implement the plug and play function without the communication ports involves, the droop control [13] for the battery is adopted. Comparing to the other methods, the droop control has its own advantages when it applies to microgrid system, either in islanding mode, SST-enable mode or transient mode, because the module can be controlled independently based on its local information (its output voltage and current) and no need to be aware of other modules' or system's information. Their droop curves are shown in Figure 14 (a). 380V is a critical value. When the DC bus voltage is greater than 380V, batteries are in charging mode; otherwise they are in discharging mode. 400V and 360V are the upper-limit and lower-limit values for the DC bus voltage. In islanding mode, all battery modules share the current to balance the power difference between the PV generation and load, and regulate the DC bus voltage according to their droop curves. Generally, the bigger rating battery, which has smaller slope will share more current than the smaller one. In SST-enabled mode, the DC bus is regulated by the SST and the batteries can be still in droop control. Battery modules are in charging or discharging mode, which are determined by the bus voltage enabled by SST. As a result, when system is in transient mode, whether the system switches from the islanding mode to SST-enabled mode or on the other way, there is no need to change the batteries control algorithm. The droop control diagrams for batteries are shown in Figure 14 (b).

(a)   (b)
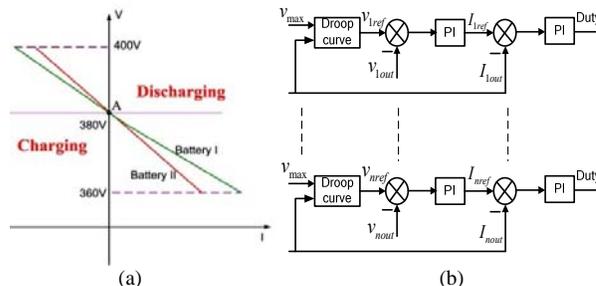
Figure 14. (a) Batteries droop curve; (b) Batteries control algorithm

C. Database

To supervise the operation of the whole system, a control center is constructed and the communication architecture is shown in Figure.15. Considering the DESD and DRER might be far from the control center, wire communication is unrealistic, thus a wireless communication port is adopted. The communication board (TS7800) [14] is interfaced with the each DSP board and transforms the SCI signals into WIFI signals, then the WIFI signals are transited to the control center. As a result, the control center of the DC microgrid collects the whole system information and a database is built to store all information. On the front-end of control center, Java Server Pages (JSP) are used to build up the webpage, which shows the real time data from each device and module's plug or unplug status, as shown in Figure 16(a). On the back-end of control center, a database is set up by Mosel to save all system history information in control center, as shown in Figure 16(b). This database can query the TS7800 board for the data based on the

customer's required period and then display the history data on the webpage. Furthermore, these data will automatically store into the database in the specified space, because each device has its own space to save data in database according to their IDs. In addition, since each device has its own ID in the control center, when the device re-plugged into the system, it can be recognized by the control center then its data can be saved into the corresponding space. Furthermore, the dynamic arrows are added on the webpage to show the module power flow direction and the battery arrow direction are bidirectional since it operates bidirectional as discussed in section III. The webpage of control center is shown as followings.

In order to verify the proposed power management strategy, experimental tests have been carried out on an experimental setup in a laboratory. A 380V DC microgrid is constructed, which consists of two PV panel modules, two batteries modules and one local resistor load bank, which share a common dc bus. A DC power supply is used to supply the DC bus voltage to emulate SST-enabled mode.
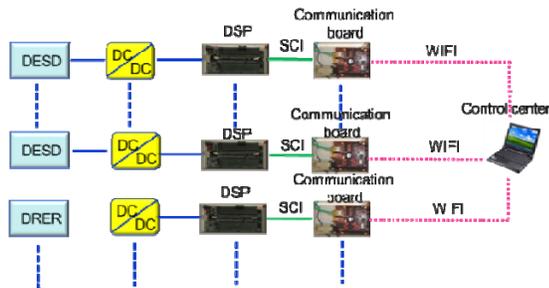
Figure 15. Communication architecture for DC microgrid

(a)

(b)
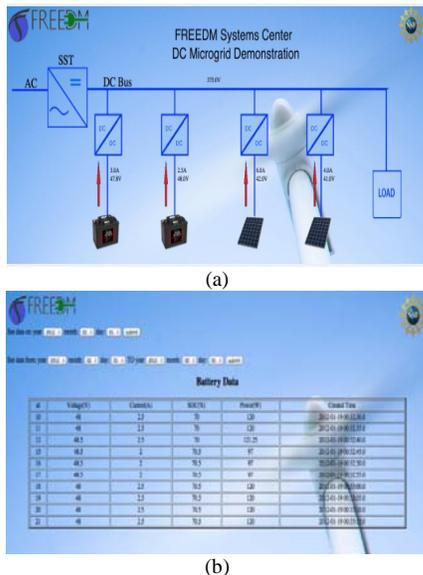
Figure 16. (a) Control center webpage; (b) system database.

Case I: islanding mode to DC source-enabled mode

Experimental results of the transition from islanding mode to DC source-enabled mode are shown in Figure 17.

(a) and Figure 17.(b). Initially the system is in the islanding mode and the DC bus voltage is regulated by the batteries. Based on the Figure 17 (b), the PV panels' output power is almost equal to the load needed, so the bus voltage is 380V and the average values of batteries output currents are almost zeros. When the system connects to the DC source, there is no obvious change of the DC bus voltage, but the batteries output current have a very small drops because the DC source output voltage has a small oscillation.

(a)   (b)
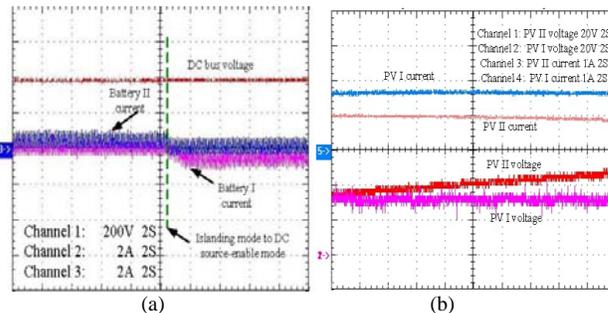
Figure 17. (a) Battery currents and bus voltage; (b) PV currents and voltages

Case II: DC source-enabled mode to islanding mode

Figure 18(a) and Figure 18(b) show the experimental results of the transition from DC source-enabled mode to islanding mode.

(a)   (b)
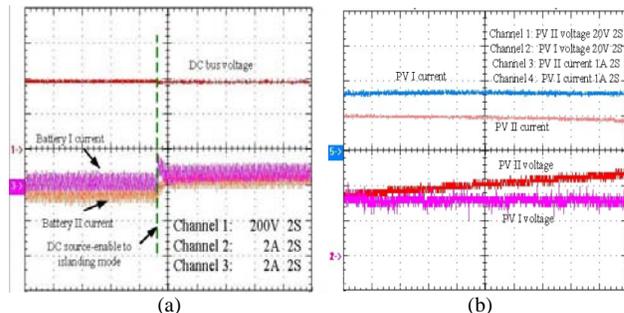
Figure 18. (a) Battery currents and bus voltage; (b) PV currents and voltages

When system is in the DC source-enabled mode, the DC bus voltage is regulated by the DC source. Since the DC source's output voltage is 380V so the batteries are almost in stand-by status (output current average value is zero). When the system goes into the islanding mode, batteries are responsible to regulate the bus voltage. Since the power generated by PVs is less than the load needed, the DC bus voltage is a little bit smaller than 380V and the batteries output very small current.

V.   SOLID STATE FAULT ISOLATION DEVICE

FID is an indispensable part of FREEDM system. The FID must provide ultra-fast isolation to fast restore the system voltage and therefore to minimize the required ride through capability of the system. It requires for a very fast operation (faults clearing within a few μS) in this power electronics based system. In the developed hardware test

bed, the FID is in series with a 12 KV distribution system. Considering the limitation of the available high voltage IGBT (6.5KV rating), the adopted topology consists of three IGBT modules connected in series, with each module containing two emitter-connected 6.5KV, 200A IGBTs together with free-wheeling diodes [14]. To absorb the over voltage caused by high di/dt in the turn-off process, mental oxide varistor (MOV) is adopted shown in Figure 19.
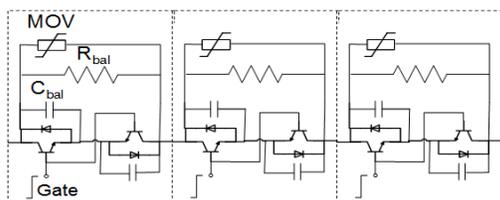
Figure 19. Gen-I FID with series connected IGBT

Test results of the developed FID prototype interrupting a 4A (peak) current in a single phase 7.2KV system is demonstrated in Figure 20. Figure 20 (a) shows the fast turn off process of the current and figure 20 (b) shows the re-closer of the FID when the fault is cleared. It is shown that the developed FID can interrupt the fault current in a short time.

(a)   (b)
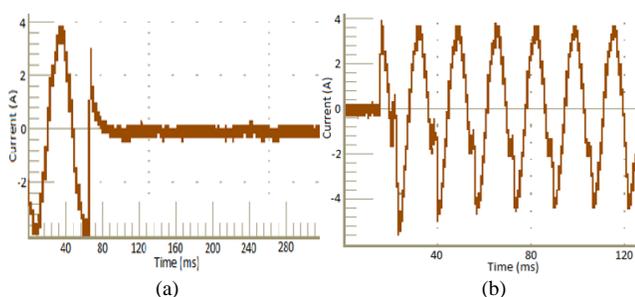
Figure 20. Test results of FID in the single phase 7.2 KV line

The Gen-II FID in FREEDM system will be focused on adoption of SiC power devices, such as SiC Emitter Turn-off Thyristor (ETO), together with hybrid switch concept to future reduce the on-state voltage, and therefore minimize the losses of the FID in the normal operation.

VI. CONCLUSION

This paper presents the promising future power distribution system architecture: the FREEDM system. As a major step towards the envisioned FREEDM system, this paper summarizes the recent development of some key technologies, including distributed grid intelligence, solid state transformer, plug-and-play DC microgrid, and fault isolation device. The proposed FREEDM system is an efficient electric power grid integrating highly distributed and scalable alternative generating sources and storage with existing power systems. Therefore, to facilitate a green and sustainable energy based society, mitigate the growing energy crisis, and reduce the impact of carbon emissions on the environment.

REFERENCES

[1] A. Q. Huang., M. L. Crow, G. T .Heydt, J. P .Zheng, and S. J .Dale, "The Future Renewable Electric Energy Delivery and Management system: The Energy Internet," in Proc. IEEE, vol.99, no.1, Jan. 2011, pp. 133-148.
[2] X.She, A.Q.Huang, S. Lukic, and M. Baran, "On integration of solid state transformer with zonal DC microgrid," IEEE Trans. Smart grid, vol.3, no.2, June 2012, pp.975-985.
[3] M.L.Crow, B.Mcmillin, W.Y.Wang, and S.Bhattacharya, "Intelligent energy management system of FREDM system," in Proc.IEEE.PES general meeting, 2010, pp.1-4.
[4] F.Wang, X.Lu, W.Y.Wang, and A.Q.Huang, "Development of Distributed Grid Intelligence Platform for Solid State Transformer," in Proc.IEEE Smart Grid Comm., 2012, pp.481-485.
[5] X.Lu, W.Y.Wang, A.Juneja, and A.Dean, "Talk to transformer: An empirical study of device communications for the FREEDM system," in Proc.IEEE.Smart Grid Comm., 2011, pp. 303-308.
[6] X.Lu, Z,Lu, W.Y.Wang, and J.F.Ma, "On network performance evaluation toward the smart grid: A case study of DNP3 over TCP/IP," in Proc.IEEE. Global Comm., 2011, pp.1-6.
[7] X.She, R.Burgos, G.Y.Wang, F.Wang, and A.Q.Huang, "Review of solid state transformer in the distribution system: from implementation to filed application," in Proc.IEEE.ECCE, 2012, pp.4077-4084.
[8] G.Y.Wang, S.Baek, J.Elliott, A.Kadavelugu, F.Wang, X.She, and A.Q. Huang, "Design and Hardware Implementation of Gen-1 Silicon Based Solid State Transformer,"in Proc.IEEE.APEC. 2011, pp.1344-1349.
[9] G.Y.Wang, A.Q.Huang, and C.S.Li, "ZVS range extension of 10A 15KV SiC MOSFET based 20KW Dual Active Half Bridge (DHB) DC-DC converter," in Proc.IEEE ECCE 2012, pp. 1533-1539.
[10] Annual report of Future Renewable Electrical Energy Delivery and Management system center, US, 2012.
[11] X.W.Yu, F.wang, and A.Q.Huang, "Power management strategy for plug and play DC microgrid," in Proc. IEEE.PES.ISGT., 2012, pp.1-6.
[12] N. Femia, G. Petrone, G. Spagnuolo and M. Vitelli, "Optimization of Perturb and Observe Maximum Power Point Tracking Method,"IEEE Trans. Power Electron., vol. 20, no. 4, 2005, pp. 963-973.

An Iterative Price-Based Approach for Optimal Demand-Response

Michalis Kanakakis, Marilena Minou, Costas Courcoubetis, George D. Stamoulis, George Thanos

Department of Informatics

Athens University of Economics and Business

Athens, Greece

{kanakakis, minou, courcou, gstamoul, gthanos}@aueb.gr

Abstract— **Demand Response (DR) programs encourage consumers to adjust their power consumption in response to DR events such as changes in electricity prices or sudden peaks in demand. While significant savings can be thus achieved, the real success of DR programs depends on the incentive compatible participation of consumers and their timely response to DR signals. In this work, we consider households that operate different types of appliances consuming electricity loads characterized by preferences on the time and quantity of the consumption. We develop a distributed optimization algorithm of practical interest, which offers both to the providers and the consumers of electrical power the opportunity to jointly interact and converge to the optimal scheduling of their appliances: this optimizes the trade-off between user-savings with the inconvenience caused by shifting consumption in time and shading load, while leading the cost of the utility company to the socially optimal level. Our approach is based on related work, and solves two key practical issues: i) we convert the consumer reaction problem to the solution of an LP using standard LP solvers, and ii) we propose a heuristic for load consolidation, i.e., regroup appliance operation on adjacent timeslots, using a randomized algorithm. We also present a substantial amount of simulation data that suggests that our heuristics are sound and reach quickly a state relatively close to the actual system optimum.**

Keywords-Demand response; dynamic pricing; energy consumption scheduling; incentives; smart grid

I.   INTRODUCTION

Demand Response (DR) programs encourage electricity end-consumers to adjust their consumption in response to DR events, such as changes in electricity prices surges in demand. Energy use in buildings, including commercial and public buildings, represents approximately 40% of total final energy consumption and 36% of CO2 emissions in Europe [1]. Space heating accounts for 67 % of household energy consumption in the EU27 and is followed by water heating (14%) and appliances/lighting (13%) [2]. Therefore, one expects significant savings to be achieved by DR. Although DR programs have been implemented in industrial environments so far, their penetration in the residential sector can result in considerable savings, due to the fact that such environments account for a large portion of the total energy demand. However, the real success of residential DR programs depends on the incentive compatible participation of consumers and their timely response to DR events.

In this work, we focus on the derivation of electricity prices for different times in a given time frame (day or week or month) by means of a mechanism that solves the social (global) welfare optimization problem, i.e. maximizes the aggregate benefit to the consumers minus the cost of provisioning. The idea is very simple and is borrowed from [3].

It defines a market for electricity consumed at different times, while prices are set to marginal costs. Since such goods are substitutable, the right prices offer the incentives for shifting and shading consumer demand, and at the equilibrium we obtain economic efficiency. Our contribution is to design the consumer response algorithms in a way that leads to simple implementations and that also capture practical constraints on the appliance side, which are hard to capture in the microeconomic model of the market proposed by [3].

Our approach presumes an infrastructure supporting consumption data gathering and processing. This information is used to reveal each consumer's preferred consumption pattern and her disutility for changing it. It has to take into account historical demand behaviours and other data revealing the context that affected the consumption (weather conditions, social events, etc.). Utilizing this knowledge, the provider may estimate the right DR signals, e.g., prices at a given time zone or timeslot. Our mechanism can be applied under multiple contexts in a straightforward manner, e.g., we differentiate weekends from weekdays, and run two separate interleaved instances of our algorithm.

Many approaches in the literature, see [2], [5-8], define algorithmic approaches that mostly focus on solving each consumer's personal optimization problem. For instance, [2] proposes a residential energy consumption scheduling framework, which attempts to achieve a desired trade-off between minimizing the electricity payment and minimizing the waiting time for the operation of each appliance in household under a real-time pricing tariff, using price prediction based on prior information. Also in [8] the authors propose a multifunctional system named "Yupik", which solves the user-side optimization problem and presents the right choices to the users by using an appropriate interface. The model involves simple linear disutility functions for delaying the scheduling of certain appliances, with parameters obtained by monitoring the history of the operation of these appliances.

Solving the global optimization problem involves the participation of the utility company. By introducing a distributed global optimization algorithm, [3] offers both the utility company and the consumers the opportunity to jointly compute the optimal prices and the demand schedules for these prices, by means of an iterative procedure. The model considers households that operate different appliances including PHEVs and batteries; the appliances are divided into types according to their operational characteristics and are described by concave utility functions and linear constraints.

Our contribution is in remedying some of the shortcomings of the approach in [3]. One concerns the complexity in solving the optimal load schedule selection from the part of the consumer. Although the problem is the one of maximizing a set of concave functions over a set of linear constraints, this requires in general the use of computationally complex gradient projection methods. We propose a simpler way to solve this problem exploiting the special property that the optimum is always on the boundary of the constraint set. This involves the local linearization of the utility functions and the use of a standard LP solver, which allows us to easily handle a general form of linear constraints. LP solvers are remarkably efficient in practice and their worst case exponential complexity is acceptable for small scale residential problems.

A second problem concerns the micro-economic modelling of appliances like washing machines, water heaters, etc., which can shift load between different timeslots. Some of them like the water heaters can consume electricity load on non-adjacent timeslots. However, other appliances like washing machines, need their consumption to be better consolidated and use a single or more adjacent timeslots. The mathematical analysis of a market where there are such combinatorial constraints on the consumption of goods is not a standard one and we don't know of simple iterative methods that work. This is the reason why our basic appliance models cannot address such load consolidation constraints and hence stay in the "computationally effective" world of simple linear constraints. This has always been a problem in models such as in [3], which trade modelling accuracy with tractability.

To tackle the above problem, we deploy a simple heuristic. Say that we have *n* consumers of the same type, which need to schedule an appliance (say a dishwasher) that requires consolidation (must run in a single time slot), and the optimal schedule for each consumer is to use 2 units of load in timeslot 1 and 1 unit in timeslot 3. Suppose we let each consumer randomly choose among the slots 1 and 3 with probabilities proportional to the required loads, i.e., 2/3 and 1/3, and run the appliance in this chosen slot using the total of the scheduled load (in this case 3 units). Then for large *n*, by the law of large numbers, at each time slot the expected value of aggregate load generated by the *n* consumers would be the same as the aggregate load in the non-randomizing case, with a relatively small variance. This load consolidation heuristic is transparent to the utility company and also offers practical schedules for the consumers. We have resorted in extensive simulations experiments to assess the speed of convergence and the efficiency of the resulting equilibrium. Our results suggest that our method has excellent convergence properties and also leads to essentially the same social optimum as the exact model, with negligible variations.

This paper is organized as follows. In Section II, we present the basic model for the utility company and the consumers and in Section III, the models for appliances and the associated utility functions, all based on [3]. In Section IV, we present our heuristics for consumers' reaction and in Section V, the experimental results validating our approach. Finally, in Section VI, we give some concluding remarks.

II. THE MODEL

As we mentioned earlier, our approach is based on the work in [3]. We present below the complete model borrowing the notation from the above paper. We describe how the utility company sets its prices dynamically, how each consumer responds, and the properties of the resulting equilibrium operating point. The basic idea is to split time into time slots and create a market for electricity in each slot. At the market equilibrium we achieve economic efficiency as expected. We exploit the inherent distributed decision-making that such a market model offers.

To make things more concrete, one can think that our basic frame of time consists of a single day. As we just mentioned, each day is divided into T timeslots of equal duration (say hours), indexed by $t \in T := \{1,2,...,T\}$. During each day the seller (utility company) posts prices for the different slots and the consumers choose their loads. This process iterates daily until equilibrium is reached. We use k to denote the kth day of the iterative process.

A. System Model

Consider a set N of households/consumers that are served by a single utility company. Our system consists of the utility company and the consumers.

1) The Utility Company:

The utility company serves as an intermediary that participates in multiple wholesale markets, including day-ahead, real-time balancing and ancillary services, to provision enough electricity to meet the demands of the N consumers. We assume that cost of electricity to the utility company is summarized by the cost function $C(Q,t)$ that specifies the cost to provide Q amount of power to the N consumers at time t, where the cost function $C(Q,t)$ is taken as convex and increasing in Q for each t. The utility company sets the prices $(p(t), t \in T)$ as the marginal cost at the total consumption level; the optimal such level and the corresponding prices are discovered by means of an algorithm described below. Setting prices equal to marginal costs does not guarantee complete cost recovery by the utility company. We implicitly assume that such issues are solved by a fixed part in the user tariffs.

2) The Consumers

Each consumer $i \in N$ operates a set A_i of appliances such as air conditioner, refrigerator, plug-in hybrid electric vehicle (PHEV), etc. (In [3] it is taken that consumers can also operate batteries and store power; for simplicity, we do not take this feature into account.) For each appliance $a \in A_i$ of consumer i, $q_{i,a}(t)$ denotes its load drawn at time $t \in T$ and $q_{i,a}$ the vector $(q_{i,a}(t), t \in T)$ of loads drawn over the whole day. An appliance a is characterized by two parameters:

- A utility function $U_{i,a}(q_{i,a})$ that quantifies the utility consumer i obtains when it consumes power according to the vector $q_{i,a}$; and,
- a set of linear inequalities $A^{i,a} q_{i,a} \le \eta_{i,a}$ on the power vector $q_{i,a}$ corresponding to technical constraints.

In Section III, we will describe in detail how [3] models various appliances through appropriate matrices $A^{i,a}$ and vector $\eta_{i,a}$. Note that an inelastic load, e.g., minimum refrigerator power, can be modelled by $q_{i,a}(t) \ge \underline{q_{i,a}}$, which implies that the appliance a of consumer i requires a minimum power $\underline{q_{i,a}}$, at all times t. This is a linear constraint and part of the set of constraints $A^{i,a} q_{i,a} \le \eta_{i,a}$.

At each time t the total power demand of consumer i is

$$Q_i(t) := \sum_{a \in A_i} q_{i,a}(t). \tag{1}$$

3) System Problem

Suppose a central planner has the objective to maximize the economic efficiency of the complete system consisting of the utility company and the consumers. Then she would solve

$$\max_q \quad \sum_i \left(\sum_{a \in A_i} U_{i,a}(q_{i,a}) \right) - \sum_t C\left(\sum_i Q_i(t) \right) \tag{2}$$

$$\text{s.t.} \quad A^{i,a} q_{i,a} \le \eta_{i,a}, \ \forall a,i \tag{3}$$

$$0 \le Q_i(t) \le Q_i^{max}, \ \forall t,i \tag{4}$$

where $Q_i(t)$ is defined in (1), the inequality (3) models the various consumer appliances (see Section III for details), the right-hand inequality of (4) imposes a bound on the total power drawn by consumer i at each slot.

By assumption the objective function is concave and the feasible set is convex, and hence there is a unique optimal point, which can in principle be computed centrally by the utility company. This, however, would require the utility company to know all the consumer utility and cost functions and all the constraints, which is clearly unrealistic.

A simple property of the optimal solution of (2) is that if each consumer was charged at each slot the marginal cost of the cost function evaluated at the optimal load vector, then she would also consume the amount of optimal load specified by (2). This suggests a distributed solution of (2) in terms of a market mechanism where the utility company posts prices and the consumers react by changing their demands in each slot in order to maximize their net benefits. At the *equilibrium* we expect the utility company to post prices $p := (p(t), t \in T)$, which induce the consumers to individually choose the consumptions q_i that correspond to the solution of (2), and these prices are the marginal costs that correspond to the aggregate of the above consumptions.

4) Distributed Algorithm

The distributed algorithm involves iterations between the consumers and the utility company.
a) Utility company: in step of the iteration, the utility company posts the price vector p that corresponds to the marginal cost of the aggregate load vector of the previous step.
b) Consumers: During each iteration, given the prices p posted by the utility company, each consumer i chooses her own power demand for all her appliances $q_i := \big(q_{i,a}(t), \forall t, \forall a \in A_i\big)$ so as to maximize her net benefit; namely, the difference of the total utility of operating each appliance a at power levels given by the vectors $q_{i,a}$ minus the total cost of electricity.

To avoid oscillations, *each consumer moves slowly her consumption in the direction of the optimal consumption*, i.e., she does not consume this optimal amount immediately but chooses an amount that corresponds to a weighted average of her old consumption and the optimal one. If this does not belong to the feasible set defined by the inequalities (3) and (4), then a projection of the aforementioned weighted average to this set is ultimately taken. It is also realistic to assume that consumers would not switch directly to the optimal consumption schedule but rather take it into account in modifying their previous schedule. An alternative approach to avoid oscillations, is to have consumers use their optimal load and prices posted by the utility company move slowly in the directions of the true marginal costs. This latter case might not be acceptable by the utility company because it may not recover its costs in the case of simple linear cost structures.

Therefore, during each iteration consumer i solves her *local problem*:

$$\max_{q_i} \sum_{a \in A_i} U_{i,a}\big(q_{i,a}\big) - \sum_t p(t) Q_i(t) \qquad (5)$$

$$\text{s. t. } (3) \text{ and } (4) \text{ apply.}$$

Note that the optimal solution of each consumer i depends on the prices $p := (p(t), t \in T)$ set by the utility company. Thus, it is denoted by $\big(q_i(p)\big) := \big(q_{i,a}(t; p), \forall t, \forall a \in A_i\big)$.

The prices p and the total consumer demand $q := (q_i, \forall i)$ are in equilibrium if $q = q(p)$, and $p = C'(q)$. At this point, we know from the first order conditions of (2) that the social welfare given by (2) is also maximized. One can show that starting from an arbitrary initial operating point, the iterative procedure mentioned earlier alternating between the utility company and the consumers converges to the unique solution of (2). As already mentioned, an important requirement for convergence is that consumers don't do abrupt load changes and move slowly towards their optimal choices.

III. APPLIANCE MODELS AND INTEGRITY CONSTRAINTS

We briefly discuss the appliance models proposed in [3]. The basic idea is to model appliances as accurately as possible while preserving the computational tractability of the resulting optimization problem, namely the concavity of the function to maximize and the linearity of the constraints.

The consumer appliances are classified into four types; each type is characterized by a utility function $U_{i,a}\big(q_{i,a}\big)$ that models how much consumer i values the consumption vector $q_{i,a}$, and a set of linear constraints on the consumption vector $q_{i,a}$, which are all involved in the optimization problems defined in Section III. Below, we mostly focus on the so-called type 2 of appliances, which exhibit important load-shifting aspects and their accurate modelling is hard (see our following discussion on this topic).

In particular, this type includes the appliances such as PHEV, dishwasher and washing machine. For these appliances, a consumer only cares about whether the task is completed before a certain time. This means that the cumulative power consumption by such an appliance must exceed a threshold by the deadline (e.g., see [4]). Let $A_{i,2}$ denote the set of type 2 appliances of consumer i. For each $a \in A_{i,2}$, $T_{i,a}$ is the set of times where the appliance can operate. For instance, for PHEV, $T_{i,a}$ is the set of times that the vehicle can be charged. For each consumer i and each appliance $a \in A_{i,2}$, we have the following constraints on the load vector $q_{i,a}$:

$$q_{i,a}^{min}(t) \leq q_{i,a}(t) \leq q_{i,a}^{max}(t), \qquad \forall t \in T_{i,a} ,$$
$$q_{i,a}(t) = 0, \ \forall t \in T \setminus T_{i,a} ,$$
$$\bar{Q}_{i,a}^{min} \leq \sum_{t \in T_{i,a}} q_{i,a}(t) \leq \bar{Q}_{i,a}^{max} , \qquad (6)$$

where $q_{i,a}^{min}(t)$ and $q_{i,a}^{max}(t)$ are the minimum and maximum power loads that the appliance can consume at time t, and $\bar{Q}_{i,a}^{min}$ and $\bar{Q}_{i,a}^{max}$ are the minimum and maximum total power draw that the appliance requires. By setting $q_{i,a}^{min}(t) = q_{i,a}^{max}(t) = 0$ for $t \in T \setminus T_{i,a}$, we can rewrite the constraints of (6) as

$$q_{i,a}^{min}(t) \leq q_{i,a}(t) \leq q_{i,a}^{max}(t), \qquad \forall t \in T$$
$$\bar{Q}_{i,a}^{min} \leq \sum_{t \in T_{i,a}} q_{i,a}(t) \leq \bar{Q}_{i,a}^{max} \qquad (7)$$

The overall utility that the consumer i obtains from a type-2 appliance a depends on the total power consumption by a over the whole day. Hence the utility function in the form introduced in Section II, reduces to:

$$U_{i,a}(q_{i,a}) := U_{i,a}\big(\sum_t q_{i,a}(t)\big) := U_{i,a}\big(Q_{i,a}\big) , \qquad (8)$$

where $$Q_{i,a} = \sum_t q_{i,a}(t) . \qquad (9)$$

It is assumed that this utility function is a continuously differentiable concave function of $Q_{i,a}$.

A key observation is that the above model satisfies the key properties required for solving it effectively (concavity of the utility function, linear constraints), but it is not realistic for most typical devices we like to model since it allows for such a device to be scheduled in non-adjacent slots and use a fraction of the total load in each slot. For instance, a washing machine requiring 2 units of load could be scheduled to use 2/3 units of load in slots 1, 3, 5. This might be practically impossible. Hence in practice, type 2 devices may have *integrity constraints*. These constraints dictate that power consumption should be in adjacent slots, i.e., the set of timeslots $T_{i,a}$ where the power is positive should constitute a single time interval. This is a combinatorial requirement and clearly destroys the linearity of our constraints making the computational solution very expensive.

Our methodology to solve this problem is a heuristic that uses the solution of the problem without these integrity constraints, and

at each step it aggregates the load at each consumer's premise in a way that looking at the aggregate load generated by all consumers it remains essentially the same without the consolidation. For simplicity we assume that each type 2 device needs to be scheduled in a *single* slot. This is a realistic assumption for residential devices, as each slot's duration offers adequate time for the accomplishment of their task. As we discuss in detail in the relevant section, each consumer picks a slot at random with probability proportional to the components of the load vector in the solution of (5) and runs the device at that slot using the total load.

Our experiments also comprise the so-called type 3 of appliances, which includes appliances such as lighting that must be on for a certain period of time. A consumer cares about how much light can get at each time t. Thus, user utility from a particular consumption vector $q_{i,a}$ is the sum of utilities due to the consumption in each timeslot; that is $U_{i,a}(q_{i,a}) := \sum_t U_{i,a}(q_{i,a}(t),t)$, subject to constraints on minimum and maximum power loads that the appliance can consume at time t.

IV.   HEURISTICS FOR CONSUMER REACTION

The local problem to be solved by each consumer is the maximization of a concave function with linear constraints, and is traditionally solved by a gradient projection type of algorithm. This makes sure that in all iterations we move without violating constraints and stay in the interior of the feasible region. In the case of general linear constraints these algorithms are computationally expensive to run. In [3] the gradient projection algorithm used in the numerical experiments is simple (we can consider each appliance individually) because the constraints across appliances (given by (4)) do not appear to be taken into account, and thus the problem obtains a special decoupling structure. This is not anymore the case if we do consider constraints on the total power consumed by all appliances at each slot.

Our idea for addressing such general constraints is as follows: at each iteration, we use a standard LP solver to find the optimum of the linearized problem of the consumer at the previous operating point. Then, we derive the updated consumption schedule by moving in this direction using a decreasing step size. This simple computational procedure works well and converges to the optimum, which in our case is always on the boundary of the simplex.

Instead of adopting the above sequence of solutions, we introduce an approximation by means of our load consolidation heuristic mentioned previously. In particular, in order to satisfy the practical integrity constraints and group the load consumption of type 2 appliances in a single slot, we use a randomization procedure that alters the load profiles proposed by the optimization algorithm. The law of large numbers suggests that in the case of many consumers of similar types, the aggregate load after the randomization has small fluctuations around its mean, which is the non-randomized aggregate load. Hence, when enhanced with this randomizing heuristic, our algorithm displays similar convergence properties and converges to the same equilibrium. Note that the heuristic can also be combined with the approach of [3], which does not conform to the practically applicable load integrity constraints.

Note also that these two heuristics may be applied either unilaterally or together, as each one of them offers a discrete benefit to the optimization procedure. However, the load consolidation heuristic may not be always feasible to be applied on each occasion, as it may violate the constraints across the appliances (see Section V-C). We summarize now these ideas in detail. In what follows, k denotes the iteration step.

A. *Linearization of Concave Utility Functions*

To simplify the presentation, we restrict attention to appliances of type 2. The utility function that represents the benefit obtained by

their use is given by equations (8) and (9). The local problem is again the optimization of each consumer's net benefit. However, at each iteration k, we resort to a linear approximation of this concave utility function that depends on the current operating point, thus obtaining:

$$\tilde{U}_{i,a}^k(q) = \gamma_{i,a}^k q , \qquad (10)$$

where $\gamma_{i,a}^k$ is the slope of the concave utility function $U_{i,a}(.)$ at the total consumption of this appliance at current iteration.

Therefore, the consumer's problem that we solve at each iteration k is as follows:

$$\max_{q_i} \ \sum_{a \in A_i} \tilde{U}_{i,a}^k(q_{i,a}) - \sum_t p^k(t) Q_i^k(t) \qquad (11)$$

s. t.   (3) and (4) apply.

Let $\tilde{q}_{i,a}^{*k}$ be the solution of the LP (11). We use "tilde" to denote the quantities related to the linearized local solution. The new consumption vector will be now computed as:

$$\tilde{q}_{i,a}^k = g^k * \tilde{q}_{i,a}^{*k} + (1 - g^k) * \tilde{q}_{i,a}^{k-1} , \qquad (12)$$

where $$g^k = \frac{g}{\sqrt{k}}, \ g < 1.$$

The selection of g influences how fast we move in the direction of the optimal load selection from the linearized local problem. As we discussed before, the actual load consumption proposed by our method is obtained by applying the consolidation algorithm to the solution $\tilde{q}_{i,a}^k$, as explained next. The formulation of the local problem as an LP (11) allows us to use a standard LP solver and thus to easily handle the various linear constraints. Therefore, we maintain feasibility in each iteration, contrary to the outcome of the gradient projection method of [3], which if necessary is projected within the feasible set at the end of each iteration.

B. *Load Consolidation*

Each consumer i chooses independently from others to sum-up the aggregate consumption $\tilde{Q}_{i,a}^k = \sum_t \tilde{q}_{i,a}^k$ for each type 2 appliance he operates, at one and only slot t. This slot is selected randomly; in particular, the probability that slot t is selected is $P_{i,a}^k(t)$, which equals the ratio of consumption $\tilde{q}_{i,a}^k(t)$ at this slot over the aggregate consumption $\tilde{Q}_{i,a}^k$, i.e., $P_{i,a}^k(t) = \frac{\tilde{q}_{i,a}^k(t)}{\tilde{Q}_{i,a}^k}, t = 1,2,\dots,T$. Hence, the actual consumption of each type 2 appliance a of a consumer i is zero in all slots except for the selected slot where the consumption equals $\tilde{Q}_{i,a}^k$. Let $\hat{q}_{i,a}$ denote the resulting loads. We use "hat" to denote the load quantities after linearization *and* load consolidation. The actual proposed reaction of consumer i corresponds to $\hat{Q}_i^k$ and we assume that a consumer always follows the above proposed schedule. Then, the utility company computes and posts its prices for the next iteration as the marginal cost corresponding to this consumption, i.e., $p^{k+1}(t) = C'\left(\sum_i \hat{Q}_i^k(t)\right)$.

The heuristic should be applied only if load consolidation does not lead to violation of the bound on the maximum total consumption in any timeslot t, which we assume in the sequel.

V.   NUMERICAL EXPERIMENTS

In this section, we present the results of a variety of experiments to illustrate the application of the proposed methodologies. First, we present the basic input parameters for the experiments, which in certain cases are partly differentiated, as explained accordingly.

Consider a system with N households (consumers) entering a DR program. Each household is assumed to have one type 2 and

one type 3 appliances, e.g., washing machine and lighting. Each type 2 appliance may operate at each slot and as already mentioned the respective consumer is interested in the aggregate consumption throughout the whole day. For the type 3 appliances, consumers are only interested during night time. As in the experiments of [3], each day is divided into 8 timeslots of equal duration (i.e., 3 hours), with the first slot corresponding to the time-zone 8-11am. Thus, type 3 appliances only operate in the slots 5-8. The basic parameters used in simulations are as follows:

For each type 2 appliance, $\bar{Q}_{i,A2}^{min}$ and $\bar{Q}_{i,A2}^{max}$ are chosen randomly and uniformly from [1400Wh, 1600Wh] and [2400Wh 2500Wh] respectively, while the values of $q_{i,A2}^{min}(t)$ and $q_{i,A2}^{max}(t)$ are 0Wh and $\bar{Q}_{i,A2}^{max}$ Wh respectively. The utility function is assumed to be logarithmic of the form:
$U_{i,A2} = w_{i,A2} \ln(Q_{i,A2})$, where $w_{i,A2}$ is randomly and uniformly distributed in the range [7700, 12320] for each consumer i.

For type 3 appliances, the minimum and maximum per slot working power requirements are 200Wh and 800Wh respectively. The utility function is assumed to be logarithmic of the form:
$U_{i,A3}(t) = w_{i,A3}(t) \ln(q_{i,A3}(t))$, where $w_{i,A3}(t)$ for each consumer i, is randomly and uniformly distributed in the range [3800, 3900] for $t = 5,6,7,8$, while $w_{i,A3}(t) = 0$, $t = 1,2,3,4$.

We also assume that the electricity cost function is of the form:
$C(Q) = cQ^2 + bQ + a$, thus $p(t) = C'(Q(t)) = 2cQ(t) + b$ and we use the values $c = 1/200N$ and $b = 0.8$ for each experiment. Except for Section V-C, we do not consider an upper bound on the total consumption per household (given by (4)).

A. Evaluation of the Linearization Method

First, we apply the linearization method (only) and evaluate it compared to the algorithm introduced in [3], with regard to: the convergence state, the iterations needed to approach it and the social welfare attained at each iteration towards the optimal. For each method, we choose different proper values for parameter g. For the algorithm in [3] we pick $g = 30$ such that it convergences within a reasonable number of iterations (order of 10). For our approach we choose $g = 0.25$ for iterations $k = 1,2,...,100$ in order to move fast towards the optimum and a lower value, $g = 0.03$, for iterations $k = 101,102,...,200$ in order to "smooth-out" the oscillations around the optimal state.

We assume a system with $N = 10$ households participating in the DR program. The initial aggregate consumption of each type 2 appliance is assumed to be allocated throughout the whole day, but biased proportionally to the number of the slot, meaning that it causes peak-demands. For instance, $\tilde{Q}_{A2}^0(8) \approx \frac{8}{7} * \tilde{Q}_{A2}^0(7)$. The initial consumption for type 3 appliances is randomly and uniformly distributed in the range [580Wh, 600Wh] for slots $t = 5,6,7,8$, while it has zero value for slots $t = 1,2,3,4$.

Fig. 1 indicates that at the convergence state, the two algorithms exhibit similar behavior with regard to the aggregate (system) per slot consumption. They distinguish the energy consumption into two parts during the day. The type 2 appliances, which can be "ON" over the entire day, gradually shift their load to the first four slots, during which type 3 appliances gain no benefit from consumption. Despite their convergence at the same optimal state, their behavior differs during iterations in the process of achieving it. Our method is subjected to oscillations due to the output of the LP solver used in (12). This is clear in Fig. 1 and 2 and has also impact on the system net benefit achieved towards the optimal (Fig. 3). Furthermore, it is obvious that the algorithm in [3] converges faster. This is not a general property but depends on the value of the parameter g employed for our method, which creates a trade-off between the convergence speed and the iterations needed for the oscillations
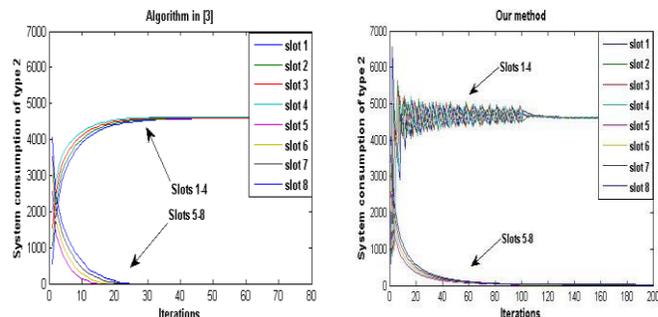
Figure 1. System consumption for type 2 appliances (left figure for the algorithm in [3] and right for the linearization method). In both cases, the load is gradually shifted at slots of no benefit for type 3 appliances.
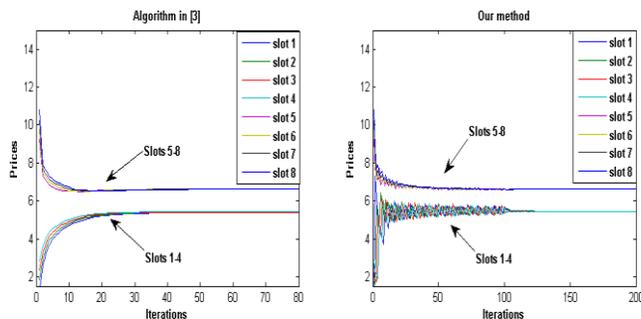
Figure 2. Prices per slot (left figure for the algorithm in [3] and right for the linearization method). Both approaches make the consumption smoother and reduce peak load demands.
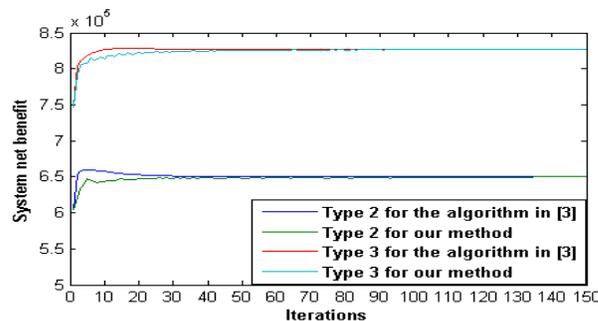
Figure 3. System net benefit for both approaches and types of appliances towards the optimal. The level achieved from our method is relatively close to that of [3] from the very first iterations.

"smooth-out" around the optimal point, as g^k tends to zero.

Fig. 2 shows the prices at each slot towards the convergence. Notice that optimal values are higher for slots 5-8 due to the weight of the utility function of type 3 appliances. Both approaches attain lower peak demands at different time periods, while balancing the need for power generation. The price vector at the convergence state has for both approaches almost equal values as follows:
$$p^*:= (5.43, 5.43, 5.43, 5.43, 6.62, 6.62, 6.62, 6.62)$$

In Fig. 3 we juxtapose the system net benefit achieved from the two methods at each iteration. The algorithm in [3] slightly outperforms ours and their difference decreases as our approach moves towards the optimal.
We additionally provide certain specific results concerning consumer 1, as described in Table I. The consumption of type 2 and type 3 appliances at the convergence state is almost equal for the two approaches, so we only present the results of our method.

TABLE I. INPUT VALUES AND RESULTS (OF (12) AT CONVERGENCE STATE) FOR CONSUMER 1. FOR BOTH APPLIANCES THE MARGINAL UTILITY EQUALS THE MARGINAL COST.

Input A2	$w_{1,A2} = 10102 , \bar{Q}_{1,A2}^{min} = 1548 , \bar{Q}_{1,A2}^{max} = 2416$
Input A3	$w_{1,A3} := (0, 0, 0, 0, 3865, 3818, 3823, 3845)$
Output A2	$\tilde{q}_{1,A2}^* := (469, 464, 462, 467, 0, 0, 0, 0)$
Output A2	$\bar{Q}_{1,A2}^{min} < \bar{Q}_{1,A2}^* = 1862 < \bar{Q}_{1,A2}^{max}$
Output A3	$\tilde{q}_{1,A3}^* := (0, 0, 0, 0, 584, 576, 577, 581)$

Notice that in both cases at convergence state $\gamma_{1,A2}^* \approx p^*(t)$, $t = 1,2,3,4$, while $\gamma_{1,A2}^* < p^*(t)$, $t = 5,6,7,8$, explaining the reason why the type 2 appliance of consumer 1 gradually refrains from consumption during these slots. This result applies for all consumers as it becomes obvious from Fig. 1. Additionally, $\gamma_{1,A3}^*(t) \approx p^*(t), t = 5,6,7,8$ meaning that for both appliances the marginal utility equals the marginal cost whenever they operate.

B. Evaluation of the Load Consolidation Method

In this section, we provide another numerical experiment, in which we apply both the linearization and the load consolidation methods. We evaluate the output compared to the same experiment, when only the linearization method is applied. We refer to the variables as "deterministic", in contrast to the case of applying also load consolidation, where we refer to these as "probabilistic".

We consider a variation of the system of Section V-A, with $N = 4000$ households, each operating one type 2 and one type 3 appliance, the utility functions of which are chosen as previously. The consumption of each type 2 appliance is initiated in a single slot (not the same for each consumer) while we keep the same ratio concerning $\hat{Q}_{A2}^0(t)$ among the slots ($\hat{Q}_{A2}^0(8) \approx \frac{8}{7} * \hat{Q}_{A2}^0(7)$), meaning that peak demanded slots arise also here. Additionally we assume that a portion of consumers (200 out of 4000) may operate the type 2 appliances only during the two latter slots of each day, i.e., $T_{i,A2} = \{7,8\}, i = 3801, 3802, ... ,4000$. The consumption of each type 3 appliance is randomly and uniformly initiated in the range [600Wh, 620Wh] for the slots 5-8. Finally, we pick for both implementations the same pair of values for the parameter g as in Section V-A (0.25 and 0.03).

In Fig. 4 we juxtapose the deterministic and probabilistic prices for three slots. Notice that both implementations reduce the peak load demands and attain the same convergence speed, as it only depends on the value of g parameter (for the same input) and the load consolidation method has no impact on it. The oscillations of the deterministic system are abruptly smoothed out at iteration
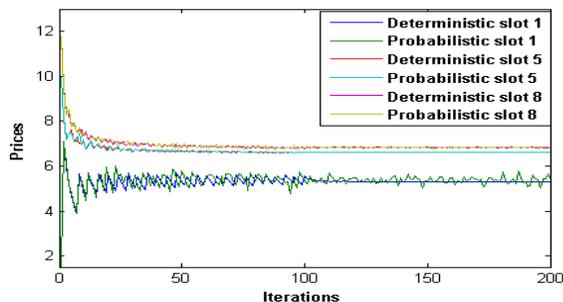
Figure 4. Deterministic and probabilistic prices for slots 1,5,8. The fluctuations at each slot depend on the load portion that type 2 appliances consume at it.
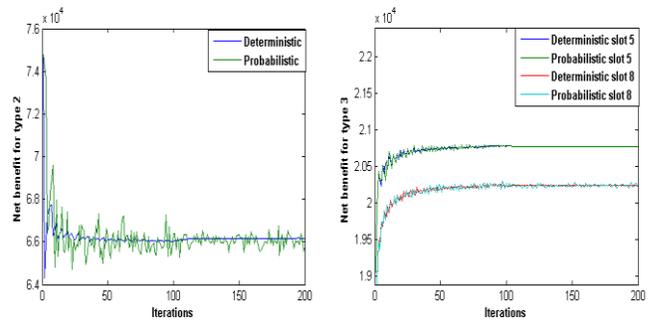
Figure 5.   Net benefit of consumer 1 for both appliances in the deterministic and probabilistic case. The left figure applies to the type 2 appliance and the right to the type 3 one for slots 5 and 8.

TABLE II. NUMERICAL RESULTS FOR CONSUMER 1 UNDER THE DETERMINISTIC AND RANDOMIZED METHODS.

Net Benefit	Deterministic	Mean	%Dif.	St. D.
A2	66165	66054	-0.17	261.3
A3 at slot 5	20780	20780	0	~0
A3 at slot 8	20244	20245	~0	16.18

101, when we decrease the value of g, while the fluctuations of the probabilistic system remain due to the randomized reaction of type 2 appliances.

For the probabilistic system, notice that the fluctuations of $p(5)$ smooth out with iterations. This is expected, as type 2 appliances gradually shift their load to slots 1-4 (as in Section V-A) and tend to refrain from consumption at slot 5. The fluctuations of $p(1)$ are the most intense among the three slots, due to the fact that only type 2 appliances consume at this. In the case of slot 8, the type 3 appliances have a stabilizing role. Their consumption depends on the probabilistic prices (as a result of type 2 reactions), but as they alter their consumption pattern slowly, and consume the greatest portion of load at this slot, they keep the price value relatively close to the deterministic one.

Fig. 5 juxtaposes net benefit of consumer 1 at each iteration for both types of appliances. The above remarks concerning fluctuations at the various slots apply to this case as well. In Table II we provide numerical results for consumer 1 applying to Fig. 5, for iterations 120-200. The "Mean" and "Standard Deviation" (St. D.) columns refer to each probabilistic value. It is apparent that applying the consolidation method yields on the average almost the same user net benefit as the deterministic method for both types of appliances. Also, the level of the standard deviations is aligned with the intensity of fluctuations at each slot.

C. Adding Constraints across Appliances

In this section, we evaluate the linearization method (only) in the presence of constraints across the appliances, so we assume that the aggregate consumption drawn by each household at each slot is upper bounded by the maximum rate Q_i^{max}, i.e., $\sum_a q_{i,a}(t) \leq Q_i^{max} \forall t$, which is the same   for all households: $Q_i^{max} = 350 \forall i$. We utilize the input parameters of Section V and adjust the initial consumption of Section V-A, so as to satisfy the aforementioned constraint. This example is not comparable with the one of Section V-B, as in the presence of this constraint the load consolidation heuristic cannot be applied.

Fig. 6 shows the aggregate consumption of type 2 appliances (left) and the prices (right) for this system. Compared to the case of Section V-A, the additional constraint causes type 2 appliances to consume a portion of their load during slots 5-8. The prices are

equal for each slot $p^*(t) = 4.3 \, \forall t$ because each user's aggregate per slot consumption is the same for all timeslots, and equal to the bound imposed by the additional constraint. These properties are further explained below on the output results for consumer 1.

In Table III we present the output results concerning the consumer 1, for comparison with those of Section V-A. In this case, type 2 appliance consumes also during slots 5-8 so as to satisfy the constraint concerning its minimum consumption value. At the convergence state $\gamma^*_{1,A3}(t) > \gamma^*_{1,A2}$, $t = 5,6,7,8$ meaning that the consumer gains more benefit from the type 3 consumption at these slots. This is the reason why type 2 appliance exactly consumes the requisite load in order to satisfy its minimum power requirements and the remaining is consumed by type 3 at each slot. Furthermore, notice that the load consumed from type 2 at each slot $t = 5,6,7,8$ is not arbitrary but depends on the corresponding weight of the utility function of type 3, so that the values of $\gamma^*_{1,A3}(t)$ are almost equal for all t. Finally, $\gamma^*_{1,A3}(t) > p^*(t)$ and $\gamma^*_{1,A2} > p^*(t) \forall t$, meaning that the consumer would gain extra benefit from further consumption at each slot from any appliance, but this is not possible due to the constraint concerning the total household consumption per slot. This is the reason why the aggregate per slot consumption equals Q_i^{max} at each slot.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have built on the work in [3], where the authors propose a way to use dynamic pricing in the retail power market. Our contribution relies on proposing two heuristics giving a more practical version of the consumer's response to the dynamic prices. Simulations have shown that the proposed linearization method for the utility functions converges to the optimal state with "acceptable" net benefit loses at each intermediate iteration compared to the well defined gradient projection algorithm introduced in [3]. The speed of convergence and the oscillations "smooth-out" around the optimal state, both depending on the value of the relaxation parameter g, which has to be appropriately chosen according to the properties of each approach. Additionally, in the presence of appliances' constraints, our approach converges to the optimal state, while that of [3] does not appear to have been numerically evaluated for this case. Furthermore, the load consolidation method offers practical schedules for the consumers, while it causes negligible variations to the prices and the net benefit of each user compared to the case that it is not applied.

There are further important issues to be considered in order to align our methodology with real life needs and the market context. Regarding our approach, we should extend the load consolidation heuristic in the presence of constraints on the maximum consumption per slot. This issue does not seem to have a trivial solution and deserves further investigation. At the market level, we should clarify the conditions, over which an iterative computation is a practical DR approach, as the context defining the utility functions may differentiate faster than the iterations needed for the convergence. As a solution, we may assume that the procedure towards the optimal state is virtually applied before each day and the consumers actually consume only the optimal schedules discovered. In this case, consumers may respond by means of sophisticated software at the edges of the grid, which automatically adjusts their consumption schedule. Finally, in the case of being forced to deploy our method on top of flat rate pricing, we should identify the mechanism that "translates" the dynamic prices to personalized rebates for each consumer, while resulting at the same consumption outcome. We leave such issues for future work.
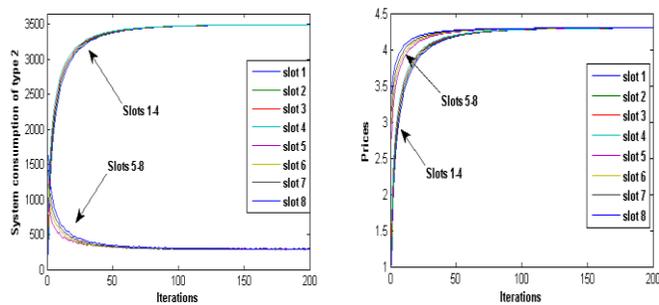
Figure 6. The left figure shows the system consumption for type 2 appliances, a portion of which is now during slots 5-8. The right figure shows the prices, the optimal values of which are determined by the additional constraint.

TABLE III. INPUT VALUES AND RESULTS (OF (12) AT CONVERGENCE STATE) FOR CONSUMER 1. THE CONSTRAINT ACROSS APPLIANCES DETERMINES THE AGGREGATE CONSUMPTION.

Input A2	$w_{1,A2} = 10102$, $\bar{Q}^{min}_{1,A2} = 1548$, $\bar{Q}^{max}_{1,A2} = 2416$
Input A3	$w_{1,A3} := (0,0,0,0,3865,3818,3823,3845)$
Output A2	$\tilde{q}^*_{1,A2} := (350,350,350,350,35,39,38,36)$
Output A2	$\bar{Q}^*_{1,A2} = 1548 = \bar{Q}^{min}_{1,A2}$
Output A3	$\tilde{q}^*_{1,A3} := (0,0,0,0,315,311,312,314)$

REFERENCES

[1] "The Smart Grid: An Introduction," U.S. Department of Energy, http://www.oe.energy.gov/DocumentsandMedia/DOE_SG_Book_Single_Pages(1).pdf, Sept. 2010 [retrieved Jan., 2013].

[2] "Energy Efficiency Trends and Policies in the EU 27". http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/DOE_SG_Book_Single_Pages%281%29.pdf, Paris, Oct. 2009 [retrieved Jan., 2013].

[3] N. Li, L. Chen, and S. H. Low. "Optimal demand response based on utility maximization in power networks." In *Power and Energy Society General Meeting, 2011 IEEE*, pp. 1-8. IEEE, 2011.

[4] A. Mohsenian-Rad, and A. Leon-Garcia. "Optimal residential load control with price prediction in real-time electricity pricing environments." *Smart Grid, IEEE Transactions on* 1, no. 2 (2010): pp. 120-133.

[5] M. Pedrasa, T. Spooner, and I. MacGill. "Coordinated scheduling of residential distributed energy resources to optimize smart home energy services." *Smart Grid, IEEE Transactions on* 1, no. 2 (2010): pp. 134-143.

[6] C.L. Su, and D. Kirschen. "Quantifying the effect of demand response on electricity markets." *Power Systems, IEEE Transactions on* 24, no. 3 (2009): pp. 1199-1207.

[7] M. Fahrioglu, and F. L. Alvarado. "Using utility information to calibrate customer demand management behavior models." *Power Systems, IEEE Transactions on* 16, no. 2 (2001): pp. 317-322.

[8] T. Bapat, N. Sengupta, S. K. Ghai, V. Arya, Y. B. Schrinivasan, and D. Seetharam, "User-sensitive Scheduling of Home Appliances" in Proceedings of the 2nd ACM SIGCOMM workshop on Green networking (GreenNets '11, Aug. 2011) pp. 43-48, Toronto, Ontario, Canada.

The Many Faces of Real-Time Scheduling Applied to Power Load Management

Tullio Facchinetti

Dipartimento di Ingegneria Industriale e dell'Informazione

University of Pavia, Italy

e-mail: tullio.facchinetti@unipv.it

Abstract—In the field of power systems there is an active research on methods to limit the peak load of power demand. Recently, a new approach to the automatic management of electric loads has been proposed, which is based on scheduling techniques derived from real-time computing systems. The research on real-time scheduling encompasses several aspects, including scheduling algorithms and analysis techniques, that may suitably be adapted to the management of sets of power loads. This paper overviews background concepts on real-time scheduling, and fosters their possible application to the power load management, with special focus on peak load reduction. Several issues regarding the power load management and possible solutions based on real-time scheduling are considered. Due to the wide scope of the proposed methodologies, only the indication of their applicability is provided in this paper. For more complex scenarios, references to previous related works are provided when available.

Keywords-Modeling; Real-time systems; Power system control; Scheduling; Smart Grid; Demand-Side Management; Direct Load Control; Load Shedding.

Figure 1. Overview of the smart distributed coordination infrastructure enabling the advanced control features based on real-time scheduling envisioned in the paper.

I. INTRODUCTION

The Smart Grid is the emerging technology in the field of electric power management, generation, distribution and usage. Its architecture combines the power distribution infrastructure with a digital communication network. Load balancing is a key challenge addressed by the Smart Grid [1]. Peak load conditions are generated by the simultaneous request of electricity by many users. Such situations may bring to severe consequences, arising technical and economic issues for both suppliers and users [2]. Therefore, an appropriate load management targeted to obtain predictable load conditions would lead to remarkable benefits.

This paper describes the potentials of a new methodology based on real-time scheduling to manage the predictable activation/shedding of power loads. The background idea is to establish an analogy between real-time computing systems and power systems, in order to use real-time modeling, scheduling and analysis methodologies to coordinate the activation of a set of loads. The goal is to delineate the characteristics of a general framework where different types of loads, physical constraints, system dynamics and objectives can be merged to allow system-wide real-time optimization of the peak load. For this purpose, electric loads are described using timing parameters derived from the real-time scheduling research domain. In this way, existing well-known real-time scheduling algorithms can be applied to predictably activate/shed each load, while guaranteeing given constraints on the underlying physical process associated with the electric load.

There are several relevant motivations to investigate the application of real-time scheduling techniques to power systems. Some of the most important benefits include:

- the predictable behavior of real-time systems, that can be a-priori guaranteed in a mathematically strong form;
- the availability of several modeling and control methodologies that can be adapted and applied to the load management;
- the automatic derivation of load priorities based upon the characteristics of both the load and the underlying physical process;
- the possibility to manage large sets of loads, thanks to powerful scheduling policies with low computational complexity;
- the availability of efficient heuristics and optimization-based methods to deal with more complex cases.

The aim of this paper is to shed the light on the chances offered by the application of real-time scheduling techniques to the coordination of power loads. In particular, the paper illustrates and categorizes the available approaches in the field of real-time systems that can be suitably applied to the management of electric loads. Therefore, each technique is

briefly described and its possible application to the management of power loads is suggested.

Fig. 1 shows the infrastructure suitable for the implementation of the proposed management system. Scheduling decisions are taken at the centralized control station and dispatched to the smart metering and control devices installed within the smart building. Such decisions trigger the activation/shedding of smart devices within the building. During the configuration, the smart meter collects configuration parameters from the controlled devices, and send them to the control station. At run-time, it measures the power usage and collect other interesting values and provides the information to the control station. In this scenario, loads belonging to different buildings can be coordinated in an aggregated manner to boost the balancing effect.

The paper is organized as follows. Section II provides a short introduction to basic concepts related with real-time scheduling. A generalization of the applicability of real-time scheduling techniques is presented in Section III. Section IV discusses scheduling-based works related with smart power management. Characteristics and features of real-time scheduling techniques suitable to electric loads management are described in Section V. Finally, Section VI concludes the paper.

II. NOTIONS OF REAL-TIME SCHEDULING

Real-Time Systems (RTSs) are studied in the domain of computing systems to allow the timely execution of concurrent processing tasks on one or more processors [3]. The peculiarity of a RTS is its strong relationship with the physical system under monitoring and control. Moreover, a distinguishing feature is the analysis of system properties typically based on the evaluation of worst case conditions, in order to achieve the desired behavior in every possible situation. This fact makes RTSs especially suitable for critical applications, such as in automotive systems, avionics, factory automation, and process control (e.g., chemical or nuclear plants).

In the traditional system model a set of computing tasks are required to be executed on one processor. Since only one task can use the processor at any given time, a scheduling algorithm assigns the use of the processor to the task having the highest priority. Differently from other approaches, the priority of tasks is not explicitly set by the system designer using some empirical assessment. Instead, it is automatically inferred by the scheduling algorithm based on values of timing parameters used to describe the task.

The simplest and most investigated system model in RTSs consists of periodically activated tasks. Each task τ_i becomes ready for the execution at the request time r_i, which is an integer multiple of the *period* T_i, i.e., the k-th instance of τ_i (called *job*) is released at time kT_i. The job duration is at most C_i time units. The parameter C_i is referred to as Worst-Case Computation Time (WCET). A relative
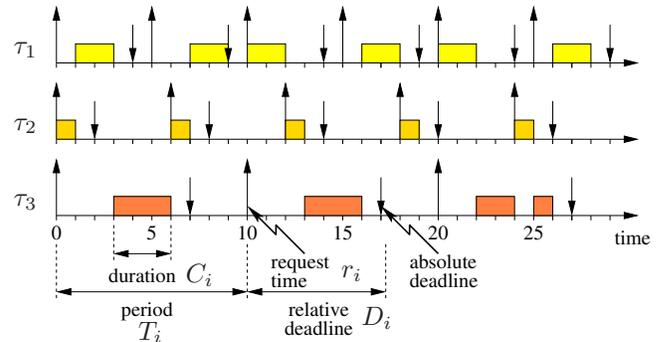
Figure 2. Example of concurrent scheduling of 3 tasks by EDF. Each task is described by the tuple (T_i, D_i, C_i). Task parameters are $\tau_1 = (5, 4, 2)$, $\tau_2 = (6, 2, 1)$ and $\tau_3 = (10, 7, 3)$.

deadline D_i is associated with the task τ_i. A deadline D_i means that the k-th job must complete no later than D_i time units after the release time, i.e., it must execute in the time frame $[kT_i, kT_i + D_i)$, where $d_i = kT_i + D_i$ is the absolute deadline of the k-th job. Often, relative deadlines are assumed equals to the period, i.e., a job must terminate before the next period. In this case, the task model is said having *implicit deadlines*. The scheduling algorithm controls the proper execution of each task within its period time frame. For instance, the Earliest Deadline First scheduling algorithm (EDF) assigns the highest priority to the task having the earliest absolute deadline. This model is suitable in several concrete applications. For instance, control applications require a periodic sensor sampling and actuator driving.

The example in Fig. 2 shows the schedule generated for 3 tasks by the EDF algorithm. The task τ_3, e.g., have a period of $T_3 = 10$ time units and a computation time $C_3 = 3$. Each job must complete no later than $D_3 = 7$ time units after the release time. Notice that, as allowed in typical RTSs, the third job of τ_3 is interrupted at time $t = 24$ after 2 time units and resumed later for being completed. This interruption is called *preemption*, and it plays an important role to let all tasks to meet their deadline. Preemptions can represent an important issue when real-time techniques are applied to the management of electric loads, as will be discussed later.

The periodic task model has several attractive properties. A very interesting one is the possibility to define a figure called *utilization*, defined as $U = \sum_i C_i/T_i$. The utilization represents the load of a processor, being the percentage of computing time required by all tasks. A remarkable use of the utilization is for the so-called *schedulability test*. A schedulability test is a condition that, whether satisfied, guarantees that every task will meet its timing constraints, i.e., the execution of every job terminates before the deadline in every possible situation (i.e., considering worst case conditions). This test can be used to obtain *a-priori guarantees* on the scheduling process. For example, when the
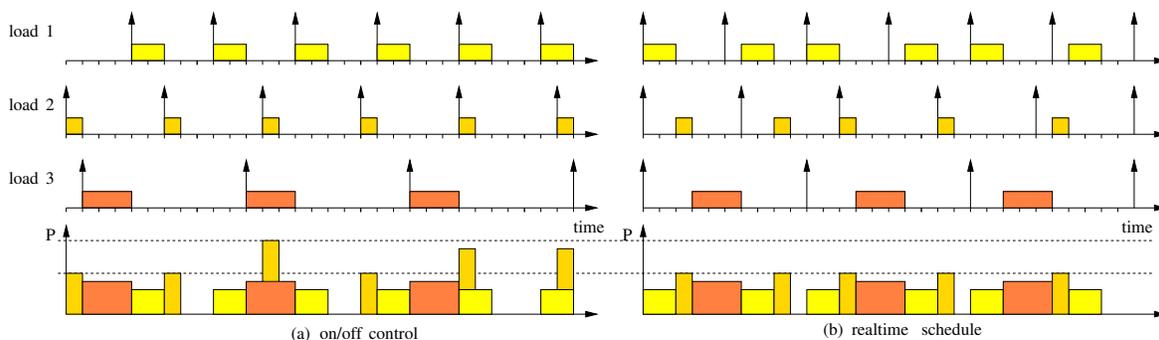
Figure 3.   Comparison between (a) a traditional on/off control method and (b) a coordinated load activation. The picture shows the comparison between the achieved peak loads. The time-base is 5 minutes.

preemptive EDF scheduling algorithm is used, and implicit deadlines are considered, timing constraints are guaranteed iff $U \leq 1$. The schedulability test formulation depends essentially on the adopted system model, task model and scheduling algorithm.

III. REAL-TIME SCHEDULING FROM COMPUTING TO POWER SYSTEMS

The problem of using real-time scheduling methods for managing sets of power loads is related with the modeling of electric loads behaviors in terms of timing parameters such as periods and deadlines – as introduced in Section II. In practice, it is necessary to determine the value of timing parameters to obtain the desired behavior of the modeled load. The derivation of timing parameters enables the application of real-time scheduling algorithms for the automatic management of loads activation/shedding. Moreover, real-time analysis methods can be used to a-priori determine the expected system behavior and to assess its performance.

The periodic model described in Section II can be easily adapted to model power loads. For instance, a HVAC (Heating, Ventilation and Air Conditioning) may be periodically activated for 5 minutes (WCET) every 30 minutes (period) to maintain the desired temperature within a room. The measurements carried out in [4] confirm the periodic nature of such type of loads. Load 2 depicted in Fig. 3 is an example of a load having such timing parameters.

The schedule represented in Fig. 2 can thus be reconsidered from the load management viewpoint. The 3 depicted tasks may represent the corresponding number of air-conditioners whose activation is periodically scheduled. Each load is characterized by a specific power demand when active. Fig. 3.b shows this application. In the figure, the schedule is completed by the time-line representing the instantaneous power consumption. The scheduling action automatically eliminates the simultaneous activation occurring in case of absence of coordination. This allows to obtain a peak load that is equal to the power consumed by the most power-consuming load. The activation pattern generated by

the real-time schedule is compared with the one produced by a classical on/off hysteresis controller (Fig. 3.a). The controller is used to regulate the variation of a physical quantity (e.g., the temperature) in order to limit its variation within a predefined range. It works by switching on the system when a range threshold is exceeded and keeping the system active until the lower threshold is reached. This control scheme leads to a behavior that can be approximated with a strictly periodic activation of the load, as in Fig. 3.a. The result is the absence of coordination among loads, leading to higher peak loads.

Clearly, this is a simplified example. For instance, the values of timing parameters need to be related to the physical variables of interest (internal/external temperatures, room insulation, open doors/windows, etc.). However, the goal of this paper is precisely to illustrate how such factors can be integrated in the model and accounted during the system functioning.

On the other hand, this approach is able to seamlessly manage large sets of heterogeneous loads. In fact, once a load is modeled in terms of timing parameters, its activation can be scheduled in coordination with other loads by an algorithm such as EDF. The type of loads that may be fruitfully managed by this approach include refrigerators, HVACs, air compressors, pumps, battery charge/discharge (e.g., in electric vehicles), lighting, and household appliances.

IV. RELATED WORKS

The techniques described in this paper can be seen as viable heuristics to face the problem of peak load reduction of power demand, as well as methods to reduce the overall electric energy demand. While there is an extended literature on power systems addressing the problem of peak load limitation and load balancing in general, in this paper we will focus on scientific results related to scheduling approaches applied to power management.

In [5] scheduling techniques are applied to the control of on/off smart loads in the Smart Grid. Several control

approaches are presented, including a priority-based and a round-robin scheduling schemes. An efficient heuristic to handle large sets of heterogeneous loads is presented in [6]. The problem is formulated as an optimization problem, and a flexible and efficient heuristic is developed to solve the problem. In [7] there is an accurate physical modeling of several kind of loads to be incorporated into a home energy management (HEM) system. The modeling is used to propose a priority-based control scheme to achieve a bound on the peak load of power demand.

Recently, the notion of *Real-Time Physical System* (RTPS) has been introduced to indicate a general class of systems where the variation of a physical value is determined by the schedule generated by a real-time scheduling policy [8]. This approach is behind the ideas described in this paper. The mentioned work addresses systems with linear dynamics, thus having an exponential behavior in the time domain, under timing and physical constraints. In [9] an optimization method is proposed to reduce the peak load when the activation of multiple loads at the same time is allowed, which is a common situation in large systems. However, the system model does not consider constraints on the state variable. In [10], constant dynamics were considered without feedback on the state variable. The concept of RTPS in absence of constraints on the physical variables (user requirements) is introduced in [11]. The paper provides a statistical evaluation of the benefit of using RTPS against the case of absence of explicit load control. Finally, in [12] RTPS are extended with a feedback scheme for the management of electric load in presence of uncertainties on the values of modeling parameters. The contribution of these papers is to provide the relationship between the variation of physical variables (temperature, air pressure, battery charge, etc.) and the timing parameters adopted to model the related load.

V. REAL-TIME CONCEPTS AND LOAD MANAGEMENT

This section introduces some of the most highly investigated issues in the field of Real-Time Systems. Their applicability and usefulness to the management of power loads is described, outlining their benefits.

A. A-priori guarantees

A peculiar feature of results developed in the field of RTSs is their strong theoretical background. This means that the vast majority of results, including analysis methodologies, characteristics of scheduling algorithms and their performance, are mathematically proved. The mathematical strength of the real-time scheduling analysis includes the derivation of schedulability tests for many task and system models [3]. Tests are based on the timing parameters of tasks. Under adequate assumptions, if the test is passed then the system will behave as expected in every possible condition. In this way, schedulability tests provide a mathematically proved *a-priori guarantee* on the system behavior.

Such guarantees can be extended to the peak load generated by a load set, thus resulting in physical processes that perform as expected while achieving an upper bound on the peak load of required power. This approach is used in [9] to determine a worst case bound of the peak load.

B. Scheduling algorithms

The scheduling algorithm encapsulates the policy that determines the sequence of activation of tasks/loads. In practice, it sets the actual priority of a task based on its timing parameters. At any given time, the task with the highest priority is selected for execution, while the remaining tasks are delayed. The study of scheduling algorithms is a central argument in the research on real-time systems, and several algorithms covering a wide range of scheduling problems variants are available.

This concept has a relevant relationship with power/energy systems. Load shedding techniques are often based on explicit specification of load priorities required to select the sheddable ones. The main advantage of real-time techniques is that the priority selection is not left to the system designer. Instead, it is deterministically based on timing and physical requirements that are translated into the timing constraints associated with tasks.

C. Aperiodic activities

Aperiodic activities are those tasks that need to be scheduled "on demand", i.e., without a fixed periodic activation. In the domain of electric loads management, examples of aperiodic tasks include TVs, elevators and ovens. In general, such loads must be activated after a user request without delay. In [7], this type of loads is said *critical*.

There are several approaches to integrate aperiodic tasks in a schedule. When such tasks need to coexist with periodic activities, the common solution is to use a *server* that periodically serves the execution of aperiodic requests. In this way, aperiodic requests can nicely be incorporated in a schedule made by periodic tasks. Other solutions exist. For example, in the gravitational task model a set of aperiodic activities need to be executed at precise time instant [13]. The gravitational model allows to minimize the offset of a task execution with respect to the desired time instant.

D. Online and offline scheduling

Most scheduling approaches in the field of RTSs are developed as online algorithms. An online algorithm dynamically generates the scheduling decisions at run-time, i.e., while the system is actually working. This approach has the strong benefit to allow a timely and flexible management of faulty conditions, such as overloads. The possibility to use online scheduling algorithms comes at the price of imposing simplifying assumptions on the system model.

Therefore, complex constraints can not be easily managed by online algorithms. For this purpose, offline scheduling is conceived as an optimization process that is performed before the actual system working, to a-priori allocate the required resources [3]. This method extends the potential of real-time scheduling, at the price of less run-time flexibility and higher computational complexity.

Similar considerations apply to loads management. Simple cases can be efficiently managed by online algorithms, while more complex situations, in terms of constraints, can benefit of off-line solutions. For example, the complex interplay between energy demand and energy production from renewable sources may require a costly computation to satisfy the imposed economic constraints.

E. Hard/soft task models: worst-case vs average cases

Almost all RTSs models are based on the assignment of *deadlines* to the task to be scheduled, as stated in Section II. When a process terminates its activity after the assigned deadline the correct system behavior may be jeopardized. Regarding the achievement of deadlines, real-time systems are either classified as *hard* or *soft*. While the former do not tolerate the violation of any deadline, in latter systems missing deadlines are allowed. Examples of hard systems are safety critical applications (e.g., avionics, automotive, etc.), while soft systems are found in the multimedia (non-critical) domain [3].

The concept of hard/soft systems can be easily extended to load management. Beside timing constraints, there are often physical requirements associated with the activation of a load. While in some cases the missing of a physical constraint is not allowed, in many cases the imposed physical constraints can be sometimes violated to account for special situations. For example, consider the room temperature regulated by an HVAC that is required to remain within the range $18 - 20°C$. A missed deadline may cause the temperature to temporarily reach $21°C$ without considerably affect the user comfort. In other cases the worst case conditions need to be achieved, since they are the potential source of power provisioning disruption (black outs) [2].

F. Scalability to large systems

The set of loads to be controlled in a realistic scenario can be composed by a large number of devices. Scalability issues may arise in large networks. In RTSs many useful techniques, including scheduling algorithms and analysis methodologies, have linear or polynomial computational complexity. Thus their application can suitably fit to large load sets. This section illustrates the approaches that can be leveraged to face the management of large load sets.

1) Multiprocessor scheduling: In RTSs, multiprocessors are those systems that allow to execute more than one task at any given time, since each task runs on a different processor [14]. Multiprocessors can clearly manage a larger number of tasks with respect to uniprocessor systems. However,

in computing systems the number of processors is imposed by the computing platform (i.e., it is a system constraint). Therefore, the typical goal is to provide an answer to the question: "Is the considered task set successfully schedulable by the available processors?", where a successful schedule is the one that satisfies the timing constraints of every task.

The application of real-time scheduling to the load management is inherently close to the multiprocessor scenario. The number of loads can be large, thus the simultaneous activation of two or more loads can hardly be avoided. However, the number of simultaneously activated loads is not a system constraint in this case. Instead, the reduction of unnecessary simultaneous activations is the objective of the approach based on real-time scheduling. Therefore, the above question can be translated to "What is the maximum number of simultaneous activated loads?". The final goal is to determine the resulting peak load in the worst case. The resulting value can be used to either determine the size of a new infrastructure or to verify the suitability of the schedule to the power constraints imposed by an existing system (e.g., to check whether contractual limits are satisfied).

2) Hierarchical scheduling: A typical method to manage large systems is to identify a hierarchy in the system components. In RTSs the hierarchical scheduling is used to aggregate heterogeneous scheduling policies in a predictable manner [15]. A schedule generated at a higher level can contain a sub-schedule, whose formal properties influence and can be used to analyze the overall system behavior.

The modeling of power infrastructure can be straightforwardly organized in a hierarchical manner. The hierarchy starts from electric loads/devices at a lower level and grows into aggregated levels as apartments/houses, buildings, districts and cities. The idea is that a predictable schedule guaranteed at lower levels, and thus a predictable power load, allows to achieve predictable aggregated power loads at higher levels. This approach has huge potentials in the management of large and heterogeneous power systems.

G. Approximated scheduling

The estimation of the value of relevant parameters to characterize a physical process is often affected by some degree of uncertainty. The uncertainty can derive from incomplete knowledge of the process, from simplified assumption to obtain simpler models, etc.

In RTSs, an approach to deal with uncertainties is the so-called probabilistic scheduling. It allows to incorporate the uncertainty regarding physical and timing parameters in the system model. For this purpose, the common timing parameters such as periods and execution times are expressed by stochastic variables. The stochastic nature of timing parameters is used to derive probabilistic guarantees on the system behavior [16].

The possibility to incorporate a sufficient degree of uncertainty in the model is essential in the practical applicability

TABLE I
A RESUME OF THE PROPOSED REAL-TIME METHODOLOGIES AND THEIR
APPLICATION TO POWER LOAD MANAGEMENT.

Feature	Application to power load management
A-priori guarantees	To determine useful system properties (e.g., the peak load) from timing parameters without the need of simulations
Scheduling algorithms	To coordinate the intelligent activation of power consuming devices, avoiding unnecessary simultaneous activations
Aperiodic activities	To efficiently manage the concurrent activation of periodic loads and aperiodic ones (e.g., ovens, dishwashers, washing machines)
Online/offline scheduling	To deal with mixed complex application constraints (offline) and dynamic adaptation of system requirements at run-time (online)
Hard/soft task models	To integrate both critical (hard) loads and less urgent (soft) ones
Multiprocessor scheduling	To manage the activation of two or more loads at the same time, when this condition can not be avoided (e.g., in large load sets)
Hierarchical scheduling	To coordinate a large power system that is decomposed into a set of sub-systems organized in a hierarchical manner
Approximated scheduling	To cope with modeling uncertainties introduced in the modeling of power loads and systems

of the scheduling approach to energy systems, where the exact value of physical parameters are always affected by errors and/or approximations. Robust scheduling methods are then required to cope with such issues and achieve reliable results under uncertain conditions.

VI. CONCLUSIONS

This paper has suggested the applicability of some existing real-time scheduling approaches with explicit indication of their possible applications to power loads management. In particular, basic concepts of real-time scheduling have been introduced to allow the subsequent coverage of more advanced topics in connection with power management.

REFERENCES

[1] A. Fazeli, E. Christopher, C. Johnson, M. Gillott, and M. Sumner, "Investigating the effects of dynamic demand side management within intelligent smart energy communities of future decentralized power system," in Innovative Smart Grid Technologies (ISGT Europe), 2nd IEEE PES Int. Conf. and Exhibition on, dec. 2011, pp. 1–8.

[2] D. Novosel, M. M. Begovic, and V. Madani, "Shedding light on blackouts," Power and Energy Magazine, IEEE, vol. 2, no. 1, pp. 32–43, jan-feb 2004.

[3] G. Buttazzo, Hard Real-Time Computing Systems: Predictable Scheduling Algorithms and Applications. Springer, 2011.

[4] G. Koutitas and L. Tassiulas, "A delay based optimization scheme for peak load reduction in the smart grid," in Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet. ACM, 2012, p. 7.

[5] G. Koutitas, "Control of flexible smart devices in the smart grid," Smart Grid, IEEE Transactions on, vol. PP, no. 99, pp. 1–11, 2012.

[6] T. Logenthiran, D. Srinivasan, and T. Z. Shun, "Demand side management in smart grid using heuristic optimization," Smart Grid, IEEE Transactions on, vol. PP, no. 99, pp. 1–9, 2012.

[7] M. Pipattanasomporn, M. Kuzlu, and S. Rahman, "An algorithm for intelligent home energy management and demand response analysis," Smart Grid, IEEE Transactions on, vol. PP, no. 99, p. 1, 2012.

[8] M. Della Vedova, M. Ruggeri, and T. Facchinetti, "On real-time physical systems," in Proceedings of the 18th International Conference on Real-Time and Network Systems (RTNS), November 2010, pp. 41–49.

[9] T. Facchinetti, E. Bini, and M. Bertogna, "Reducing the peak power through real-time scheduling techniques in cyber-physical energy systems," in Proceedings of the First International Workshop on Energy Aware Design and Analysis of Cyber Physical Systems (WEA-CPS), April 2010, pp. 18–25.

[10] T. Facchinetti and M. L. Della Vedova, "Real-time modeling and control of a cyber-physical energy system," in Proceedings of the First International Workshop on Energy Aware Design and Analysis of Cyber Physical Systems (WEA-CPS), April 2010, pp. 10–17.

[11] M. L. Della Vedova, E. Di Palma, and T. Facchinetti, "Electric loads as real-time tasks: an application of real-time physical systems," in 7th International Wireless Communications and Mobile Computing Conference (IWCMC), July 2011, pp. 1117–1123.

[12] T. Facchinetti and M. L. Della Vedova, "Real-time modeling for direct load control in cyber-physical power systems," IEEE Transactions on Industrial Informatics, special issue on Information Technology in Automation, vol. 7, no. 4, pp. 689–698, November 2011.

[13] R. Guerra and G. Fohler, "A gravitational task model with arbitrary anchor points for target sensitive real-time applications," Real-Time Systems, vol. 43, no. 1, pp. 93–115, 2009.

[14] R. I. Davis and A. Burns, "A survey of hard real-time scheduling for multiprocessor systems," ACM Computing Surveys, vol. 43, no. 4, pp. 1–44, Oct. 2011.

[15] J. Regehr, A. Reid, K. Webb, M. Parker, and J. Lepreau, "Evolving real-time systems using hierarchical scheduling and concurrency analysis," in Proceedings of the IEEE Real-Time Systems Symposium, dec. 2003, pp. 25–36.

[16] A. Burns, S. Punnekkat, L. Strigini, and D. Wright, "Probabilistic scheduling guarantees for fault-tolerant real-time systems," in Dependable Computing for Critical Applications 7, 1999, nov 1999, pp. 361–378.

Efficient Multicast Authentication in Energy Automation Environments

Rainer Falk, Steffen Fries

Corporate Technology
Siemens AG
Munich, Germany
e-mail: {rainer.falk|steffen.fries}@siemens.com

Abstract—**Information security is gaining increasingly more importance for real-time industrial automation networks. Multicast communication is used widely especially on field and process level to cope with performance requirements and to ease the handling of communication peers as the destinations need not to be known by the sender. A security design must not interfere with these communication types. For these reasons, a solution is required allowing to perform an efficient authentication of field level multicast communication.**

Keywords–security; device authentication; multicast; real-time; network access authentication; firewall

I. INTRODUCTION

Decentralized energy generation (e.g., through solar cells or wind power) is becoming increasingly important to fight global warming and to better exploit existing energy resources. Introducing decentralized energy generators into the current energy distribution network poses great challenges for energy automation (EA) in the smart grid scenario, e.g., secure communication between a control station and equipment of users (e.g., decentralized energy generators) but also secure communication on decentralized field equipment must be addressed. Standard communication technologies as Ethernet and IP are increasingly used in energy automation environments down to the field level. Guaranteed real-time communication plays an essential role for many industrial control applications.

IEC 61850 is one popular standard for communication in the domain of energy automation. It is assumed to be the successor of the currently used standards IEC 60870-4-104 and DNP3 also for the North American region. IEC 61850 enables interoperability between devices used in energy automation, i.e., two IEC 61850 enabled devices of different manufacturers can exchanged a set of clearly defined data and the devices can interpret and use these data to achieve the functionality required by the application due to a standardized data model. In particular IEC 61850 enables continuous communication from a control station to decentralized energy generators by using a standardized data format.

Today, IEC 61850 is mainly used for reporting status and sampled value information from Intelligent Electronic Devices (IED) to Substation automation controller as well as for command transport from Substation automation controller to IEDs. It also addresses the communication directly between IEDs using the Generic Object Oriented Substation Event (GOOSE) instead of dedicated wires. Necessary tasks comprise also configuration of equipment as well as control of circuit breakers. The following figure shows a typical example scenario in which IEC 61850 can provide a benefit.
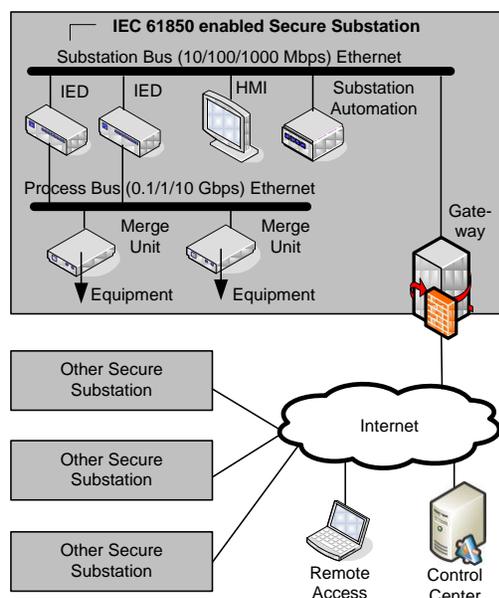
Figure 1. Typical IEC 61850 Scenario

Security is increasingly important in energy automation as on part of the Smart Grid. Here, IEC 62351 kicks in, providing security services for IEC 61850 based communication covering different deployment scenarios using serial communication, IP-based communication and also Ethernet communication. The latter one is used for in substation automation to cope with the real-time requirements. While these messages may not need to be encrypted, they need to be protected against manipulation on one hand and allow for source authentication on the other. Note that besides pure communication security there is also the need to address security in the physical environment and

also in the processes connected with communication. This is being addressed for instances in IEC 62443 (ISA 99) or in ISO TR 27019 for the automation environments. Both standards are stated here to underline, that security is not only restricted to the field communication and also applies to the embedding environment The paper itself does not address these standards and concentrates on the specific problem of multicast authentication on field level.

The remainder of this paper is structured as follows: Section II provides an overview on real-time control networks on the example of GOOSE in substation automation. Section III describes the problem statement and the existing security solution. Section IV gives an overview about multicast authentication schemes. This is used later on in section V and section VI by applying them to substation automation. Section VII concludes the paper and provides an outlook.

II. SUBSTATION AUTOMATION COMMUNICATION

Real-time systems typically consist of hardware and software that are subject to time constraints regarding execution of commands. This comprises the initiation of a command, the execution itself and the acknowledgement of the execution. Real-time in the context of this paper refers to systems with a deterministic behavior, resulting in a predictable maximum response time. These systems will handle all events at appropriate (context-dependent) speed, without loss of events. Automation networks are typically shared networks connected in a ring, star, or bus topology or a mixture of these. Most often, the time critical part is performed on a dedicated network, while the rest of the communication supporting the automation systems is performed on networks with lower performance requirements. An example may be the connection of the process network to a SCADA (office) network. For example, a ring topology is shown in Figure 2.
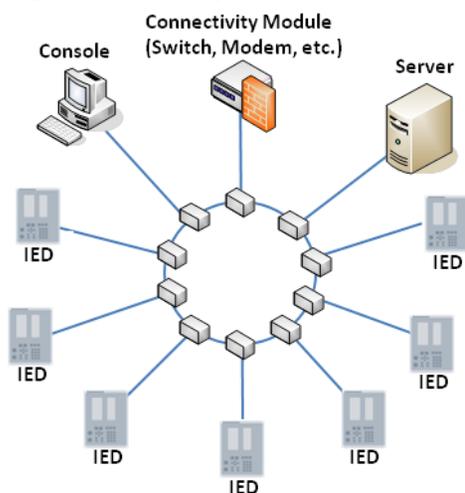
Figure 2. Ring topology in a substation

One of the protocol sets used in substation automation is IEC 61850, which provides Generic Object Oriented Substation Events (GOOSE) on process bus level. It is a control model mechanism in which any format of data (status, value) is grouped into a data set and transmitted as set of substation events, such as commands, alarms, or indications. It aims to replace the conventional hardwired logic necessary for intra-IED (Intelligent Electronic Device) coordination with station bus communications. Upon detecting an event, field devices use a multi-cast transmission to notify those devices that have registered (subscribed) to receive the data. GOOSE messages or Sampled Values (SV) are re-transmitted multiple times by each field device. The reaction of each receiver depends on its configuration and functionality.
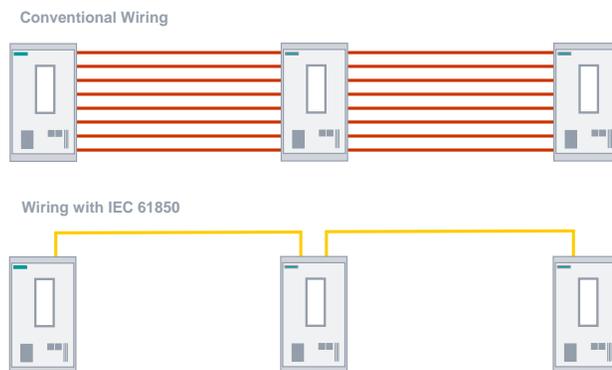
Figure 3. Advantage of using IEC 61850 GOOSE

Following mechanisms are used to ensure the specified transmission speed and reliability:

– GOOSE data is directly embedded into Ethernet data packets and works on publisher-subscriber mechanism on multicast or broadcast MAC addresses

– GOOSE uses VLAN and priority tagging as per IEEE 802.1Q to have a separate virtual network within the same physical network and to set an appropriate message priority level

– Enhanced retransmission mechanisms – the same GOOSE message is retransmitted with varying and increasing re-transmission intervals. A new event occurring within any GOOSE dataset element will result in the existing GOOSE retransmission message being stopped. A state number within the GOOSE protocol identifies whether a GOOSE messages is a new message or a retransmitted message.

IEC 61850-5 [3] defines message types and their performance classes. The following performance classes are supported:

– P1 applies typically to a distribution bay (or where low requirements can be accepted),

– P2 applies typically to a transmission bay (or if not otherwise specified by the customer),

– P3 applies typically to a top performance transmission bay.

The following table shows the different message types and their timing requirements based on IEC 61850-5 [3].

TABLE I.   GOOSE TRANFER TIMES

Type	Definition	Timing Requirements
1	**Fast messages** contain a simple binary code containing data, command or simple message, examples are: "Trip", "Close", etc.	See Type 1a and 1 b below
1A	**TRIP** – most important message	– P1: transfer time shall be in the order of half a cycle. → 10 ms – P2/3: transfer time shall be below the order of a quarter of a cycle. → 3 ms
1B	**OTHER** – Important for the interaction of the automation system with the process but have less demanding requirements than trip.	– P1: transfer time < 100ms – P2/3: transfer time shall be below the order of one cycle. → 20 ms
2	**Medium speed messages** are messages where the time at which the message originated is important but where the transmission time is less critical.	– Transfer time < 100ms
3	**Low speed messages** are used for slow speed auto-control functions, transmission of event records, reading or changing set-point values and general presentation of system data.	– Transfer time < 500ms

The definition of transfer time, according to IEC 61850-5, is shown in Figure 4 below. The transfer time includes the complete transmission of a message including necessary handling at both ends. The time counts from the moment the sender feeds the data content into transmission stack till the moment the receiver extracts the data from its transmission stack. As shown in TABLE I. the transfer time of GOOSE messaging for a TRIP command shall be such that the command should arrive at the destination IED within 3ms. For a single IED, by assuming the time for the publishing process and the subscribing process are approximately equal and if t_b can practically be ignored, then at least half of the defined time is needed for the IEDs to process the message (i.e., 1.5ms for TRIP).
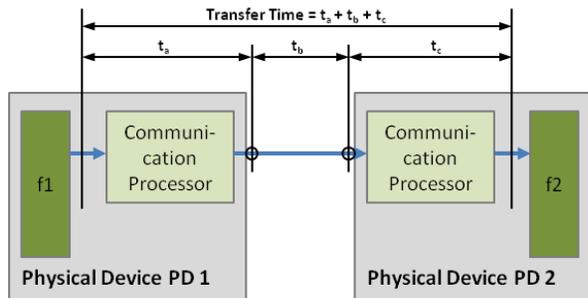
Figure 4. Transfer Time [3]

As shown in Figure 5 below, if a signal, e.g., the pick-up "Overcurrent I>picked up", is configured in a GOOSE message, the IED sends this message cyclically every 0.5 seconds as a telegram with high priority over the Ethernet network. The content of this telegram communicates the state of pick-up ("not picked up" or "picked up") to the subscribers of the GOOSE message. The cyclic transmission enables each of the subscribers to detect a failure using a logic block when a transmitter has failed or a communications channel has been interrupted.

This approach provides constant monitoring of the transmission line because the subscriber expects to receive a telegram at several-second intervals. This can be compared with pilot-wire monitoring in conventional wiring. On a pick-up, i.e., a signal change, a GOOSE telegram is transmitted spontaneously and is repeated after 1 ms, 2 ms, 4 ms etc. before returning to cyclic operation.
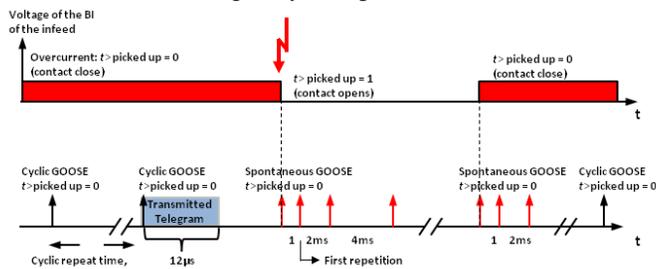
Figure 5. Transmission of binary states with GOOSE messages

Typical examples for GOOSE application in substation automation comprises:
− Tripping of switchgear
− Starting of disturbance recorder ("Störschrieb")
− Providing position status of interlocking

Security considerations for the GOOSE communication are available and are discussed as part of the next section.

III.   SECURITY FOR SUBSTATION AUTOMATION MULTICAST MESSAGES

Security has been acknowledged as a basic requirement for substation automation. The main security requirements especially for GOOSE and SV communication have been determined as message integrity and source authentication.

Within the standard IEC 62351-3 a security solution is provided, which exactly addresses these requirements for the transfer of GOOSE and SV messages in multicast Ethernet networks. The basic approach taken here builds on digital signatures. They are used to basically calculate a checksum over the payload of the Ethernet PDU (Protocol Data Unit). The transport of the security related part is defined as an extension to the existing definition of the GOOSE or SV PDU. Digital signature calculation presents a higher load to the IED, especially if retransmissions are taken into account. Moreover, at a sample rate of 80 samples per power cycle, there are up to 4000 packets per second for the common frequency of 50 Hz. If those messages carry a digital signature, it places a high burden for the sender during the generation of the digital signature and also on the receiver

for verifying the signature. IEDs are typically not built to handle this type of operation at that speed. This has been verified by prototypes building on FPGAs. Therefore, there exists a demand for an alternative solution.

Beside the discussion of exchanging GOOSE and SV packets within a substation, there is also a request to transmit this information for synchrophasor application in distributed environments over wide area networks. This is depicted in the technical report IEC 61850-90-5 (cf. [20]). Here, Ethernet will not be the base for communication but UDP/IP instead, which also allows for multicast. A new requirement arising here is the provisioning of confidentiality for the data. This requirements stems from the fact, that the synchrophasor information is interesting to determine the load and stability of a dedicated electricity network. While this information is protected in a substation by physical means, it needs to be protected when communication over wide area networks based on sound cryptographic methods. Note that the discussion of confidentiality is not part of this paper. To better cope with the required performance, IEC 61850-90-5 proposes to rely on integrity check values (ICV), which are calculated using HMAC-SHA256 or AES-GMAC involving a shared key, rather than using digital signatures. This shared key is supposed to be a group based key, shared among the configured participants of a group. A key distribution center is responsible for authenticating the group participants and generating and distributing the shared group key to authenticated peers. The underlying key distribution protocols is Group Domain of Interpretation (GDOI, cf. [21]), which has already proven its feasibility in many router implementations to distribute group keys for multicast services in the Internet. The integrity check is applied in the processing in a similar way as the digital signature. The sender creates the ICV, while the receiver checks the ICV upon receiving the message, before executing a command.

The following subsections discuss multicast authentication options in general and propose the application of authentication schemes for dedicated messages that allow for the delayed verification of message integrity of already received messages.

IV. EXISTING APPROACHES FOR MULTICAST AUTHENTICATION

Many widely used security protocols as IPSec [4] and SSL/TLS [5] are designed mainly for point-to-point communication. However, the communication type of multicast requires specific handling. The objective of security within substation automation is to ensure the integrity and authenticity of messages. Protection the confidentiality is not required, however.
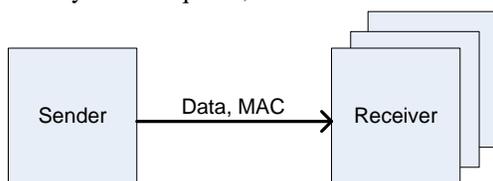
Figure 6. Broadcast/Multicast Sender Authentication

Figure 6 shows the basic set-up. A sender sends a message containing data protected with a message authentication code MAC. Several receivers verify the received message. Cryptographic authentication of multicast communication comprises to main parts:

- Message protection: A data packet or frame has to be protected (encryption and/or message authentication). A cryptographic checksum (message authentication code) is applied to a message that is verified by the receivers.

- Multicast Key management: The cryptographic keys required by the sender and by the receiver have to be established.

Conceptually, the problem would be solved by applying a digital signature scheme based on public key cryptography, e.g., PKCS#7 [6] or DSA [7]. However, the computational requirements of these algorithms render them inadequate for the targeted field level devices as already discussed in section III above. So a message level protection based on symmetric algorithms as AES-CBC-MAC, AES-GMAC, or HMAC-SHA256 [7] is rather used. The sender and the receiving nodes apply the same secret key for creating and for verifying the cryptographic checksum.

Various protocols have been designed for group key management, e.g., the Group Key Management Protocol (GKMP) [8] and Scalable Multicast Key Distribution [9]. Group Secure Association Key Management Protocol (GSAKMP) [10]. A survey [11] of group key management protocols describes different options for group key management in centralized environments. Also common wireless communication standards support secure multicast/broadcast communication, e.g., IEEE 802.11 WLAN [12] and 3GPP Multimedia Broadcast/Multicast Service [13]. The basic design idea is to rely on a group key management server that authenticates group members and establishes group keys for protecting communication within the group. There exist also decentralized approaches for group key establishment that do not require a group key server, e.g., Group Diffie–Hellman Key Exchange [14]. All these approaches result in symmetric group key shared between the members of the group. So each node can send and verify protected group messages. No authentication of the sending node is achieved, as each group member knows the group key that can be used for both sending and receiving messages. A specific key management based on key chains can be used to achieve sender authentication. An element of the key chain is valid for sending only during a limited, defined time period. During that time period, it is known only by the sender. Only after the time validity has passed, the key is released to receiving nodes. To verify a received message, a receiving node has to store the received message until it has received the corresponding key. Only after receiving also the key, the receiver can verify the received messages. This leads to a delay in processing of the messages.

The Timed Efficient Stream Loss-tolerant Authentication protocol (TESLA) [15] provides sender authentication. TESLA is based on loose time synchronization between the sender and the receivers. Source authentication is realized in

TESLA by using Message Authentication Code (MAC) [7] using a symmetric key of a one-way key chain.

Figure 7. Hash Key Chain

Figure 7 illustrates the concept of a hash key chain. The hash key chain of length n is determined by the sender starting with a randomly chosen key K_{init} that is valid during a time period tn-1. The sender computes the keys K_i using a cryptographic hash function H as the hash of the key K_{i+1}, i.e., $K_i := H(K_{i+1})$. The key Ki is valid for sending messages only during the time period t_i. But the sender releases the key K_i only after the time period t_i has already passed, i.e., when the key is not valid for sending anymore. A receiver can verify messages received during the time period ti only after ti has passed, i.e., after having received the key. However, a malicious receiver cannot forge messages on behalf of the sender as the key is already invalid.

The sender provides the first key K_0 to receivers in a secure way (i.e., protected by a digital signature or provided over a protected communication channel). Each receiving node stores the key K_0. Further keys K_{i+1} are released by the sender in clear as a receiver can verify the authenticity of the released key efficiently by computing its hash value. Due to the one-way property of the hash function H, a receiver cannot practically determine a key K_{i+1} from a known K_i.

The important property of the one-way key chain is that once the receiver has obtained a single authenticated key of the chain, subsequent keys of the chain are self-authenticating. This means that the receiver can easily and efficiently authenticate subsequent keys of the one-way key chain using the one authenticated key. The initially distributed message is protected using a well-known digital signature.

µTESLA addresses sensor network scenarios and optimizes the TESLA protocol for this use case [16]. The general setup assumes a base station, which has an authenticated connection to sensor nodes based on a shared secret. As the digital signature for the initial message protection in TESLA is too costly for sensor nodes, µTESLA addresses this by using the node-to-base-station authenticated channel to bootstrap the authenticated broadcast. The remainder of the protocol is similar to the original TESLA approach.

V. Enhancements for Substation Automation Multicast Security

A new solution is proposed for the authentication and integrity protection of broadcast/multicast control messages. It combines hash key chains with digital signatures. This solution can be applied in particular to a field-level energy control protocol (e.g., a substation controller).

To avoid a centralized node as single point of failure each sending node manages its own key chain. As in TESLA, the initialization information of a hash key chain is protected by the sender using a digital signature. Synchronized time is already available in energy automation

using Network Time Protocol (NTP) [17] per substation. A GPS receiver is attached to the substation controller to provide the reference time for all connected components. If a GPS device is not available, the time information may also be received from a hierarchically higher system component like a control center over other signaling channels.

Known enhancements to the basic TESLA scheme support immediate authentication by using buffering by the sender [17]. However, this requires that the sending node has to already have the information about the contents of future packets. This makes it unsuited for real-time control applications where the future changes in the physical world are not known in advance. Furthermore, the usage of multiple key chains has been proposed where a sending node manages multiple hash chains for receivers observing different network delays.

The following subsections describe new enhancements to TESLA to cope with the specific requirements of a real-time control network.

A. Multiple Message-class specific Hash Chains

A sending node manages multiple hash key chains. A hash key chain message is bound to a certain class of control messages. The class of control messages is specified by the sender as part of the hash chain's initialization information. This allows a receiver to determine whether an announced hash chain includes potentially control commands relevant for the receiver. Only if this is the case, the receiver has to store the initialization information. A receiver may also verify that a received control message is in fact of the class as announced in the hash chain initialization information.

B. Hierarchical Hash Key Chains

In TESLA, each hash key chain initialization information is protected by a separate digital signature. It is proposed to establish a first hash key chain that is used to protect initialization information of further hash key chains. This is in particular advantageous if several hash key chains are established for different message classes. Also, hash chains which have to be established frequently as they may have a short time delta between hash chain values can be established efficiently.

C. Early control command execution

When using a hash chain, a receiver can verify the cryptographic checksum a received control message only after a certain delay (when the next element of the hash chain is disclosed by the sender). This leads to a non-negligible delay. It is therefore proposed that for some classes of commands the receiver performs the control action immediately after receiving the message, i.e., before verifying the command's cryptographic checksum. However, roll-back information is stored by the receiver. Should the checksum be invalid (once it is verified later), an inverse control operation is performed, neutralizing the effect of the invalid control command. If the checksum is valid, the roll-back information is deleted to free occupied memory. In an enhancement, this early command execution is performed

only for certain control commands, e.g., for which parameter values have passed a plausibility check.

D. Evaluation

The properties of the new enhancements are evaluated. Performance requirements on field level devices are reduced even further as a device processing only low data rate or low real time requirements has to process only messages of a corresponding hash key chain. The number of digital signature verifications is kept low as the hash key chain initialization information of the multiple key chains is protected by a hash key chain itself. The design fits with the existing solutions, supporting publish/subscribe communication, and avoiding any central controller. It is one option that can be used in combination with currently defined options.

However, still support for digital signatures is required. This may be avoided by using the µTESLA approach in such cases where a substation controller is available to distribute the initial group key in an authenticated way. Also the time delay caused by the period of uncertainty between reception and verification of a message is still occurring, making it inappropriate for control traffic requiring a very short reaction time (e.g., an emergency power switch off in case of overload). So, there is basically a trade-off whether immediate reaction to a control command is more important than sender authentication. The described approach of defining different security solutions for different message classes allows addressing application-specific side conditions by the security solution. For example, it is possible that a power on command is accepted only with sender authentication, while emergency power off is performed using normal group membership authentication. The susceptibility to denial-of-service attacks is not necessarily increased as control equipment could also provide wrong, manipulated measurements or control command by themselves (independent of any cryptographic authentication scheme).

VI.   INTEGRATION IN SUBSTATION AUTOMATION PROTOCOLS

The described approach for multicast sender authentication can be integrated in existing field level energy automation protocols transmitting GOOSE or SV information. This has the following implications on field level devices:

- Each field device requires a public private key pair to protect the initialization information. The public key is certified and available for other field devices.

- A disclosure schedule is known to all entities upfront, e.g., fixed or defined during engineering.

- The field device has to generate a hash key chain of determined length n $(h_0, h_1, h_2, ..., h_{n-1})$. The length is determined by the time interval t_A that shall be covered by the overall hash key chain. Other factors are the storage requirements of messages at the receiver side. This time interval t_A is then divided into subintervals t_I.

Each subinterval is associated with a key from the hash chain $(t_0, h_n, t_1, h_{n-1} ... t_n, h_0)$.

The operation proceeds as follows:

- Step 1: Initialization of the Hash Chain by an IED

The field device sending GOOSE or SV broadcast/multicast messages provides the last value of the hash chain as part of a GOOSE or SV message and protects this message before sending it. The field level device uses a digital signature, or a higher-hierarchy hash key chain. The field device includes a description (manifest) of the message type protected with this hash chain. All subscribers will receive the message, and upon successful verification they will store the hash value together with an identifier of the sender. This identifier may be a MAC address, a serial number or similar.

- Step 2: Sending protected broadcast/multicast messages by a field device

After step 1, the time interval t_1, starts that is associated with the hash value h_{n-1}. The field device now uses a keyed hash for this time interval to protect the integrity of the GOOSE or SV values. The receiver has to store the messages until the sender has released the hash value h_{n-1}. This value can be released after the time interval has ended. The value can be released in clear. The receiver can now calculate the integrity check value of the stored message to achieve a delayed authentication of these messages.

An advanced variant of the key disclosure schedule may alternatively depend on the number of messages sent. Another advanced variant of the key disclosure schedule may alternatively depend on the priority (e.g., depending on the performance class) of the message sent.
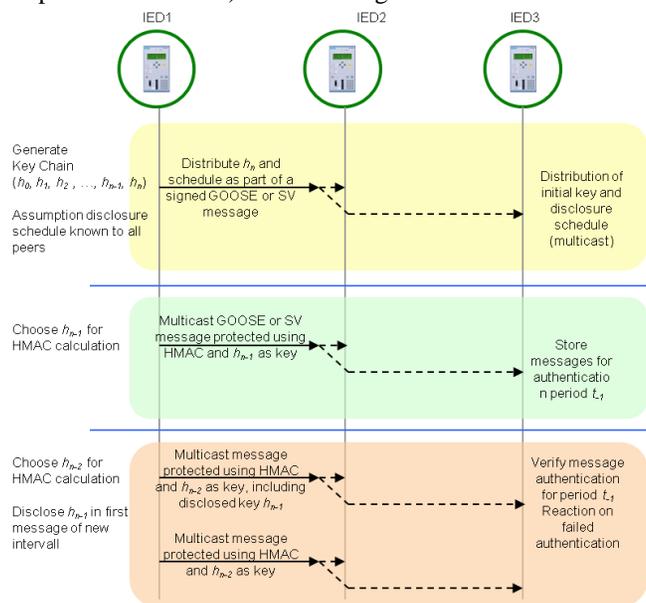

Figure 8. Broadcast/Multicast Control Message Sender Authentication in Field Level Energy Automation

As shown above, the general approach for protection of the distribution of the initial group key can be followed, allowing for authentication based on digital signatures (as in TESLA or as in IEC 61850-90-5) while the handling of the actual messages is protected using symmetric key application.

VII. CONCLUSIONS AND OUTLOOK

This paper described an application space, were multicast authentication is used in energy automation environments like substation communication to protect commands or sampled values send via GOOSE as defined in IEC 61850. As shown the currently provided security mechanisms in IEC 62351-6 to ensure source authentication and message integrity provide for very good security. The application of this approach is hindered by the typical hardware used in IEDs, which does not cope with the performance requirements of the implied cryptographic operations matching the time restrictions of the deployment environment.

This paper analyzed other multicast authentication schemes as alternative solutions for the intended use case. It investigates especially into the application of TESLA and mapped the protocol to the substation automation use case. TESLA provides a solution for delayed authentication allowing an IED to perform a dedicated action in real-time and to perform the associated security check later on. It is obvious that there is a period of uncertainty between reception and verification of a message, making it inappropriate for control traffic requiring a very short reaction time (e.g., an emergency power switch off in case of overload) for actions, which may not be reversible. So, there is basically a trade-off whether immediate reaction to a control command is more important than sender authentication. It is also possible to support different multicast authentication schemes within one technical solution and to use the described approach only for timely critical messages, while other messages may use the typical approach verifying a message, before operating on the content. Additionally, combining solutions allows for in-time authentication as a group member, while the delayed authentication can be used to identify an individual sender.

The described approach has not been implemented, yet. Hence, performance numbers and especially performance comparisons of the different approaches cannot be delivered at this time.

REFERENCES

[1] M. Felser, "Real-time Ethernet – industry prospective," Proc. IEEE, vol. 93, no.6, June 2005, pp. 1118-1128, http://www.felser.ch/download/FE-TR-0507.pdf [retrieved: Oct. 2012]

[2] T.S. Sidhu, M.G. Kanabar, and P. Palak, "Implementation issues with IEC 61850 based substation automation systems," Proc. Fifteenth National Power Systems Conference (NPSC), Dec. 2008, http://romvchvlcomm.pbworks.com/f/p274.pdf [retrieved: Oct. 2012].

[3] IEC 61850-5 – "Communication requirements for functions and device models", July 2003.

[4] S. Kent and K. Seo, "Securit Architecture for the Internet Protocol", Internet RFC 4301, Dec. 2005, http://tools.ietf.org/html/rfc4301 [retrieved: Oct. 2012].

[5] T. Dierks and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", Internet RFC 5246, Aug. 2008, http://tools.ietf.org/html/rfc5246 [retrieved: Oct. 2012].

[6] B. Kaliski, "PKCS#7 Cryptographic Message Syntax Version 1.5, Intenret RFC2315, March 1998, http://tools.ietf.org/html/rfc2315 [retrieved: Oct. 2012].

[7] C. Paar, J. Pelzl, and B. Preneel, "Understanding Cryptography", Springer, 2010.

[8] H. Harney and C. Muckenhirn, "Group Key Management (GPMP) Architecture", Internet RFC 2094, July 1997, http://tools.ietf.org/html/rfc2094 [retrieved: Oct. 2012].

[9] A. Ballardie, "Scalable Multicast Key Distribution", Internet RFC 1949, May 1996, http://tools.ietf.org/html/rfc1949 [retrieved: Oct. 2012].

[10] H. Harney, U. Meth, and A. Colegrove, "GSAKMP Group Secure Association Key Management Protocol", Internet RFC 4535, June 2006, http://tools.ietf.org/html/rfc4535 [retrieved: Oct. 2012].

[11] S. Rafaeli and D. Hutchison, "A Survey of Key Management for Secure Group Communication", ACM Computing Surveys, Vol. 35, No. 3, pp. 309-329, Sep. 2003, http://merlot.usc.edu/cs530-s08/papers/Rafaeli03a.pdf [retrieved: Oct. 2012].

[12] IEEE 802.11 "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Part 11", 2007.

[13] 3GPP TS33.246, "3G Security; Security of Multimedia Broadcast/Multicast Service (MBMS)", 2012, http://www.3gpp.org/ftp/Specs/html-info/33246.htm [retrieved: Oct. 2012].

[14] M. Steiner, G. Tsudik, and M. Waidner, "Diffie-Hellman Key Distribution Extended to Group Communication", Proceedings of the 3rd ACM conference on Computer and communications security, pp. 31 – 37, ACM CCS96, 1996, http://corsi.dei.polimi.it/distsys/2007-2008/pub/p31-steiner.pdf [retrieved: Oct. 2012].

[15] A. Perrig, D. Song, R. Canetti, J.D. Tygar, and B. Briscoe, "Timed Efficient Stream Loss-Tolerant Authentication (TESLA): Multicast Source Authentication Transform Introduction", Internet RFC 4082, June 2005, http://tools.ietf.org/html/rfc4082 [retrieved: Oct. 2012].

[16] A. Perrig, R. Szewczyk, D. Tygar, V. Wen, and D. Culler, "SPINS: Security Protocols for Sensor Networks", Proceedings of the 8th Wireless Networks, pp 521-534, July 2002, http://www.csee.umbc.edu/courses/graduate/CMSC691A/Spring04/papers/spins-wine-journal.pdf [retrieved: Oct. 2012].

[17] D. Mills, U. Delaware, J. Martin, and W. Kasch, "Network Time Protocol 4: Protocol and Algorithms Specification", Internet RFC 5905, June 2010, http://tools.ietf.org/html/rfc5905 [retrieved: Oct. 2012].

[18] A. Perrig, R. Canetti, D. Song, and J.D. Tygar, "Efficient and Secure Source Authentication for Multicast", Network and Distributed System Security Symposium, NDSS '01, 2011, http://users.ece.cmu.edu/~adrian/projects/tesla-ndss/ndss.pdf [retrieved: Oct. 2012].

[19] "Efficient Energy Automation with the IEC 61850 Standard Application Examples", Siemens AG, December 2010, http://www.energy.siemens.com/mx/pool/hq/energy-topics/standards/iec-61850/Application_examples_en.pdf [retrieved Oct. 2012].

[20] IEC 61850-90-5 – "Use of IEC 61850 to transmit synchrophasor information according to IEEE C37.118", December 2011

[21] M. Baugher, B. Weis, T. Hardjono, and H. Harney, "The Group Domain of Interpretation", RFC 3547, July 2003, http://tools.ietf.org/html/rfc3547 [retrieved: Oct. 2012]

A Energy Balancing Control Strategy for Microgrid with Storage Systems

Jing Tianjun, Niu Huanna, Wang Jiangbo, Yang Rengang

College of Information and Electrical Engineering

China Agricultural University

Beijing, China

jtjy11@cau.edu.cn

Abstract— **In a microgrid, there is usually an energy storage system, while there is no continuous energy supply. The storage system does not only be used as a power supply in islanded state, but also could be taken as the demand side management devices in grid-connected state. In this paper, a system energy managing strategy is proposed. It can balance the energy in autonomous state, and manage power in grid-connected state under time-of-use pricing. Next, the droop control and economic dispatch method are introduced into the power management strategies to improve the power supply reliability and perform peak load shaving for maximum benefits. Finally, a simulation is shown to analyze the benefit from the strategies. And the applications of the economic dispatch method are discussed in the results under different operating conditions and market policies.**

Keywords-microgrid; storage system; energy balancing.

I. INTRODUCTION

For electric loads that are described as the unpredictable fluctuation with obvious difference between peak and off-peak load, power system planners must consider the maximum load in the plan of the electric systems expansion. In order to utilize the existing equipment sufficiently and reduce the distribution equipments investment, electric administrative department usually take measures of 'peak load shaving' (also called time-of-use pricing (TOU), load control and management, etc.) to solve this problem. Now, it is available to store electrical energy on a large scale with the development of energy storage devices and control methods.

Microgrid are the distribution networks comprising various distributed generators (DGs) and storage devices that can operate either interconnected or isolated from the main distribution grid, as in [1]. The microgrid voltage could cover the whole distribution networks, but only the low-voltage is considered in this paper. Storage devices in microgrid can be controlled flexibly to minimize microgrid operation. Some power management strategies are presented, as in [2, 3], for optimal economic operation of the microgrid, including making power and planning, according to the prediction for PV power production and the load forecasting. It could reduce the costs of microgrid and optimise battery charge states. Sortomme and El-Sharkawi [4] used particle swarm optimisation to achieve optimal dispatch of controllable loads and generators as well as effectively utilizing the battery storage of each microgrid. Kim et al. [5] proposed a cooperative control strategy of energy storage system and micro sources for stabilizing the microgrid in islanded operation mode [6-8], consisting of the centralized

controller and local controllers interconnected with communication bus [9-11]. Now the storage system has been used to microgrid control, but has not yet been used to energy balancing control.

A detailed analysis of the generic control scheme for energy storage system (ESS) is presented. Based on the analysis, this paper proposes a coordinated control strategy, using a ω-P/V-Q and P-ω/Q-V droop control method, to avoid the real and reactive power coupling problem cased by the high R/X line impedance ratio [12-14]. An economic dispatch method for energy storage systems, under time-of-use pricing, is presented to maximise benefits, and the spare economic benefits of ESS is considered.

In the paper, the control strategies of the ESS in both grid-connected and island mode is proposed. And then, the ESS dispatch method based the control strategies is discussed. At last, the dispatch method is tested in the case study.

II. CONTROL STRATEGIES OF THE ESS

A. The ESS in a microgrid system

ESS plays a vital role in the reliable operation of a microgrid. Figure 1 shows a single-line diagram of a typical distributed ESS in a microgrid. The centralized control system of the ESS comprises local micro source controllers (MCs) and central controller. The MC uses local information to control the voltage and the frequency of the ESS in transient operation.
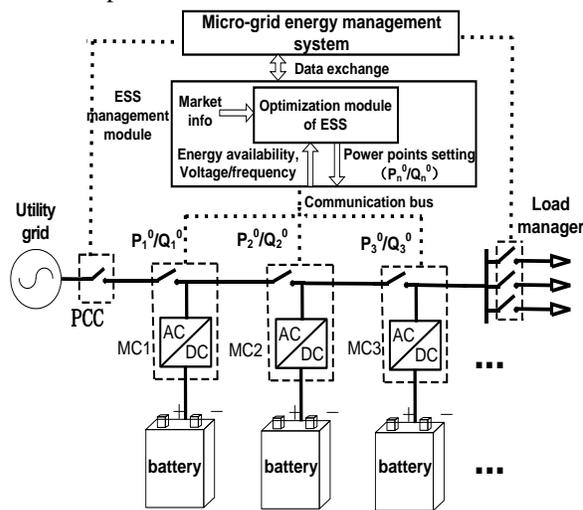
Figure 1. Typical ESS structure in a microgrid

The MC also follows the demands from the central controller to optimize the operation. The decentralized control method for the parallel operation of inverters working in island mode is the frequency and voltage droop method, and the power management strategies are based on local measured signals without communications assistance. The ESS management module is responsible for the dispatch of the ESS and the maximization of the ESS's value. The dispatch method is related to the structure and operation modes of the microgrid, and has an effect on the microgrid stability. The proposed hierarchical control system of ESS is shown in Figure 1. The battery storage inverter with island detection function changes the control scheme based on the microgrid mode of operation, i.e. islanded or grid-connected, and the ESS management module sends dispatch commands to the local controllers according to the amount of available energy, electricity price and system electric parameters.

B. Control methods of the ESS in grid-connected mode

In the grid-connected mode, the utility grid is expected to supply the difference in real/reactive power. Similar to a conventional utility system, each DG unit can be controlled to generate pre-specified real and reactive power components (PQ-bus) or generate pre-specified real power and regulate its terminal voltage (PV-bus). Optimization power management of the ESS can help to minimize the tie line power (peak shaving), and maximize the ESS's value under TOU price.

Figure 2 shows a block representation of a generic control system of the ESS inverter in grid-connected mode. A signal processing block and a phase locked loop (PLL) block are used to process the measured currents and voltages of the ac-bus and to estimate the local frequency. The frequency estimation is used to synchronization and to track relative angle of the inverter reference frame. Controls of the real/reactive power are implemented in a dq0 reference frame that determines d- and q-axis components of the ac-side currents. The power set-point is determined by the ESS power management module.
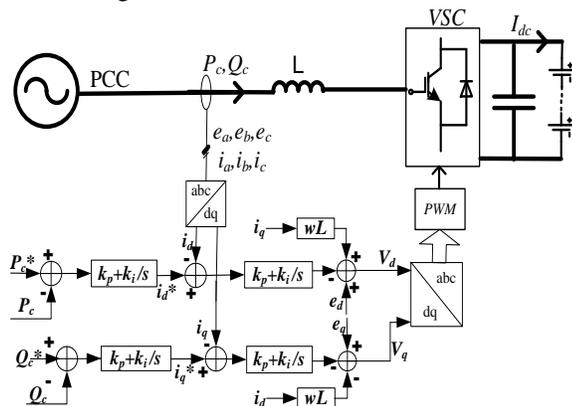
Figure 2. P-Q control scheme for the ESS

The dispatch method of ESS power management module in grid-connected mode is based on electricity price, to achieve economic benefit maximization of ESS and load

peak shaving. A ω-P/U-Q droop method of ESS control is shown in Figure 3.
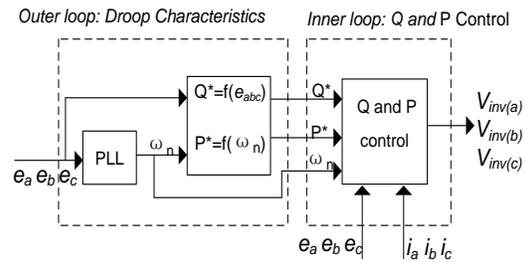
Figure 3. ω-P / V-Q droop controller for ESS in grid-connected mode

The method in Figure 3 improves the microgrid stability. The outer loop determines the set points for P*and Q*, according to the frequency measured by PLL and voltages of ac-side. The droop characteristics can be mathematically represented as

$$P_n = -1/K_{wn}(w_n - w_0) + P_n^0$$
$$Q_n = -1/K_{Vn}(U_n - U_0) + Q_n^0 \quad (1)$$

where K_{wn} and K_{Vn} are the droop characteristic slope for the nth battery unit, ω_0 and U_0 are the reference frequency and reference voltage of the microgrid, respectively, and the P_n^0 and Q_n^0 represent the initial real/reactive power generation assigned to the unit, respectively. The inner loop utilizes independent real/reactive power control method to ensure the output power to be the pre-specified P* and Q*. The control method is similar to the one used in conventional power systems for the correct load sharing among generators feeding to ac system, and can help to support the frequency and voltage of the microgrid.
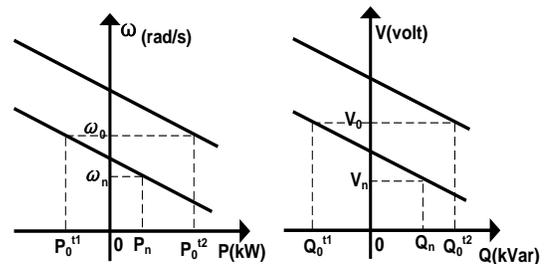
Figure 4. Frequency and voltage droop control characteristics

This droop method of ESS can also be applied in islanded microgrid with a slack bus to minimize the dynamics and system oscillations. The frequency and voltage droop control characteristics are shown in Figure 4.

C. Control method of the ESS in islanded mode

In grid-connetcted mode, the ESS power management discussed in this paper can be applicable to other controllable sources as well, and the unstable sources are controlled based on optimal power generation schemes to deliver maximum available power. Whereas in islanded mode, the microgrid central controller is responsible to the regulation of the charge and discharge power of ESS, output power of DG units, and load shedding, to match generation and load

demand. The objective of the ESS control system is to maintain the system frequency and the voltage magnitude, rather than load peak shaving.

The power control system of ESS encompasses real power generation control and the reactive power control block. The P-f and Q-V droop characteristic is introduced as the primary control to share load and regulate the voltage without communication among the fast acting inverters. This cooperative control method of ESS and micro sources is chosen because none of the inverters can dominantly support the frequency and voltage of the system in islanded microgrid. Base on mainly inductive line impedance between the inverters, the frequency control is coupled with the real power flow control, and the inverter voltage is controllable through Q.

Figure 5 shows a block representation of a generic control system of the ESS inverter in islanded microgrid. The control structure of parallel voltage sources converter adopted droop method encompasses two control loops, as shown in Figure 6. The outer loop determines the reference ω^* and E^*, and expressed as in (2).

$$w_n = -K_{wn}(P_n - P_n^0) + w_0$$
$$E_n = -K_{Vn}(Q_n - Q_n^0) + E_0 \qquad (2)$$

where K_{wn} and K_{Vn} are the droop characteristic slope for the nth battery unit, ω_0 and U_0 are the reference frequency and reference voltage of the microgrid, respectively, and the P_n^0 and Q_n^0 represent the initial real/reactive power generation assigned to the unit, respectively. The inner loop controls the ac-bus frequency and the voltage of converter according to ω^* and E^*. In islanded microgrid, the output real/reactive power of ESS inverters is vary depending on the ω_n and E_n, not pre-specified, and the cardinal control objective is reliability of power supply, rather than the economic dispatch of ESS.
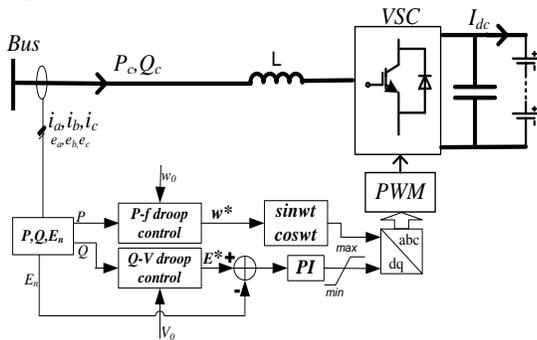
Figure 5. V-f control scheme for the ESS
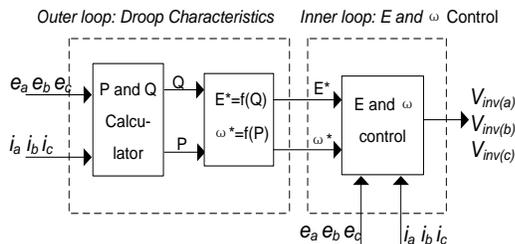
Figure 6. P-ω/ Q-V droop controller for ESS in islanded mode

However, when the traditional droop method is implemented in a low voltage microgrid, the different inverter output impedances and the high R/X line impedance ratio lead to severe real and reactive power control coupling and system stability concerns. So the droop method should be improved. This paper presents the combination control strategy of ω-P/V-Q droop and P-ω/Q-V droop method to solve the real/reactive power coupling problem and improve the stability performance. In the distributed storage system structure shown in Figure 1, the MC1 adopts P-ω/Q-V droop control method after the transition from grid-connected mode to islanded mode, while the other MCs adopts ω-P/V-Q droop control regardless of the microgrid mode of operation. This combination control strategy uses the system frequency and voltage as a communication means without causing real/reactive power coupling problem in islanded operation.

III.   ECONOMIC DISPATCH OF THE ESS

The objective of the microgrid energy management module is to generate suitable set-points for the storages and micro sources in such a way that economically optimised power dispatch will balance the certain load demand. The output of the ESS management module defines the direction and amount of the power flow between storages, especially in grid-connected mode, to realize the maximum economic benefits of the ESS. The economics of the ESS in islanded microgrid are particular complex, because it related to other controllable sources and loads. But in grid-connected mode, the ESS can be dispatched according to the TOU price. Thus, the energy stored in the ESS is used as the state variable. Time stages are calculated with unit 1h, and most stage more than 1hour. Energy stored in the batteries is expressed as follows.

$$Q_{n,t+1} = \begin{cases} Q_{n,t} + A_n * P_{n,t} * \Delta T & (\text{ if the ESS is charging }) \\ Q_{n,t} - A_n * P_{n,t} * \Delta T & (\text{ if the ESS is discharging }) \\ Q_{n,t} & (\text{ if the ESS is idle }) \end{cases} \quad (3)$$

$$A_n = \begin{cases} 1/\eta_C \\ 0 \\ \eta_D \end{cases}$$

where $Q_{n,t}$ is the energy stored in the nth storage units at hour t, $P_{n,t}$ is the real power of the nth storage units output at hour t, A_n is the nth storage units charging/discharging efficiency(η_C:charging factor, η_D:discharging factor).

The ESS sells stored energy at high price and shaves the peak loads of the large system, and stores it when the electricity is abundant and cheap. The objective function for each hour intervals and the physical constraint are expressed as in (4), and it aimed at maximizing the generation benefit.

$$MaxF(P_{n,t}) = \sum_{t=1}^{T} \sum_{n=1}^{N} * c_t * P_{n,t} * \Delta T$$

$$s.t. \begin{cases} Q_{\min} \le Q_{n,t} \le Q_{\max} \\ P_{\min} \le P_{n,t} \le P_{\max} \\ V_{\min} \le V_{n,t} \le V_{\max} \\ f_{\min} \le f_{n,t} \le f_{\max} \end{cases} \qquad (4)$$

Where c_t is the electricity price at time t, $\triangle T$ is the time stage, T is total number of hours, and N is total number of storages. The constraints in optimization module includes unit capacity, rated power, and the variation of the V and f of the system, because of the applied of droop control.

The ESS increases the reliability of power supply by supporting certain loads demand in microgrid, when the utility electric grid is unavailable. Thus, the benefits of the emergency reserve of the ESS should be calculated. The benefits encompass the spinning reserve capacity $C_{rsc,n}$ and the reserve electricity $C_{rsp,n}$. The spinning reserve capacity benefit is related to both the remaining capacity of the battery and the rated power of the ESS inverters. And it can be expressed as in (5).

$$C_{rsc,n} = Q_{rsoc} * P_{rp} * \rho_{rp,n} \qquad (5)$$

P_{rp} is the rated power of the ESs inverter, $\rho_{rp,n}$ is the price of the spinning reserve capacity. Consumers that chose the reserve services should pay the reserve capacity fee according to the load no matter grid failures occur or not. When the grid failures occurs, the consumers should pay the reserve electricity benefit as

$$C_{rsp,n} = \rho_{rsp,n} * Q_{rsoc} * \eta_D \qquad (6)$$

Thus, the total reserve benefits of the ESS is expressed as

$$C_{r,n} = C_{rsc,n} + P_{rh}\% * C_{rsp,n} \qquad (7)$$

Where $\rho_{rsp,n}$ is the price of the reserve electricity, $P_{rh}\%$ is the probability of the grid failure happens. The objective function of the ESS dispatch module taking account of the reserve benefits is expressed as

$$MaxF(P_{n,t}) = \sum_{t=1}^{T}\sum_{n=1}^{N} c_t * P_{n,t} * \Delta T + C_{r,n} \qquad (8)$$

IV. CASE STUDY

In this section, it takes a 500Ah/600V lithium batteries system rated power 100kW as a study case to analyse the economic dispatch in the paper. The energy management of a distributed storage system in a grid-connected microgrid is formulated as a linear programming problem. The market prices of a day is given in Table 1, and the parameters is set as follows, $\eta_C = \eta_D = 0.9$, $Q_{min} = 90kWh$，$Q_{max} = 300kWh$, $P_{max} = 100kW$, and the initial remaining energy of the batteries is $Q_s = 90kWh$. Ignoring the reserve benefits, it can be noted that benefits can be made only if ESS efficiency is greater than the off-peak/peak price ratio, i.e.

$$\eta_D * \eta_C > P_{off-peak} / P_{on-peak} \qquad (9)$$

TABLE I. Electricity TOU PRICE

Time/h	0-8	8-12	12-14	14-18	18-22	22-24
Price	25%	100%	50%	100%	75%	50%

When 1kWh power is stored from the grid in the off-peak, no more than 1 kWh can be return to the grid in the peak, because of the convert loss. So, the benefit ratio for the storage should be higher than the loss.

TABLE II. POWER SETTINGS OF THE ESS

Time/h	0-8	8-12	12-14	14-18	18-22	22-24
P_t/kW	29.2	-47.2	100	-40.5	0	0

According to formula 4, the power set points of the ESS of the day are given in Table 2, and the benefits of the ESS is 192.7 standard kWh per day calculated as follow.

$$MaxF(P_{n,t}) = \sum_{t=1}^{T} c_t * P_{n,t} * \Delta T = 192.7 kWh \cdot$$

TABLE III. POWER SETTINGS OF THE ESS INCLUDING RESERVE SERVICE

Time/h	0-8	8-12	12-14	14-18	18-22	22-24
P_t/kW	18.7	-37.5	75	-37.5	0	0

If the reserve benefits is considered, the parameters is set as follows, $\rho_{rp,n}=0.1\%$, $\rho_{rsp,n}=150\%$, $P_{rh}\%=0.5\%$, $Q_{rsoc}=60kWh$, and the available capacity for dispatch under market price is $Q_a = Q_{max} - Q_{rsoc} - Q_{min}$. According to formula 8, the power set points of the ESS of the day is given in Table 3, and the benefits of the ESS offering reserve service is 151.4 standard kWh per day calculated as follow.

$$MaxF(P_{n,t}) = \sum_{t=1}^{T}\sum_{n=1}^{N} c_t * P_{n,t} * \Delta T + C_{r,n} = 151.4 kWh$$

V. CONCLUSION

In this paper, the detailed control method analysis of the ESS is presented in island and grid-connected mode. The droop control method can be applied in both modes operation to improve the grid transient performance. Mathematic model for analyzing the economic effect of the energy storage system is proposed under the TOU price. Also, the reserve benefits of the ESS are also taken into account in this model. The ESS management module can make an optimal dispatch schedule according to the electricity price to maximize benefits.

REFERENCES

[1] A. G. Tsikalakis and N. G. Hatziargyriou, "Centralized Control for Optimizing Microgrids Operation," IEEE Trans. Energy Convers., vol. 23, no. 1 , March. 2008, pp. 241-248.

[2] C. Chen, S. Duan, T. Cai, B. Liu, and G. Hu, "Smart Energy Management System for Optimal Microgrid Economic Operation," IET Renew. Power Gener., vol. 5, no. 3, May. 2011, pp. 258-267.

[3] H. Kanchev, D. Lu, F. Colas, V. Lazarov, and B. Francois, "Energy Management and Operational Planning of a Microgrid with a PV-Based Active Generator for Smart Grid Applications," IEEE Trans. Ind. Electron., vol. 58, no. 10, October. 2011, pp. 4583-4592.

[4] E. Sortomme and M.A. El-Sharkawi, "Optimal Power Flow for a System of Microgrids with Controllable Loads and Battery Storage,". IEEE/PES Power Systems Conf. and Exposition,March.2009, pp. 1–5.

[5] J. Kim, J. Jeon, S. Kim, C. Cho, J. H. Park, H. Kim, and Y. Nam, "Cooperative Control Strategy of Energy Storage System and Microsources for Stabilizing the Microgrid during Islanded Operation," IEEE Trans., Power Electron., vol. 25, no. 12, 2010, pp. 3037-3048.

[6] M. C. Chandorkar, D. M. Divan, and R. Adapa, "Control of Parallel Connected Inverters in Standalone AC Supply Systems," IEEE Trans. Ind. Applicat. vol. 29, no. 1, January. 1993, pp. 136-143.

[7] K. D. Brabandere, B. Bolsens, J. V. D. Keybus, A. Woyte, J. Driesen, "A Voltage and Frequency Droop Control Method for Parallel

Inverters," IEEE Trans. Power Electron., vol. 22, no. 4, July. 2007, pp. 1107–1115.

[8] L. Yan and W. Yun, "Power Management of Inverter Interfaced Autonomous Microgrid Based on Virtual Frequency-Voltage Frame," IEEE Trans. Smart Grid, vol. 2, no. 1, March. 2011, pp. 30-40.

[9] C. Marnay, M. Stadler, A. S. Siddiqui, R. Firestone, and B. Chandran, "Optimal Technology Selection and Operation of Commercial-Building Microgrids," IEEE Trans. Power syst., vol. 23, no. 3, 2008, pp. 975-982.

[10] F. Katiraei and M. R. Iravani, "Power Management Strategies for a Microgrid With Multiple Distributed Generation Units," IEEE Trans. Power syst., vol. 21, no. 4, November. 2006, pp. 1821-1831.

[11] F. Katiraei, M. R. Iravani, and P. W. Lehn, "Small-signal Dynamic Model of a Micro-grid including Conventional and Electronically Interfaced Distributed Resources," IET Gener. Trans. Distrib., vol. 1, no. 3, 2007, pp. 369-378.

[12] Y. Guan, W. Wu, and X. Guo, "Control Strategy for Three-phase Inverters Dominated Microgrid in Autonomous Operation," Proceedings of the CSEE, vol. 31 no. 33, November. 2011, pp. 52-60.

[13] X. Yang and J. Su,"Voltage Control Strategies for Microgrid With Multiple Inverters," Proceedings of the CSEE 32 no. 7, March. 2012, pp. 7-13.

[14] Y. Li and C. Kao, "An accurate power control strategy for power electronics-interfaced distributed generation units operating in a low voltage multibus microgrid," IEEE Trans. Power Electron., vol. 24, no. 12 , 2009, pp. 2977–2988.

Smart Grid Enabled and Enhanced by Broadband Powerline

Strategic value and inherent technical characteristics of Broadband Powerline

Nadine Berezak-Lazarus

bmp Telecommunications Consultants
Weil a.R, Germany
nberezak@bmp-tc.com

Abstract—**For Smart Grid applications, Broadband Powerline presents inherent Unique Selling Points compared to other technologies. Many have promoted various technologies for Smart Grids. In Europe, new generation of Narrowband Powerline seem to gain ground. However, a growing number of utility and industrial companies, concerned about the limits of such widespread technologies, have tested and implemented Broadband Powerline- some with the help of bmp as strategy consultancy. The maturity coupled with the performance of this technology allows to forecast Broadband Powerline as one of the key technologies for Smart Grid, which will enable utility companies to overcome numerous challenges posed by the next generation energy distribution grid. Through this paper, we will describe the technologies forming the Powerline communications and especially, the place of Broadband Powerline among the Smart Grid technologies and its advantages.**

Keywords-Broadband Powerline; Smart Grid; Smart Metering; comparative performance of AMR technologies; USPs.

I. INTRODUCTION

Traditional grid used by Utility companies is facing new challenges concerning the efficiency and flexibility of its applications. The development of Automatic Meter Reading (AMR) with Narrowband Powerline Technologies or other technologies was a first step. However, the needs of real time and permanent connectivity for all grid applications suggest the use of a technology offering higher performances and offering long term viability. Broadband Powerline appears as a technology meeting these needs and offering a global solution.

Through this paper, we will first focus on the overall market of Powerline Communications. Secondly, we will offer an overview of the Broadband Powerline technology, and then its advantages as unique selling points. Finally, we will describe some representative cases concerning Broadband Powerline and assess its competitive position.

II. BROADBAND POWERLINE WITHIN THE LANDSCAPE OF POWERLINE

Using the electrical network to build a two-way communication Internet infrastructure, Broadband over Power Line technology (BPL) is currently taking a leading position [1] for the energy providers which want to upgrade their networks into "Smart Grids".

BPL transmits signals over the distribution networks in the 1 MHz-34 MHz (and beyond) via Orthogonal Frequency Division Multiplexing modulation (OFDM) [2]. This multi-channel modulation allows transmitting the signal over various frequencies, taking maximum advantage of the available frequency bandwidth.

Recent similar developments are being undertaken focusing on the CENELEC A band (under 148 kHz), but even here BPL presents significant and valuable advantages over such Narrowband Powerline technologies.

Used as part of the largest ongoing European projects, such as Linky by ERDF in France or Telegestore in Italy by Enel [3], the narrowband Powerline technology is a suitable solution for basic applications such as meter reading or remote billing. Supported by many international players, we are witnessing the emergence of several industrial standards, seeking to become European or even global references, such as PRIME [4] (led by Iberdrola), G3-Powerline [5] (led by ERDF) or Meters & More [6] (led by Enel). Some leading companies in PLC technology, such as Echelon [7] or Siemens [8], are following this trend and announced their willingness to open their protocols to all manufacturers.

TABLE I.   PERFORMANCE OF NARROWBAND PLC

Narrow-band PLC	Performance [9]			
	Modulation	*Data rates*	*Reliability of data transmis-sion*	*Distance LV/LV*
1st Generation	FSK, S-FSK, BPSK, DCSK, DBPSK, DQPSK	Until 5 Kbps	97% on a 3 days period	700 m - 1,5 km
3rd Generation	OFDM	Prime: 19,2 -128 Kbps G3 up to 300 Kbps	>95% overnight readings	Prime: 4 km G3: 6 km

The market is characterized by a quick evolution of the narrowband Powerline performances. As referred in Table I, the technologies using simple carrier modulations (FSK [10], S-FSK [11], BPSK [12], DCSK [13]) show today a commercial maturity and are industrialized by industry heavyweights such as Echelon [14], Enel or Yitran [15]. However, the low data rates – not exceeding 5kbps – allowed by this 1st generation constrain the implementation of future new applications.

More recently, yet still in development, a new generation using a multi-carriers modulation (OFDM) has been introduced on the markets, enabling higher data rates (from a dozen up to 128 kbps) and using the CENELEC A frequency band. As shown in Figure 1, this technology also increases the robustness of the signal, allowing crossing the transformers in order to communicate over the low and medium voltage lines, which is a clear advantage for electrical network including remote parts of rural areas. Definitely, more suited to the future, this new generation of OFDM-based products emerges, enabling for instance, the use of IPv6.
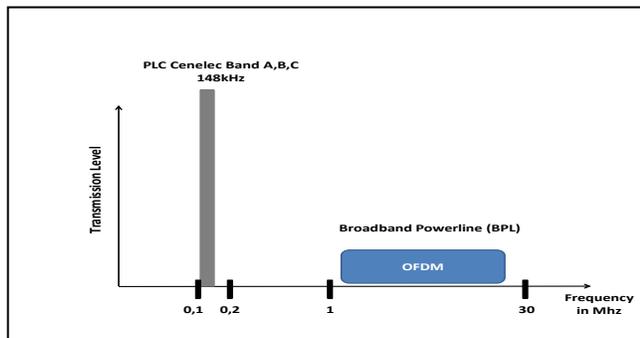
Figure 1.   PLC and BPL basics [16]

To implement the most suitable Smart Grid network regarding the needs of each energy supplier, the key question is to clearly identify the limits of this architecture enabled by narrowband Powerline. In order to ensure the sustainability of the network, it is primordial to take into account the possible evolution of these needs, in addition to the required applications to date. These might entail the implementation of alternative technologies allowing other data rates, and permanent connectivity throughout the energy distribution network.

However, BPL reaches significantly more speed, as, in most of the cases, injecting 200 Mbps on the low voltage enables to benefit from Mbps up to the meter, and enhances the connectivity in enabling real time applications.

Performances of BPL depend on the distance to cover, the electricity environment (type of cable, cables' isolation methods, and type of secondary substation), skills of installer/integrators and the effective noise levels. Until recently, 200/300 Mbps was the maximum gross bandwidth reached by companies; however, current bandwidth can reach 400/500 Mbps or even 1 Gbps but higher speeds being addressed for the inhome market.

BPL current performances are unequalled on the medium and low voltage lines. On one hand, low voltage lines offer 200 Mbps speed over 250-300m without repeater and physical data rates under 1 Mbps over a long range transmission. As seen in the recent CEZ pilot [17] on Figure 2, where more than 1,000 meters have been connected through BPL, all meters have been read, with speeds up to 30 times higher than narrowband PLC having been tested in the same project in 2011. On the other side, BPL on medium-voltage underground lines is performing extremely

well, allowing up to several tens of Mbps physical throughput. Such performances have been achieved with a number of tests and projects such as in Morocco with ONEE- Office National de l'Electricité et de l'Eau- [9] as early as 2008, or more recently in several deployments as backhaul or smart metering data as seen with Iberdrola in Spain [9].
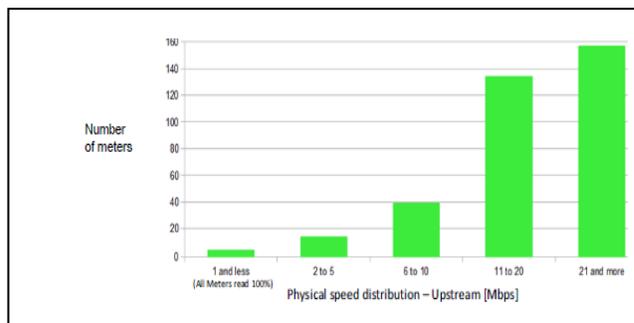
Figure 2.   Physical speed distribution on a secondary substation sample equipped with BPL at CEZ as in October 2011 [17]

Today, BPL on overhead medium voltage lines are in some cases seen to allow 2 Mbps speed over 4 km and at least 200 kbps over 6 to 8 km.

III.   REVIVAL OF BROADBAND POWERLINE

After years of unrealized hopes, BPL activities are now intensifying though the growing interest shown by industrial and utility companies, or venture capitalists.

BPL has long seemed to be unsuccessful, with cases such as IBEC [18], which after having announced a partnership with IBM in order to connect undeserved US American areas via BPL technology, suddenly ceased its activities at the beginning of 2012. Major companies such as Endesa and Itochu, which in 2005 decided to invest in DS2 [19], have disinvested later on, or Earthlink & Consolidated Energy, which, after investing in Ambient [20], has seen the company stopping its broadband Powerline activities meanwhile.

But, the maturity and newest developments of the BPL technology have led various players to re-consider their position towards this specific market.

The first positive factor was the achieved standardization in the field of BPL. Indeed 2010 was the year when two major standards for the BPL industry were reached.

The IEEE 1901 standard is based on the industrial norms Homeplug [21] and HD PLC [22].

The technology specified by IEEE 1901 aims at using transmission frequencies of less than 100 MHz, and allows data rates rating from 100 Mbps. It is also used by all classes of BPL devices, such as BPL devices used for the first-mile/last-mile connection, and broadband services as well as BPL devices used in buildings for LANs, Smart Energy applications, transportation platforms (vehicle) applications, and other data distribution (<100m between devices). Though the standard does not provide interoperability with the two technologies it is based on, it allows benefiting from

the coexistence between these technologies supported by two different MAC layer protocols.

The ITU G.hn [23] standard, based on DS2/Marvell technology norm using OFDM access method, has first focused on the inhome market and can be supported by any wire (coaxial cable, electrical wire and telephone line). Its technology offers a data rate speed up to 1 Gbps but though being non-interoperable with IEEE 1901 based products, it has secured co-existence with these thanks to the Inter System Protocol (ISP), which coordinates the usage of the frequencies among the various domains of BPL implementation.

Such a context has been securing the environment for BPL and has enabled the emerging of new players, such as, just to name a few: Broadcom [24], Mstar [25], Qualcomm Atheros [26], which are chipset providers using IEEE 1901 standard or Lantiq [27], Marvell [28], Metanoia [29] using the G.hn standard. Sigma Designs [30] is one of the main chipset providers using both standards.

Furthermore, heavyweights have adopted BPL as part of their portfolio or strategy.

Alstom Grid [31] has partnered with a DS2 integrator to include BPL as a SCADA solution for medium voltage, integrating digital information together: telephony, video images, and grid data over 20 kV networks. Several deployments have been conducted by Alstom: in France, Brazil, India, etc.

In 2012, Siemens Financial Services Venture Capital has become shareholder of PPC [32] to strengthen their portfolio in BPL communications.

Alcatel-Lucent has been studying the opportunities, which might be realized by this technology, but up to now, it has not come out with a specific product or strategy.

Itron cooperates with PPC for integrating BPL in their modems. Other modem suppliers have also undertaken some first developments with BPL: Landis & Gyr [33], IUSA [34], ZPA [35] and Mikroelektronika [36]. In 2009, Elster [37] announced its collaboration with Defidev [38] for the first meter prototype with BPL and radio being integrated.

ETDE (groupe Bouygues) [39] is using BPL technologies for smart street lightning in various French cities.

Utility companies as well have chosen BPL for various segments: Iberdrola has decided to implement BPL on its medium voltage lines for gathering the AMR data, British Gas which invested in PPC company in 2012 to strengthen its presence in the European smart grid market will start a pilot project with BPL on various smart grid applications for new opportunities.

In addition, numerous utility companies all over the world are testing the technology for their own needs, in Europe (CEZ, EnBW, EDF, etc.), in Asia (Meralco, RESB, TPC, etc.) in Africa (CEET in Togo ONEE in Morocco, etc.), or in Latin America (Light, etc.).

The interest for BPL has been intensified due to the potentials of the technology for supporting new applications within Smart Grid.

IV. UNIQUE SELLING POINTS OF BPL FOR SMART GRID INFRASTRUCTURE

Implemented on the low and medium voltage networks, BPL allows a real time and permanent connectivity and thus, continuous monitoring of the entire electrical grid. The technology works without creating interferences with the energy distribution activities and allows energy utilities to achieve immediate operational gains, such as:

In the short term:
- Technical and non-technical losses on low and medium voltage are identified and restrained.
- Manual meter readings are no longer required, replaced by remote metering enabling immediate and reliable readings.
- Invoices are automatically generated, preventing any entry error and reducing invoicing delay.
- Low and medium voltage network and substations remote monitoring allows improved network operations (preventive and reactive maintenance, etc.) and optimised investment decisions.

In the medium term:
- Quality of service is improved thanks to error prevention, innovative offers and tariffs meet specific needs of clients.
- Gathering information in real time allows optimised network operation and highly reactive maintenance services.
- Customer service can offer a highly efficient support, improved with real-time network information transmitted through the Broadband Powerline infrastructure, and thus, enabled to answer to any incident or demand accurately.

In contrast to the generated benefits, Broadband Powerline induces limited additional operational costs, essentially related to the IP (Internet) management tools and the equipment maintenance. Further, the achieved commercial maturity of the technology has allowed manufacturers to develop a wide range of dedicated products for applications related to energy distribution.

The flexibility of BPL communications and their ability to complement with any other telecommunication technology enable the use of existing infrastructures such as optical fiber deployed on High (or even Medium) Voltage networks or wireless technologies, and thus, to interface with any usual management system already deployed. Therefore, the investments required by the deployments are clearly defined and controlled.

The energy provider is able to activate its infrastructure and to remotely manage each of its BPL devices, the meter and other connected devices (sensors, alarms, cameras, etc.) as soon as first infrastructures have been implemented. Thus, Return on Investment of Broadband Powerline businesses is immediately enabled and attractive and grows along with the increase of the Powerline coverage.

Broadband Powerline deployment might be divided in phases and cells, namely by group of LV equipments behind a transformer or a HV/MV line. The CAPEX can be spread across the first years of the project, in order to smooth out the required investments specific LV or MV groups. Such flexibility and incremental deployment permit previous analysis to identify the best potentials for operating savings.

The analysis of Broadband Powerline projects deployed until today has brought proof of the rapid profitability reached by the implementation of a Smart Grid network through this technology. This is being achieved through the incremental savings made available by Broadband Powerline, the control of the correlated investments and the limited operating costs. Coverage of the dense areas ensures a return on investment within a few years and easily balances out the lower profitability of more rural areas, to eventually reach a profitability worth up to five times initial investments.

These advantages are not the only benefits brought by Broadband Powerline, as this technology allows:

- A widening spectrum of possible applications thanks to the versatile nature of Broadband Powerline
- Possibility of upgrading the delivered speeds according to future needs and emerging new applications
- An efficient and simplified coverage of acentric, less dense areas

V. HIGHLIGHTS ON SOME REPRESENTATIVE BPL CASE STUDIES- RETURN ON EXPERIENCE

Representative projects implementing BPL have been set up within Smart Grid and show the viability and performance of BPL compared to other technologies.

The most quoted as an example for AMR is the Advanced Metering Management Pilot Project, illustrated in Figure 3, and conducted by CEZ which has enabled to analyse and compare various powerline solutions for Smart Grid technology's installation.

Alongside various narrowband PLC solutions, such as:

- Landys&Gyr using PLC modem from local PLC vendor Modemtec,
- ZPA Smart energy which is a local company using a first generation PLC solution / licence from Renesas [40],
- ADD Grup Moldova [41]
- Echelon
- as well as a few GPRS from ZPA and a few RF from EHM,

all being deployed at 28 000 meter locations and in almost 1,000 distribution substations, 5000 BPL meters have been deployed with Corinex' technology in 2 regions Vrchlabí and Pardubice in underground cables. Over 100 BPL couplers have been installed on medium voltage technology, as well as several dozen repeaters.
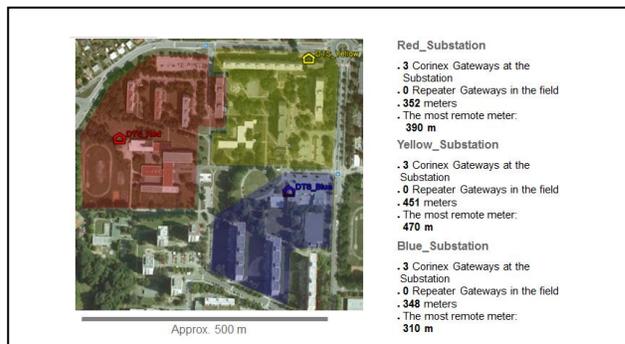
Figure 3. Pardubice BPL test with CEZ in Q1 2012 [17]

General issues appeared along the project deployment such as network optimization, LV nodes identification with high level crosstalks, to name only few of them. But on the whole, the project implementation was mastered and allows drawing significant conclusions about the various technologies' performance.

In the Pardubice BPL project, where no repeaters have been installed, a final average meter reading success rate of 99.705% was registered. Indeed, after the installation by CEZ (which has been done by the utility without assistance by Corinex), some 1.6% of the 1220 meters were not connecting first, but this was resolved quickly after reviewing.

Thus, BPL appeared as having superior connectivity and performance to any other PLC deployments.

Furthermore, CEZ states that the costs are comparable with narrow band solution in absolute terms, and hundreds times lower in price per Mbps.

Thus, though emerging as smart metering solution, BPL - among the numerous technologies tested or deployed for Smart Grid applications (Narrowband Powerline, Wireless, Radio, GPRS, etc.) - possess unique selling points to be pinpointed in the direct control of the energy distribution network, through a permanent and continuous connection (without concentrator for example), under truly controlled costs and allowing the possibility to increase data rates for future applications.

Another Smart Grid area is the utilization of BPL for MV lines, allowing SCADA applications, monitoring of the distribution substations and gathering the data flow from the meters. The appetence of various utilities and industrialists indicate a soon to come maturity for BPL. Many Utility companies are testing (EDF, EnBW, Stadtwerke Ratingen, Scottish Power, etc.); but already some have decided to implement this in a full deployment. Iberdrola in Spain is a good example as already 40% of the targeted MV substations are connected through BPL.

In France, one project has as well come to maturity with the SICAE de la Somme et du Cambraisis [42], which has deployed, jointly with Alstom Grid, innovative BPL technology for MV power grid applications such as telephony, video surveillance, remote meter reading, real-time data exchange with grid controls and active demand response management.

The BPL platform has been successfully deployed over 65 km of underground and overhead lines with approximately 50 communication points on grid nodes for its French rural municipalities.
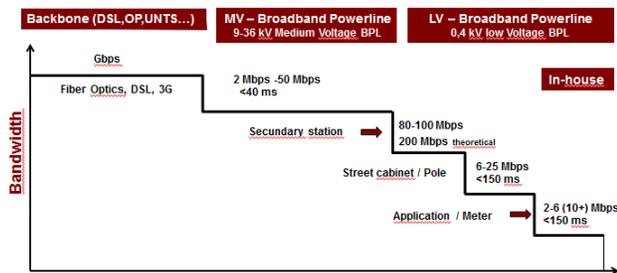
Figure 4.   General overview of BPL performance [9]

Unequalled performance on underground and overhead MV lines has been attained in this project, showing that BPL has pushed further its first limits related to distance and sufficient bandwidth, especially related to overhead MV lines.

VI.   CONCLUSION

On the facts gathered from numerous projects and as highlighted by Figure 4, Broadband Powerline, enabling permanent, reliable and fast connectivity onto the entire electrical grid, is therefore considered as a future proof technology, particularly relevant for the implementation of the Smart Grid. However, it is foreseen that hybrid architectures will best meet the specific objectives and priorities of each energy provider, where as Broadband PLC/BPL can and is expected to play a central role.

Some expect the European market potential for BPL in Smart Grids to reach €25bn by 2020 in the Low Voltage segment (metering, e-Mobility, Home Gateway,etc.), while the MV (mainly network automation) is forecasted at €3.5bn. [43].

Though BPL will have to face specific challenges such as imposing an interoperable standard, while using BPL, one should as well consider the use of the excess of capacities which could be used to create high data rates Internet connections and make them available for various applications.

In this respect, one could think of Internet and telecom services, as well as the usage of BPL for smart street lightning networks and the set up of smart cities and buildings.

The deployment of BPL solutions has allowed companies such as Bouygues, ETDE to offer additional services basing on a platform developped first to monitor the street lightning system. Thus, excess capacities enable video-surveillance, traffic monitoring, and the creation of hot spot WIFI as seen in some French cities in the Parisian region.

VII.   REFERENCES

[1] S. Galli, A. Scaglione, and Z. Wang "For the Grid and Through the Grid: The Role of Power Line Communications in the Smart Grid," Proc. IEEE Symp. IEEE Press, June 2011, pp. 8-10.

[2] R. W. Chang, "Orthogonal frequency division multiplexing," U.S. Patent 3 488 445, 1970.

[3] Open Public Extended Network Metering. [Online]. Available: www.openmeter.com, 11/02/2013.

[4] Prime Alliance Homepage, www.prime-alliance.org, 11/02/2013.

[5] G3-PLC Homepage, www.g3-plc.com, 11/02/2013.

[6] Meters & More Homepage, www.metersandmore.com, 11/02/2013.

[7] Echelon Homepage, www.echelon.com/applications/smart-metering/, 11/02/2013.

[8] PowerPlusnet [Online]. "Siemens to Open Up Smart Metering Communications for Meters Produced by Other Manufacturers," Available:www.powerpulse.net, 21/09/2010.

[9] bmp TC research 2012, www.bmp-tc.com/res1.htm, 11/02/2013.

[10] B. Watson, "FSK: Signals and Demodulation," Watkins-Johnson Company, Vol. 7, No. 5 September/October 1980.

[11] T. Schaub, "Spread Frequency Shift Keying," IEEE Transactions on Communications, Vol. 42, Feb/Mar/Apr 1994, pp.1056 – 1064.

[12] J.A Ávila Rodríguez, "Binary Phase Shift Keying Modulation (BPSK)," University FAF Munich, Germany, 2011.

[13] G. Kolumbán, B. Vizvári,W. Schwarz, and A. Abel, "Differential chaos shift keying: A robust coding for chaos communication," Proc. Int. Workshop Nonlinear Dynamics of Electronic Systems, pp. 87–92, June 1996.

[14] Echelon Homepage, www.echelon.com, 11/02/2013.

[15] Yitran Homepage, www.yitran.com, 11/02/2013.

[16] bmp TC, SG Paris, Smart Metering session, www.sgparis.fr/, Paris, 06/2012.

[17] Mereni, Muller, and Mamula, CEZ, [Online]. www.cez.cz/edee/content/file/o-spolecnosti/dcerine-spolecnosti/cez-reni/konference2012/CEZ-MereniMullerMamulaEN.ppt, 18/04/2012.

[18] International Broadband Electric Communications, www.telecompetitor.com/ibec-shutdown-deals-latest-blow-to-bpl/, 11/02/2013.

[19] Design of Systems on Silicon, Marvell Homepage, www.marvell.com, 11/02/2013.

[20] Boston Business Journal, [Online]. Available: http://www.bizjournals.com/boston/stories/2004/02/16/daily53.html, 20/02/2004.

[21] Homeplug Homepage, www.homeplug.org, 11/02/2013.

[22] HD-PLC Homepage, www.hd-plc.org, 11/02/2013.

[23] ITU Gh.n Standard Homepage, www.itu.int, 11/02/2013.

[24] Broadcom Homepage, www.broadcom.com, 11/02/2013.

[25] Mstar Homepage, www.mstarsemi.com, 11/02/2013.

[26] Qualcomm Atheros Homepage, www.qca.qualcomm.com, 11/02/2013.

[27] Lantiq Honepage, www.lantiq.com , 11/02/2013.

[28] Marvell Homepage, www.marvell.com, 11/02/2013.

[29] Metanoia Homepage, www.metanoia.com.tw, 11/02/2013.

[30] Sigma Designs Homepage, www.sigmadesigns.com, 11/02/2013.

[31] Alstom Grid Homnepage, www.alstom.com/grid, 11/02/2013

[32] Power Plus Communication Homepage, www.ppc-ag.de, 11/02/2013.

[33] Landis & Gyr Homepage, www.landisgyr.com, 11/02/2013.

[34] IUSA Homepage, www.grupo-iusa.com, 11/02/2013.

[35] ZPA Homepage, www.zpa.cz, 11/02/2013.

[36] Mikroelektronika Homepage, www.mikroe.com, 11/02/2013.

[37] Elster Homepage, www.elstermetering.com, 11/02/2013.

[38] Defidev Homnepage, www.defidev.com, 11/02/2013.

[39] ETDE Homepage, www.etde.fr, 11/02/2013.

[40] Renesas Homepage, www.renesas.eu, 11/02/2013.

[41] ADD Grup Homepage, www.addgrup.com, 11/02/2013.

[42] SICAE de la Somme Homepage, www.sicaesomme.fr, 11/02/2013.

[43] I. Schönberg, PPC "Next Generation Smart Grid"[Online]. Available: www.ppc-ag.de, 23/11/2010.

Sensors for Smart Grids

Smart Grids Technologies

Francisc Zavoda

Electric Equipment Expertise (EEE)
Research Institute of HQ (IREQ)
Varennes, Canada
e-mail: zavoda.francisc@ireq.ca

Chris Yakymyshyn

FieldMetrics, Inc
Seminole, Florida, USA
e-mail: yakymyshyn@fieldmetricsinc.com

Abstract—**The Smart Grid will enable the customers to actively decide their energy choices and to accommodate their generation and storage options. The same grid will provide higher reliability and consistent power quality, which are required by our digital economy, by optimizing the use of assets and the grid operation. The future power grids will come into reality by enabling intelligent communication across sensing, measurement, and control layers that are embedded into the existing power systems. This paper discusses the results from tests performed on two sensors, a combined voltage and current sensor and an optical voltage sensor.**

Keywords-Smart Grid; Distribution Automation; Volt & VAR Control; Fault Location; Power Quality; monitoring; sensor; accuracy, magnitude, phase angle.

I. INTRODUCTION

The Smart Grid will enable the customers to actively decide their energy choices and to accommodate their generation and storage options. The same grid will provide higher reliability and consistent power quality, which are required by our digital economy, by optimizing the use of assets and the grid operation. The future power grids will come into reality by enabling intelligent communication across sensing, measurement, and control layers that are embedded into the existing power systems.

Smart distribution systems will increasingly be dependent on monitoring of the system conditions for both real time management and improved maintenance strategies [1]. Integrated distribution monitoring systems will require various types of sensors and transducers to help understand system conditions and respond to disturbances affecting it. Voltage and current monitoring will be critical for a large number of applications [2]. Specific sensor requirements will include support for:

- Advanced voltage control functions for voltage optimization, voltage reduction, Volt/VAR control,
- Load current monitoring for reconfiguration strategies, asset management and fault location applications,
- Protection and reconfiguration applications (fault current monitoring, coordination of protection characteristics, etc.),

- Waveform acquisition for fault location and other diagnostic applications including incipient fault detection and location,
- Harmonic monitoring for power quality assessments and other diagnostic applications.

Nowadays, several manufacturers offer performing sensors to improve network performance.

The paper is divided into several sections. Section II discusses the need for new sensors. Sections III and IV describe optical sensor techniques. Section V describes the details of the sensors that were tested. Section VI describes the test methods used. Section VII presents the test results. Section VIII provides conclusions.

II. SMART GRID AND SMART DISTRIBUTION

Today, advances and falling prices of the communication and control technologies, allow their embedding into distribution grid and use for grid monitoring and remote control of major distribution equipment, (switches, capacitors banks, reclosers and voltage regulators), thus replacing the old Distribution Automation (DA), consisting mainly in automatic operation of medium voltage (MV) equipment.

A better knowledge of what is happening in the power system is crucial for an improved distribution grid management. New technologies such as sensors, IEDs, software and telecommunication can provide data required by smart distribution applications to improve the power system efficiency through utility's business needs. It is a necessary feedback loop to improve the distribution system performance.

The new way of thinking and operating the Smart Grid created the premises for an integrated distribution monitoring system, based on sensors and Intelligent Electronic Devices (IED) capable of providing accurate and reliable data required by new smart distributed applications.

III. SENSORS

For Smart Grid and Smart Distribution (SD) applications, vendors propose new "Smart Sensors", which can work autonomously, when equipped with a communication interface, or in combination with IEDs. Among them,

voltage and current optical sensors offer a number of major advantages, but the speed of their integration into the grid depends on price reduction.

IV. OPTICAL SENSORS

Conventional methods to monitor overhead electrical lines use iron core instrument transformers, which are not easy to mount because of their significant size and weight, and can present safety constraints to personnel as well as to the interconnected equipment. The physical hazards associated with conventional iron-core devices, when coupled with the complexity of installation, have typically precluded their use within medium-voltage distribution systems.

The Optical Sensors [3] represent a significant improvement over conventional iron-core instrument transformers or Rogowski coils. These sensor devices are positioned for use in switchgear, power distribution, electronics and medium voltage applications (< 37 kV) where the conventional current and voltage transformers may be undesirable due to space and weight concerns.

A. Optical Current Sensor (OCS)

The technology used by optical current sensors is based on the Faraday effect. When a linearly polarized light travels through a transparent material that is exposed to a magnetic field, its plane of polarization rotates (see Figure 1). In sensor systems that exploit the Faraday Effect, a sensor assembly is placed into a magnetic field. By monitoring the rotation of the incident polarization state, a direct measurement of the magnetic field intensity, and consequently the associated current, can be inferred.
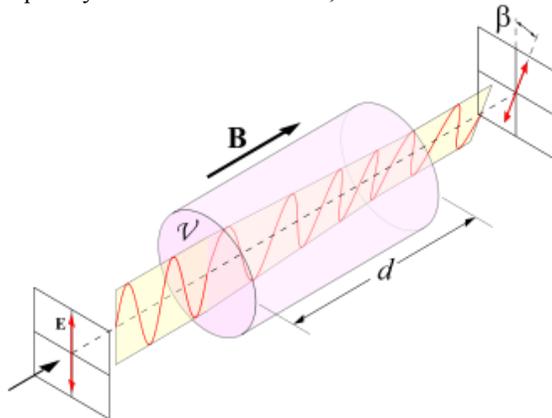
Figure 1. Faraday effect (source Wikipedia)

Currently, there are several types of optical current sensors on the market, which use radically different technologies.

Some of them use either bulk optical crystals or long lengths of optical fiber, coiled around a current-carrying conductor. When an optical path completely encircles a conductor, as is the case of coiled optical fiber sensors, a numerical integration can be performed about the optical path that directly relates the Faraday rotation to the current

flowing through that optical path. The measurement sensitivity is dependent on the number of optical fiber turns around the conductor being monitored.

The sensors utilizing bulk glass take advantage of the Faraday Effect exhibited by the bulk glasses [4]. The sensor can be fabricated from materials, which are more sensitive to the influencing magnetic field than those used in coiled optical fiber sensors.

Fiber optic magneto-optic field sensors [5] use ferromagnetic materials more sensitive than materials used in simple fiber optic cable or bulk-optic crystals. The straightforward result is that the sensor requires much smaller Faraday rotator to measure a given magnetic field strength, and offers versatility in physical size and the range of measurement properties.

B. Optical Voltage Sensor (OVS)

The OVS technology [6] is based on the Pockels effect. It is used to make Pockels cells, which are voltage-controlled wave plates (see Figure 2).
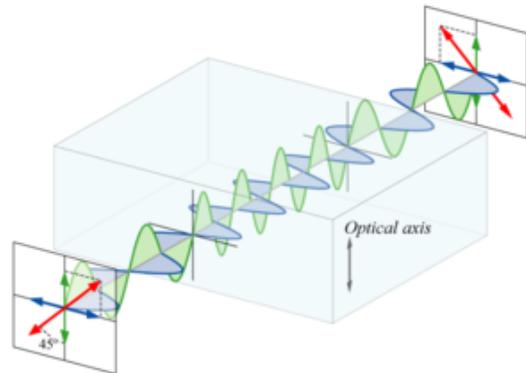
Figure 2. Half wave plate (source Wikipedia)

When a steady electric field is applied to certain electro-optic materials, their refractive indices change, roughly in proportion to the strength of the applied field. The Pockels effect occurs only in noncentrosymmetric optical crystals, such as lithium niobate or gallium arsenide. In combination with a polarizer, Pockels cells can be used as an extremely fast shutter that can respond in nanoseconds. In sensor systems that exploit the Pockels effect, a sensor assembly is placed into an electric field. When polarized light is projected through the Pockels cell that is in this voltage field, the state of polarization of the light traveling in the crystal is rotated. From the measured polarized light rotation the voltage across the crystal can be determined.

V. SENSORS UNDER TEST

Both sensors used in this study, the MetPod [7] and the Optical Voltage Transducer (OVT), were rated for 25kV class operation. Prior to frequency response testing, the prototype devices passed dielectric type testing [8,9]. This included a dry AC withstand test at 50 kV for 1 minute, followed by BIL testing of 15 full wave impulses (1.2 x 50 microseconds) at 150 kV peak, along with two or more chop wave impulses (1.2 x 3 microseconds) at 175 kV peak. Both

polarities were applied in the BIL tests. Thermal cycling was performed in an insulating temperature chamber while the device was energized with high voltage, or simultaneously with high voltage and moderate current. The thermal ramp rate was kept at <30 ºC/hour to allow the unit under test to thermally equilibrate with the ambient temperature in the test chamber.

1) MetPod/Combined Voltage & Current Transducer (CVCT)

The MetPod (see Figure 3) is a self-powered sensor solution that combines a voltage sensor, a current sensor and a power supply to provide output signals suitable for interfacing with IEDs. As shown in Figure 3, the MetPod uses a hollow core composite silicone polymer insulator to support a high voltage conductor. The opposite end is tied to neutral. An optical fiber carries digitized current and voltage data from the MetPod to an interface module, where it is converted back to either low voltage signals or 120V and 1A signals suitable for interfacing with a variety of power quality measurement devices.

Figure 3. MetPod sensor

The current sensor approximates Ampere's law by integrating the spatial magnetic field around a closed path that encircles the high voltage conductor using a plurality of point sensors. This approach generates a sensor output that is independent of: conductor position within the sensor window; conductor size; current density distribution within the conductor; or magnetic field perturbations caused by adjacent phases or nearby ferrous objects. The use of an air core permits an open slot on one side of the current sensor that allows installation without breaking an existing high or medium voltage conductor.

The voltage sensor uses a precision impedance divider collinearly located near a second impedance string that supplies AC current to an integrated power supply. Because

of the relatively low impedance of the voltage divider (<10 Megohms at 60 Hz), and the electrostatic shield provided by the second impedance string, the voltage measurement is insensitive to external electric field perturbations caused by adjacent phases, pollution or ice on the insulator surface, etc. The hollow insulator is pressurized to 15 psi with dry nitrogen gas.

At the high voltage end of the MetPod, two 24-bit A/D converters digitize and interleave the sensor signals, while providing a signal bandwidth of 20 kHz. The single digital stream is coupled onto a conventional multimode fiber and transferred out of the neutral end of the MetPod. The analog signals that are reconstructed at the interface module have a fixed group delay of approximately 650 microseconds relative to the high voltage conductor. Due to the interleaved sampling, the relative phase error between the voltage and current signals is <0.2 degrees at 60 Hz.

As declared in the sensor's technical specification, the voltage and current measurements meet 0.3% accuracy or better over a temperature range of -20 ºC to +60 ºC. The current sensor's rated current range extends from 200 A to 10,000 A. In all cases, the current sensor is not damaged by currents with crest values as high as 20,000 A.

2) Optical Voltage Transducer (OVT)

Figure 4. OVT sensor

The OVT (see Figure 4) has the same physical footprint as the MetPod. A collimated light source located at the high voltage end is circularly polarized before passing through an aperture. The optical beam then passes through one or more optical crystals located along the length of the hollow core insulator. Each electro-optic crystal modulates the state of polarization of the light in direct proportion to the voltage drop across that crystal. Provided the modulation depth is kept small, the final state of polarization is a linear sum of

the modulation imparted by each crystal along the vertical optical beam path. This provides an excellent approximation to the line integral of the electric field between high voltage and neutral, which is the definition of potential difference. The polarization-modulated optical beam is sampled by two photodiodes fed by the orthogonal outputs of a polarizing beamsplitter. The two detected signals undergo separate normalization and temperature compensation prior to being differenced to create the final output signal. Differencing removes common-mode noise that may arise from a variety of sources. The analog signal is digitized and transferred onto standard multimode tele-communications optical fiber. The resulting sensor system operates within a factor of 2 of the shot noise limit, with a signal bandwidth of >20 kHz.

A co-located impedance string feeds the integrated power supply to energize the light source at the high voltage end, and a second power supply at the neutral end to power the optical receiver electronics and digitizer.

The OVT provides 0.3% accuracy from 80% - 120% of rated line voltage, over a temperature range of -20 ºC to +60 ºC. The system bandwidth is >20 kHz.

VI. EXPERIMENTAL SETUP

The measurement of the harmonic response of the MetPod and the OVT was performed using two separate setups.

A. Voltage Measurement Setup

As shown in Figure 5, the MetPod voltage response and the OVT response were evaluated by applying a harmonic tone superimposed on a 60 Hz carrier signal. One arbitrary waveform generator (HP 33120A), operating at 60 Hz, was used to phase-lock a second HP 33120A operating at a harmonic of 60 Hz. The amplitude of each generator was independently adjustable. The two sinusoids were combined and amplified by a 2400W power amplifier, which drove the secondary winding of a 300:1 Potential Transformer. The high voltage signal was applied to the device under test (DUT), and monitored using a reference divider having a bandwidth of >100 kHz. The low voltage output of the interface module was amplified by a precision amplifier with a gain of G=20. An identical amplifier was used to amplify the output of the reference divider.
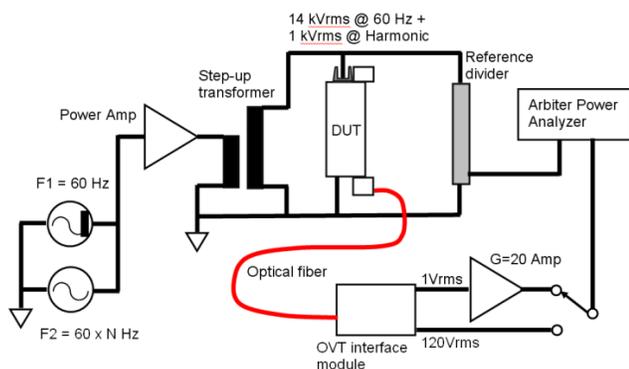
Figure 5. Schematic diagram of test setup used for harmonic response of the MetPod Voltage sensor and the OVT

The outputs of the interface module and the reference divider were fed into a power quality analyzer (Arbiter Systems 931A) with a maximum frequency of 3 kHz, or the 50th harmonic. The reported amplitude and phase are relative to the fundamental component present. Data were recorded by adjusting the harmonic amplitude to be 2% - 10% of the fundamental amplitude. The 60 Hz fundamental was kept constant at 14 kV. The amplitude error was typically +/- 0.05%, while the phase error was typically +/- 0.2 degrees.

B. Current Measurement Setup

The current response of the MetPod was measured by providing external 8VDC power to the MetPod through an auxiliary power input port that was installed for these tests. This allows the current sensor to be functional without having to apply high voltage to activate the power supply. A sine wave from an arbitrary waveform generator was amplified by a 2400W power amplifier, which then drove the primary winding of a 122:1 turns ratio, 22 kVA transformer. The transformer secondary was connected to a short length of stranded copper cable that formed 5 turns through the MetPod current sensor slot. An open-loop split-core sensor (LEM series HBT-200, 50 A/V, 50 kHz bandwidth) was used as the reference. The optical fiber output of the MetPod was coupled to the interface module to provide a low voltage reconstruction of 5 times the current in the conductor. The multiple turns were used to maintain adequate signal to noise ratio while operating the test setup at relatively low conductor currents (at high frequencies, the conductor current was <10 A).

The MetPod and reference signals were monitored with an HP 3478A true RMS multimeter, and phase angle was measured with a Krohn-Hite 6500 phase meter. The magnitude error with this test setup was typically +/- 0.5%, while the phase angle error was typically +/- 0.2 degrees. The phase error was extracted from the raw phase measurement by subtracting a baseline phase delay caused by the group delay of the optical fiber link. The delay resulted in a phase angle offset in degrees that is given by -0.23f, where f is the harmonic frequency in Hz. For example, the phase offset is -13.8 degrees at 60 Hz.

VII. TEST RESULTS

Results from the tests performed on above sensors are presented below. In all cases, the reported magnitude error or ratio is a relative to the magnitude of the measured harmonic. Since the harmonic signal amplitude in many cases is only a few percent of the fundamental signal, the reported errors are <0.1% of the total applied signal (fundamental + harmonic).

A. Voltage Measurement of OVT

The results for the OVT are shown in Figure 6 (amplitude) and Figure 7 (phase angle).
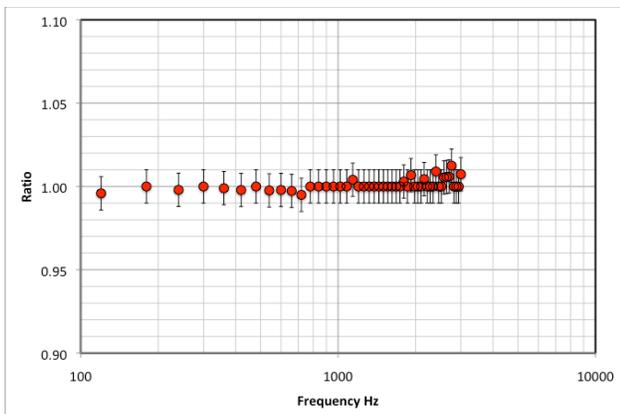
Figure 6. Amplitude response of OVT voltage sensor versus frequency

The OVT voltage sensor maintains 1% or better accuracy over the entire range of frequencies tested, up to 3 kHz or the 50th harmonic. The phase angle does not exceed 10 degrees lag at the 50th harmonic. The phase response can be modeled as a second-order low-pass system response with a bandwidth of 18 kHz.
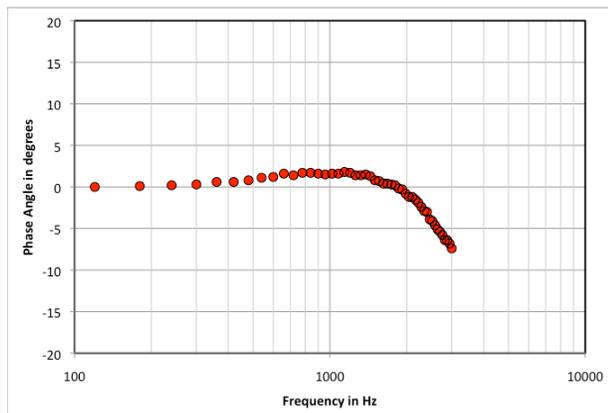
Figure 7. Phase response of OVT voltage sensor versus frequency
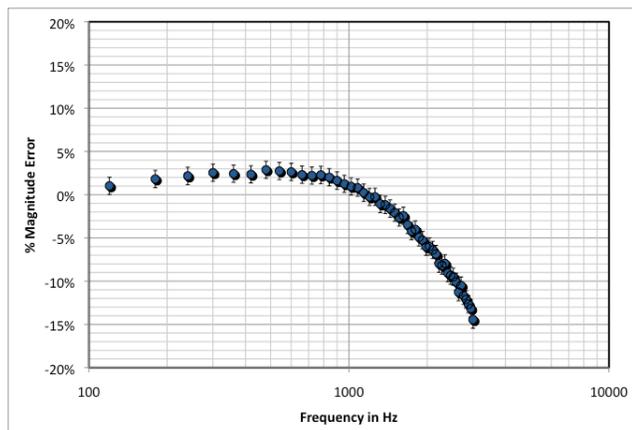
Figure 8. Amplitude response of MetPod voltage sensor versus frequency

B. Voltage Measurement of MetPod

The magnitude response versus frequency of the MetPod's voltage sensor using the 120 Vrms output from the interface module is shown in Figure 8, and the phase angle response is shown in Figure 9. The MetPod voltage sensor maintains better than 5% accuracy up to 1.8 kHz or the 30th harmonic of 60 Hz. The phase error remains less than 30 degrees at the 30th harmonic. The response indicates an overall 3 dB voltage sensor bandwidth of approximately 3.2 kHz. The bandwidth is currently limited by the interface module, and could be increased to the bandwidth of the digital optical link, or 20 kHz, if desired.
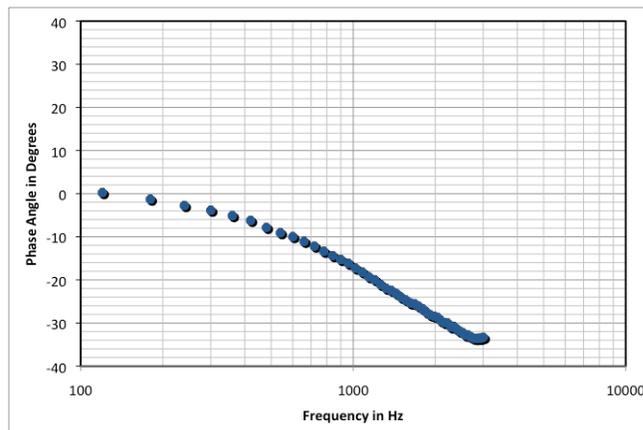
Figure 9. Phase response of MetPod voltage sensor versus frequency

C. Current Measurement of MetPod

The magnitude response versus frequency of the MetPod's current sensor is shown in Figure 10, and the phase angle in Figure 11.
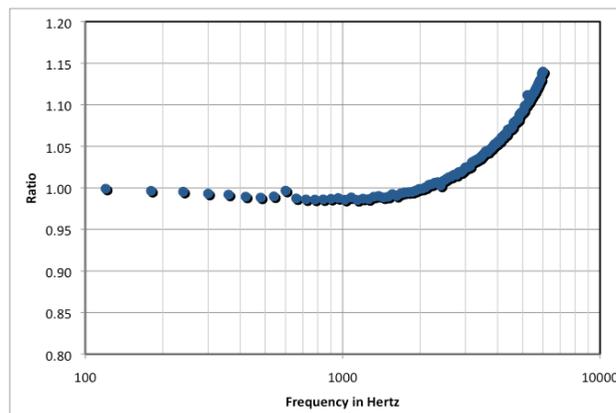
Figure 10. Amplitude response of MetPod current sensor versus frequency.

The magnitude accuracy of the current sensor remains within 5% at 4 kHz, which is the 66th harmonic of 60 Hz. The phase error remains within +/- 5 degrees up to 6 kHz, or the 100th harmonic.
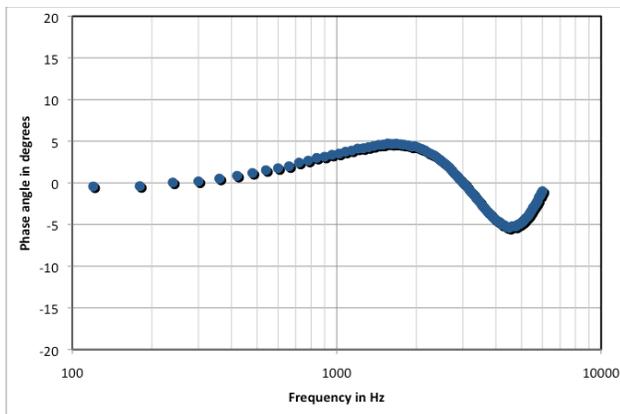
Figure 11. Phase angle versus frequency for MetPod current sensor

The MetPod current sensor frequency response was also evaluated by comparing the response with a reference CT (Ritz extended range, 0.15% from 1% - 150% I_{nom}, 1000:5) using the Arbiter 931A power quality analyzer. In this case, the applied voltage was purely fundamental 60 Hz at 14 kV, whereas the current was 100% at the harmonic frequency being tested. The response is shown in Figure 12.
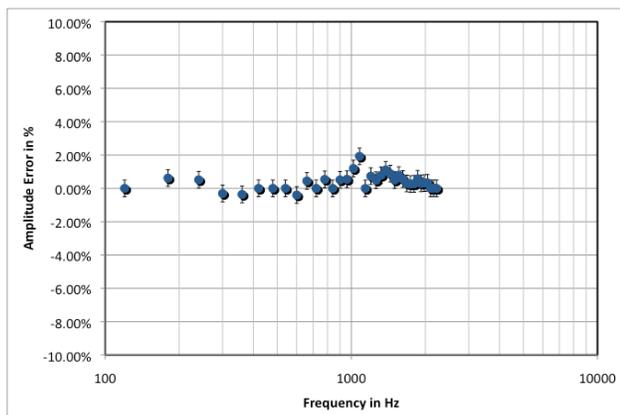
Figure 12. MetPod current sensor amplitude response versus harmonic frequency

The frequency response remains within 2% out to 2100 Hz, the highest frequency tested in this configuration.

VIII. CONCLUSIONS

Optical sensors are proving their value, especially in applications where accurate measurement over wider dynamic range is required, or where ability to retrofit, and improved safety are of main concern. They are suited for the advanced functionality of leading-edge protective relays and meters and for compatibility with digital communications in modern substations. Signal processing inside the current and voltage electronics of the sensor is inherently digital in nature and is accessible in a format consistent with IEC standards such as IEC 61850-9-1.

The intrinsic insulating properties of the non-conductive optical fiber allow the optical transducer to be used on 15 kV systems as well as 460 V systems. Because the optical sensors monitor the external electric or magnetic fields, they do not burden the generating system, resulting in savings by eliminating the electrical losses and production of heat within the monitoring system. The use of optical sensors brings several significant advantages such as:

- Reduced substation costs (combined VTs and CTs are suitable for both metering and protection).
- Reduced installation and commissioning costs (the sensor's lighter weight has a direct impact on foundation size, ease of equipment installation, and transportation and handling).
- Improved operating performance:
- Accuracy over a wide dynamic range,
- Intrinsically safe, no ferroresonance or open-circuit secondary concerns.
- Reduced maintenance and overhaul costs including reduced end of life disposal costs (no oil or SF6).

REFERENCES

[1] F. Zavoda, "Advanced Distribution Automation (ADA) Applications and Power Quality in Smart Grids," CICED 2010, Nanjing, China, September 2010, pp. 1-7.

[2] B. K. Duncan and B. G. Bailey, "Protection, Metering, Monitoring and Control of Medium Voltage Power Systems," IEEE Transactions on Power Delivery, Vol. 40, No. 1, January/February, 2004, pp. 33-40.

[3] Handbook for Electricity Metering, 10th Edition, 2002 (Edison Electric Institute), pp. 381 – 391.

[4] T. W. Cease and P. Johnston , "A Magneto-Optic Current Transducer," IEEE Transactions on Power Delivery, Vol. 5, April 1990, pp. 548 – 555.

[5] J. Blake, P. Tantaswadi, and R. T. de Carvalho, "In-line Sagnac Interferometric Current Sensor," IEEE Transactions on Power Delivery, Vol. 11, January 1996, pp. 116 – 121.

[6] C. P. Yakymyshyn, S. Weikel, P. Johnston, and M. Brubaker, "362 kV Optical Voltage Transducer- Final Report," ESEERCO Research Report EP91-04, December, 1998.

[7] M. A. Brubaker, P. J. Hamilton, and C. P. Yakymyshyn, , "Clamp-On Untethered Power Interface Pod", U. S. Department of Energy, SBIR Grant Number DE-FG03-01ER832228/A002, Phase II Final Report, August, 2004.

[8] IEEE Standard Requirements for Instrument Transformers, IEEE Std. C57.13-1993.

[9] IEC 60044-7, Instrument Transformers, Part 7: Electronic Voltage Transformers, 1992-12.

Grid Monitoring: Bounds on Performances of Sensor Placement Algorithms

Muhammad Uddin
Anthony Kuh
Aleksandar Kavcic
Dept. of Electrical Engineering
University of Hawaii
Honolulu, HI, USA
Email: {uddin, kuh, kavcic}@hawaii.edu

Toshihisa Tanaka
Dept. of Electrical and
Electronic Engineering
Tokyo University of Agriculture
and Technology
Tokyo, Japan
Email: tanakat@cc.tuat.ac.jp

Danilo P. Mandic
Dept. of Electrical and
Electronic Engineering
Imperial College
London, U.K.
Email: d.mandic@imperial.ac.uk

Abstract—The objective of this paper is to find numerical bounds on the performances of algorithms for the placements of phasor measurement units (PMUs) in the power grid. Given noisy measurements and knowledge of the state correlation matrix, we use a linear minimum mean squared error estimator as the state estimator to formulate the PMU placement problem as an integer programming problem. Finding the optimal placements of a fixed number of PMUs in a large network is prohibitively complex, forcing us to look for suboptimal algorithms and bounds on the optimal performance. In this paper, we present a sequence of nested bounds using matrix pencils and generalized eigenvalues, that upper bound the optimal performance. Finally, we numerically compare the performances of the suboptimal solutions with the upper bounds using the IEEE 30- and 57-bus test systems, showing that the proposed bounds provide a valid basis for determining the quality of different suboptimal solutions.

Keywords–phasor measurement units; matrix pencils; generalized eigenvalues.

I. Introduction

State estimation (SE) is a key function in modern energy management systems, where various crucial control tasks depend on the accurate snapshots of the system state [1]. Conventional state estimators rely on the redundant measurements captured by supervisory control and data acquisition (SCADA) systems [1], which can only take non-synchronized measurements. These measurements are too infrequent to capture the dynamics of the power grid [1]. With the advent of phasor technology, time synchronized measurements can be obtained using phasor measurement units (PMUs) [2]. These devices take advantage of the global positioning system (GPS) technology to provide time-stamped measurements of the bus voltage magnitudes and phase angles [2].

Traditional SE using SCADA measurements is nonlinear, and is solved using iterative algorithms [3]. The PMUs, on the other hand, can directly measure the states at the PMU-installed buses, and the states of all the connected buses (if enough channels are available). In fact, given the high measurement precision and reliability of the PMUs, we can consider the PMU measurements to be low-noise

refinements of certain states (exactly those states that are measured by the PMUs) [2]. Since the PMUs can refine only a small subset of all state estimates, a common task is to refine the remaining state estimates (corresponding to the buses not carrying PMUs) using the sparse PMU measurements.

To measure all the state variables, the PMUs need to be installed at around one third of all the buses [4]. Since this goal is unlikely to be achieved in the near future, researchers look for the best solutions to deploy PMUs at a smaller subset of the buses, such that the state estimation error is minimized.

In this paper, we consider the optimization problem where we have n bus locations (where we can deploy PMUs) and m PMUs to place ($m << n$). We formulate the optimization problem to minimize the mean squared estimation error. Finding the optimal solution for the PMU placement problem is very difficult. In fact, it has been shown that the problem is NP-complete [4]. This means that there is no known efficient method to solve this problem with computational complexity that is polynomial in n. For this reason, heuristic approaches (e.g., greedy algorithm [5], gradient projection algorithm [6] etc.) are typically applied to search for good suboptimal solutions. But the question is, how can we guarantee that a heuristic solution is close to the optimal one, when we have no computationally feasible method of computing the optimal solution? The only way to guarantee the quality of a heuristic solution is to compare it to a provable and computationally feasible performance bound. However, for the PMU placement problem, no tight bounds are available either. Hence, in this paper, we propose upper bounds on the optimal solution that allow us to bound the difference between optimal and suboptimal solutions. Thereby, we significantly extend our prior work in [7] by presenting nested upper bounds with complete analysis of the bounds using matrix pencils and their generalized eigenvalues.

Related Work on PMU Placement: There has been much work done on finding reasonable suboptimal solutions to the PMU placement problem. In [5], the PMU placement

problem is formulated as an optimal experimental design problem. A greedy approach is proposed to solve this problem suboptimally and a lower bound is obtained on the performance of the greedy solution when the objective function is submodular. An estimation-theoretic approach is proposed in [6]; after posing the optimization problem as a linear regression problem, a convex relaxation is developed to find a suboptimal solution. Other research has considered PMU placement for system observability and measurement redundancy [8][9]. Dua et al. [9] considered the phasing of PMU deployment in an integer linear programming framework, while [8] proposed a binary particle swarm optimization based algorithm. In [10], the authors present a unified description of different algorithms proposed to solve the PMU placement problem. However, for a large network, in the absence of the optimal solution there is no benchmark to compare the performances of these suboptimal algorithms.

Paper Organization: The paper is organized as follows: Section II gives the measurement model used for our analysis. The optimization problem is formulated in Section III. We review two suboptimal algorithms in Section IV. In Section V, a family of upper bound on the optimal solution is found and the results are verified numerically using IEEE 30- and 57-bus test systems in Section VI. Section VII summarizes the results of this paper and suggests directions of future work.

Notation: Upper and lower case letters denote random variables and their realizations, respectively; underlined letters stand for vectors; boldface upper case letters denote matrices, and $\mathbf{I}_n$ denotes the $n \times n$ identity matrix; $\langle \mathbf{A}, \mathbf{B} \rangle$ denotes a matrix pencil formed by matrices $\mathbf{A}$ and $\mathbf{B}$; $(\cdot)^T$ and $\mathbf{E}(\cdot)$ stand for transposition and expectation, respectively.

II. MODEL

We consider voltage magnitudes and phase angles as state variables that are initially estimated using nonlinear SE from SCADA data, and then further refined using sparse PMU measurements. For introduction and justification of this approach, see [5]. We further explain this scenario below.

Assume there are n_b buses. Let V_k and Δ_k denote the voltage magnitude and angle of the kth bus, $k = 1, \ldots, n_b$. Let $\underline{V} = [V_1, V_2, \ldots, V_{n_b}]^T$ be the state vector representing the bus voltage magnitudes and $\underline{\Delta} = [\Delta_1, \Delta_2, \ldots, \Delta_{n_b}]^T$ be the state vector representing the corresponding phase angles. To make the state estimation more efficient in terms of storage and computational costs, we assume the voltage magnitudes and phases to be statistically independent random vectors [1][11]. We further assume that all the PMUs are identical and take statistically independent voltage magnitude and phase measurements with variances σ_v^2 and σ_δ^2, respectively.

Let m_v and m_δ be the number of PMU voltage magnitude measurements and the number of PMU phase angle mea-

surements, respectively, where $m_v \leq n_b$ and $m_\delta \leq n_b$. Let $\underline{Z}_v \in \mathbb{R}^{m_v}$ and $\underline{Z}_\delta \in \mathbb{R}^{m_\delta}$ be the PMU voltage magnitude measurement vector and PMU phase angle measurement vector, respectively. Then the PMU measurement model is:

$$\underline{Z}_v = \mathbf{C}_v(\underline{V} + \sigma_v \underline{N}_v), \tag{1}$$

$$\underline{Z}_\delta = \mathbf{C}_\delta(\underline{\Delta} + \sigma_\delta \underline{N}_\delta), \tag{2}$$

where $\underline{N}_v$ and $\underline{N}_\delta$ are random noise vectors with mean zero and covariance matrix $\mathbf{I}_{n_b}$. We assume that the random noise vectors $\underline{N}_v$ and $\underline{N}_\delta$ are statistically independent of the state vectors. $\mathbf{C}_v$ and $\mathbf{C}_\delta$ are matrices that represent the positions of the PMU placements (see Example 1). These are binary matrices with orthonormal rows, where each row has one '1'. The positions of ones in the matrix $\mathbf{C}_v$ and $\mathbf{C}_\delta$ denote the position of the sensors. In order to provide a reference point for the phase angle measurements, we assume that a PMU is always placed at the swing bus [2].

Example 1. *Figure 1 shows a* 4-*bus system [3] and matrices* $\mathbf{C}_v$ *and* $\mathbf{C}_\delta$ *when two PMUs are placed on buses* 1 *and* 3.

$$\mathbf{C}_v = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{C}_\delta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

● PMU
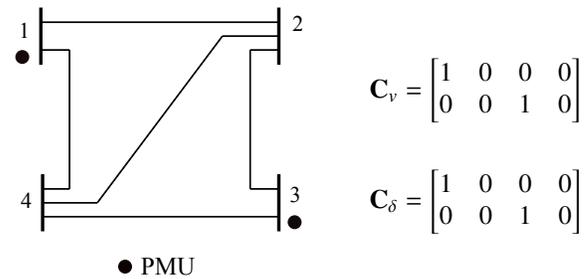
Figure 1.   PMU placement in a 4 bus system

Equations (1) and (2) represent separate models for PMU voltage magnitude and PMU phase angle measurements. We can combine (1) and (2) into a single model equation as

$$\begin{bmatrix} \underline{Z}_v \\ \underline{Z}_\delta \end{bmatrix} = \begin{bmatrix} \mathbf{C}_v & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_\delta \end{bmatrix} \left(\begin{bmatrix} \underline{V} \\ \underline{\Delta} \end{bmatrix} + \begin{bmatrix} \sigma_v \underline{N}_v \\ \sigma_\delta \underline{N}_\delta \end{bmatrix} \right). \tag{3}$$

Next, we argue that we can capture the natures of models (1)-(3) using a single state vector $\underline{X} \in \mathbb{R}^n$ and a single measurement vector $\underline{Z} \in \mathbb{R}^m$ ($m \leq n$) as

$$\underline{Z} = \mathbf{C}(\underline{X} + \sigma \underline{N}), \tag{4}$$

where $\underline{N}$ and $\underline{X}$ are statistically independent zero-mean random vectors with covariance matrices $\mathbf{I}_n$ and $\Sigma_{\underline{X}}$, respectively. The following two examples illustrate this concept.

Example 2. *Under the following transformations,* (1) *and* (4) *are equivalent.*

$$\begin{aligned} \underline{X} &= \underline{V} - \mathbf{E}(\underline{V}), & \sigma &= \sigma_v, \\ \underline{Z} &= \underline{Z}_v - \mathbf{E}(\underline{Z}_v), & n &= n_b, \\ \underline{N} &= \underline{N}_v, & m &= m_v, \\ \mathbf{C} &= \mathbf{C}_v. \end{aligned}$$

Example 3. *Under the following transformations,* (3) *and*

(4) *are equivalent.*

$$\underline{X} = \begin{bmatrix} V - \mathbf{E}(V) \\ \frac{\sigma_v}{\sigma_\delta} (\Delta - \mathbf{E}(\Delta)) \end{bmatrix}, \qquad \sigma = \sigma_v,$$

$$\underline{Z} = \begin{bmatrix} \underline{Z}_v - \mathbf{E}(\underline{Z}_v) \\ \frac{\sigma_v}{\sigma_\delta} (\underline{Z}_\delta - \mathbf{E}(\underline{Z}_\delta)) \end{bmatrix}, \qquad n = 2n_b,$$

$$\underline{N} = \begin{bmatrix} \underline{N}_v \\ \underline{N}_\delta \end{bmatrix}, \qquad m = m_v + m_\delta,$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_v & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_\delta \end{bmatrix}.$$

We assume that the network is observable using conventional SCADA measurements and $\boldsymbol{\Sigma}_{\underline{X}}$ is the state covariance matrix (estimated using the traditional nonlinear SE approaches), $\underline{X}$ and $\underline{N}$ are statistically independent and $\mathbf{C}$ is composed of m rows of the $n \times n$ identity matrix $\mathbf{I}_n$.

III. Sensor placement problem

We desire to utilize the newly obtained low-noise PMU measurement $\underline{Z}$ to make a refined estimate $\hat{\underline{X}}(\underline{Z})$ of the entire state vector $\underline{X}$. To this end, we use the linear minimum mean squared error estimator $\hat{\underline{X}}(\underline{Z})$ given by [12]

$$\hat{\underline{X}}(\underline{Z}) = \mathbf{E}(\underline{X}|\underline{Z}) = \mathbf{E}(\underline{X}\,\underline{Z}^T)\mathbf{E}(\underline{Z}\,\underline{Z}^T)^{-1}\underline{Z}. \qquad (5)$$

The estimation error vector is defined as $\underline{\mathcal{E}} = \underline{X} - \hat{\underline{X}}(\underline{Z})$ and the error covariance matrix is given by [12]

$$\mathbf{E}(\underline{\mathcal{E}}\,\underline{\mathcal{E}}^T) = \boldsymbol{\Sigma}_{\underline{X}} - \mathbf{E}(\underline{X}\,\underline{Z}^T)\mathbf{E}(\underline{Z}\,\underline{Z}^T)^{-1}\mathbf{E}(\underline{Z}\,\underline{X}^T), \qquad (6)$$

where $\mathbf{E}(\underline{X}\,\underline{Z}^T) = \boldsymbol{\Sigma}_{\underline{X}}\mathbf{C}^T$ and $\mathbf{E}(\underline{Z}\,\underline{Z}^T) = \mathbf{C}\boldsymbol{\Sigma}_{\underline{X}}\mathbf{C}^T + \sigma^2\mathbf{I}_m$.

Our task is to find the matrix $\mathbf{C}^*$ that minimizes the total expected estimation error $\mathrm{tr}\,\mathbf{E}(\underline{\mathcal{E}}\,\underline{\mathcal{E}}^T)$. We now define the optimization problem as an integer programming problem of choosing m rows of $\mathbf{I}_n$ that minimize $\mathrm{tr}\,\mathbf{E}(\underline{\mathcal{E}}\,\underline{\mathcal{E}}^T)$.

Definition 1. *Let $C^{[m \times n]}$ denote the set of all $m \times n$ matrices composed of m rows of $\mathbf{I}_n$.* ∎

The optimization problem is then given by

$$\mathbf{C}^* = \underset{\mathbf{C} \in C^{[m \times n]}}{\arg \min}\, \mathrm{tr}\,\mathbf{E}(\underline{\mathcal{E}}\,\underline{\mathcal{E}}^T) = \underset{\mathbf{C} \in C^{[m \times n]}}{\arg \min}\, \mathbf{E}(\underline{\mathcal{E}}^T\underline{\mathcal{E}}). \qquad (7)$$

Since the first term in (6) (i.e., $\boldsymbol{\Sigma}_{\underline{X}}$) does not depend on the choice of matrix $\mathbf{C}$, we can express the minimization problem in (7) as an equivalent maximization problem using the following definition.

Definition 2. *Let the **efficacy** of a matrix $\mathbf{C}$ be defined as*

$$J(\mathbf{C}) \overset{\Delta}{=} \mathrm{tr}\left\{ \mathbf{E}(\underline{X}\,\underline{Z}^T)\mathbf{E}(\underline{Z}\,\underline{Z}^T)^{-1}\mathbf{E}(\underline{Z}\,\underline{X}^T) \right\} \qquad (8)$$

$$= \mathrm{tr}\left\{ \left[\mathbf{C}(\boldsymbol{\Sigma}_{\underline{X}} + \sigma^2\mathbf{I})\mathbf{C}^T\right]^{-1} \mathbf{C}\boldsymbol{\Sigma}_{\underline{X}}^2\mathbf{C}^T \right\}. \qquad (9)$$

[Note: the form in (9) is a generalized Rayleigh quotient.] ∎

The optimization problem in (7) is then equivalent to

$$\mathbf{C}^* = \underset{\mathbf{C} \in C^{[m \times n]}}{\arg \max}\, J(\mathbf{C}), \qquad (10)$$

which is an integer programming problem of choosing m rows of the identity matrix $\mathbf{I}_n$ that maximize the efficacy. The optimum solution to (10) requires an exhaustive search by testing all $\binom{n}{m}$ possible choices of m rows. Even for a moderately sized n and m, this becomes computationally infeasible. In fact, the sensor placement problem in the power grid is NP-complete [4].

IV. Suboptimal solutions

Since the optimization in (10) is difficult to perform, several algorithms that seek suboptimal but computationally feasible solutions have been reported [5][6][7][10]. If $\mathbf{C}$ is a suboptimal solution to (10), then it provides a lower bound on the optimal efficacy $J(\mathbf{C}^*)$, i.e., $J(\mathbf{C}) \leq J(\mathbf{C}^*)$. Therefore, the search for suboptimal solutions to closely approach (10) is equivalent to constructing tight lower bounds on $J(\mathbf{C}^*)$. A number of algorithms can be applied to find suboptimal solutions. Here, for completeness, we review two such algorithms requiring a much lower search complexity than $O\left(\binom{n}{m}\right)$.

A. Expedient solution

This is a trivial approximate solution to consider [7]. Let $J(\underline{e}_k)$ be the efficacy of the k-th unit row vector, i.e., the efficacy of the sensor placed at the k-th bus when $m = 1$. Then using (9),

$$J(\underline{e}_k) = \sum_{i=1}^{n} \frac{\left(\underline{e}_k \boldsymbol{\Sigma}_{\underline{X}} \underline{e}_i^T\right)^2}{\underline{e}_k \boldsymbol{\Sigma}_{\underline{X}} \underline{e}_k^T + \sigma^2}. \qquad (11)$$

We rank the vectors $\underline{e}_k$ in descending order of their efficacies $J(\underline{e}_k)$. For any arbitrary m, we pick the m highest ranked vectors $\underline{e}_k$ and stack them to be the rows of the approximate solution $\mathbf{C}_E$. Since this algorithm requires sorting and picking m highest ranked vectors $\underline{e}_k$, it has search complexity at most $O(n \log n)$.

This algorithm provides a useful insight into problem (10). From (11), it is clear that we want to place sensors at buses where the measured state is maximally correlated to the remaining states.

B. Greedy solution

A greedy algorithm obtains an approximate solution to (10) by making a sequence of choices [13]. At each step t, it assumes that t sensor locations are fixed, and makes a greedy choice where to place the $(t + 1)$-st sensor. Let $\mathbf{C}_G$ denote the solution provided by the greedy algorithm. The algorithm can be described by the following.

Greedy Algorithm [13]

1) Set $t = 1$ and $\mathbf{C}_t = \underline{e}^*$ such that $\underline{e}^* = \arg\max_{\underline{e} \in C^{[1 \times n]}} J(\underline{e})$.

2) Find $\underline{e}^* = \underset{\underline{e} \in C^{[1 \times n]}:\, \mathbf{C}_t \underline{e}^T = \mathbf{0}}{\arg \max} J\left(\begin{bmatrix} \mathbf{C}_t \\ \underline{e} \end{bmatrix}\right)$.

3) Set $\mathbf{C}_{t+1} = \begin{bmatrix} \mathbf{C}_t \\ \underline{e}^* \end{bmatrix}$.

4) Increment: $t \leftarrow t+1$.

5) if $t = m$ set $\mathbf{C}_G = \mathbf{C}_t$ and stop, else go to 2. ∎

Note that the greedy solution may not be optimal even for $m = 2$, but it has search complexity $O(mn)$, which is much smaller than $O\left(\binom{n}{m}\right)$ required to find the optimal solution $\mathbf{C}^*$.

V. Upper bounds on the optimal efficacy

In Section IV, we considered algorithms to obtain lower bounds on the optimal efficacy $J(\mathbf{C}^*)$. However, it is hard to evaluate how well these algorithms perform compared to $J(\mathbf{C}^*)$ since the optimal solution is not available for comparison. Therefore, we want to obtain numerically computable upper bound, $\bar{J}$, on the optimal efficacy $J(\mathbf{C}^*)$ such that

$$J(\mathbf{C}^*) - J(\mathbf{C}) \leq \bar{J} - J(\mathbf{C}).$$

We devote this section to finding a family of upper bounds $\bar{J}_k$ on $J(\mathbf{C}^*)$ by relaxing conditions on $\mathbf{C}$.

A. Definitions

To develop a family of upper bounds on the optimal efficacy, we generalize the reward function (efficacy), and generalize the optimization problem and its constraints. Instead of considering two matrices $\Sigma_{\underline{X}}^2$ and $\Sigma_{\underline{X}} + \sigma^2 \mathbf{I}$, in this section we consider a general matrix pencil $\langle \mathbf{A}, \mathbf{B} \rangle$, where $\mathbf{A}$ and $\mathbf{B}$ do not necessarily equal $\Sigma_{\underline{X}}^2$ and $\Sigma_{\underline{X}} + \sigma^2 \mathbf{I}$, respectively. Next, instead of considering matrix $\mathbf{C}$ whose entries take values in the set $\{0, 1\}$, in this section we consider an unconstrained matrix $\mathbf{F}$ whose entries take values in $\mathbb{R}$. Finally, we introduce a modified optimization problem (different from the one in Section III) that leads to the upper bounds. The following definitions set the stage.

Definition. *For two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, define the efficacy of a matrix $\mathbf{F}$, with respect to the pencil $\langle \mathbf{A}, \mathbf{B} \rangle$, as*

$$J_{\langle \mathbf{A}, \mathbf{B} \rangle}(\mathbf{F}) \stackrel{\Delta}{=} \operatorname{tr}\left\{ \left(\mathbf{FBF}^T \right)^{-1} \mathbf{FAF}^T \right\}, \quad (12)$$

[should the inverse $(\mathbf{FBF}^T)^{-1}$ exist]. ∎

Definition. *For $m \leq n$, let $\mathcal{F}^{[m \times n]}$ be the set of all $m \times n$ matrices with rank m.* ∎

Definition. *We define $\mathbf{F}_{\langle \mathbf{A}, \mathbf{B} \rangle}^*$ to be the argument that solves the following optimization problem*

$$\mathbf{F}_{\langle \mathbf{A}, \mathbf{B} \rangle}^* \stackrel{\Delta}{=} \arg\max_{\mathbf{F} \in \mathcal{F}^{[m \times n]}} J_{\langle \mathbf{A}, \mathbf{B} \rangle}(\mathbf{F}) \quad (13)$$

$$= \arg\max_{\mathbf{F} \in \mathcal{F}^{[m \times n]}} \operatorname{tr}\left\{ \left(\mathbf{FBF}^T \right)^{-1} \mathbf{FAF}^T \right\}. \quad ∎$$

Definition. *We define $J_{\langle \mathbf{A}, \mathbf{B} \rangle}^*$ as the solution to the optimization problem in (13).*

$$J_{\langle \mathbf{A}, \mathbf{B} \rangle}^* \stackrel{\Delta}{=} \max_{\mathbf{F} \in \mathcal{F}^{[m \times n]}} J_{\langle \mathbf{A}, \mathbf{B} \rangle}(\mathbf{F}) = J_{\langle \mathbf{A}, \mathbf{B} \rangle}\left(\mathbf{F}_{\langle \mathbf{A}, \mathbf{B} \rangle}^* \right). \quad ∎$$

B. Canonic theorem

Definition. *For a matrix pencil $\langle \mathbf{A}, \mathbf{B} \rangle$, if a number γ_j and a vector $\underline{u}_j$ satisfy the equation*

$$\mathbf{A}\underline{u}_j = \gamma_j \mathbf{B}\underline{u}_j, \quad (14)$$

then γ_j is called the generalized eigenvalue and $\underline{u}_j$ is called the generalized eigenvector of the pencil $\langle \mathbf{A}, \mathbf{B} \rangle$. ∎

If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ symmetric matrices, there exist n generalized eigenvectors $\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_n$, corresponding to generalized eigenvalues $\gamma_1, \gamma_2, \ldots, \gamma_n$. [Note: $\gamma_1, \gamma_2, \ldots, \gamma_n$

need not be distinct.] We arrange the generalized eigenvalues as the diagonal elements of a diagonal matrix $\mathbf{D}$,

$$\mathbf{D} \stackrel{\Delta}{=} \begin{bmatrix} \gamma_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \gamma_n \end{bmatrix}, \quad (15)$$

and we arrange the generalized eigenvectors as the columns of a matrix $\mathbf{U}$,

$$\mathbf{U} \stackrel{\Delta}{=} \left[\underline{u}_1, \ldots, \underline{u}_n \right]. \quad (16)$$

Theorem 1 (see [14]). *Let $\mathbf{A}$ and $\mathbf{B}$ be symmetric and $\mathbf{B}$ be positive definite, and let $\mathbf{D}$ and $\mathbf{U}$ denote the generalized eigenvalue matrix and generalized eigenvector matrix as in (15) and (16), respectively. If the eigenvalues are ordered as $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n \geq 0$, then*

$$J_{\langle \mathbf{A}, \mathbf{B} \rangle}^* = \operatorname{tr}\left\{ \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \end{bmatrix} \mathbf{D} \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \end{bmatrix}^T \right\} = \sum_{j=1}^{m} \gamma_j, \quad (17)$$

and

$$\mathbf{F}_{\langle \mathbf{A}, \mathbf{B} \rangle}^* = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \end{bmatrix} \mathbf{U}^T = \left[\underline{u}_1, \ldots, \underline{u}_m \right]^T. \quad (18)$$

Remark 1 (see [7]). *If $\mathbf{A} = \Sigma_{\underline{X}}^2$ and $\mathbf{B} = \Sigma_{\underline{X}} + \sigma^2 \mathbf{I}$, and $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ are the eigenvalues of $\Sigma_{\underline{X}}$, then the generalized eigenvalues of the pencil $\left\langle \Sigma_{\underline{X}}^2, \Sigma_{\underline{X}} + \sigma^2 \mathbf{I} \right\rangle$ are $\gamma_j = \lambda_j^2 / (\lambda_j + \sigma^2)$. Thus, using Theorem 1 we can write*

$$J_{\left\langle \Sigma_{\underline{X}}^2, \Sigma_{\underline{X}} + \sigma^2 \mathbf{I} \right\rangle}^* = \sum_{j=1}^{m} \frac{\lambda_j^2}{\lambda_j + \sigma^2}. \quad (19)$$

To develop a family of upper bounds, we find it useful to solve a series of modified efficacy maximization problems for all $k \leq m$. The next definition addresses the modified efficacy maximization problem.

Definition. *For any $k \leq m$, we define $\mathbf{F}_{\langle \mathbf{A}, \mathbf{B} \rangle}^{(k)*}$ and $J_{\langle \mathbf{A}, \mathbf{B} \rangle}^{(k)*}$ as the solution pair of the following modified efficacy maximization*

$$\mathbf{F}_{\langle \mathbf{A}, \mathbf{B} \rangle}^{(k)*} \stackrel{\Delta}{=} \arg\max_{\mathbf{F} \in \mathcal{F}^{[(m-k) \times (n-k)]}} J_{\langle \mathbf{A}, \mathbf{B} \rangle}\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \right), \quad (20)$$

and

$$J_{\langle \mathbf{A}, \mathbf{B} \rangle}^{(k)*} \stackrel{\Delta}{=} \max_{\mathbf{F} \in \mathcal{F}^{[(m-k) \times (n-k)]}} J_{\langle \mathbf{A}, \mathbf{B} \rangle}\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \right) \quad (21)$$

$$= J_{\langle \mathbf{A}, \mathbf{B} \rangle}\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{\langle \mathbf{A}, \mathbf{B} \rangle}^{(k)*} \end{bmatrix} \right). \quad ∎$$

In order to solve the modified efficacy maximization problem in (20) - (21), it is convenient to split the efficacy

$$J_{\langle \mathbf{A}, \mathbf{B} \rangle}\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \right)$$

into two terms such that

1) the first term does not depend on $\mathbf{F}$, and
2) the second term equals the efficacy of $\mathbf{F}$ with respect to a modified pencil of lower dimensions.

We formulate the split in the following lemma.

Lemma A. *In the optimization problem (21), the efficacy can be expressed as*

$$J_{\langle \mathbf{A}, \mathbf{B} \rangle}\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \right) = t_k + J_{\langle \mathbf{A}_k, \mathbf{B}_k \rangle}(\mathbf{F}), \quad (22)$$

where the term t_k and the modified pencil $\langle \mathbf{A}_k, \mathbf{B}_k \rangle$ satisfy

$$t_k = \mathrm{tr}\left\{ \mathbf{A} \begin{bmatrix} \mathbf{P}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right\} \tag{23}$$

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{P}_k^{-1}\mathbf{Q}_k \\ -\mathbf{I}_{n-k} \end{bmatrix}^T \mathbf{A} \begin{bmatrix} \mathbf{P}_k^{-1}\mathbf{Q}_k \\ -\mathbf{I}_{n-k} \end{bmatrix} \quad and \quad \mathbf{B}_k = \mathbf{R}_k - \mathbf{Q}_k^T \mathbf{P}_k^{-1} \mathbf{Q}_k \tag{24}$$

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0} \end{bmatrix}^T \mathbf{B} \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0} \end{bmatrix} \ and \ \mathbf{Q}_k = \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0} \end{bmatrix}^T \mathbf{B} \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{n-k} \end{bmatrix} \ and \ \mathbf{R}_k = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{n-k} \end{bmatrix}^T \mathbf{B} \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{n-k} \end{bmatrix} \tag{25}$$

Proof: For any $\mathbf{F} \in \mathcal{F}^{[(m-k)\times(n-k)]}$, using (12) we have

$$J_{\langle \mathbf{A}, \mathbf{B} \rangle}\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \right) = \mathrm{tr}\left\{ \left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \mathbf{B} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix}^T \right)^{-1} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \mathbf{A} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix}^T \right\}. \tag{26}$$

Using (25), we express $\mathbf{B}$ in terms of $\mathbf{P}_k$, $\mathbf{Q}_k$ and $\mathbf{R}_k$ as

$$\mathbf{B} = \begin{bmatrix} \mathbf{P}_k & \mathbf{Q}_k \\ \mathbf{Q}_k^T & \mathbf{R}_k \end{bmatrix}. \tag{27}$$

Substituting $\mathbf{B}$ into (26) and using the partitioned matrix inversion lemma [15], we write (26) as

$$J_{\langle \mathbf{A}, \mathbf{B} \rangle}\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \right) = \mathrm{tr}\left\{ \mathbf{A} \begin{bmatrix} \mathbf{P}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right\}$$
$$+ \mathrm{tr}\left\{ \begin{bmatrix} \mathbf{P}_k^{-1}\mathbf{Q}_k\mathbf{F}^T \\ -\mathbf{I}_{m-k} \end{bmatrix} \left\{ \mathbf{F}\left(\mathbf{R}_k - \mathbf{Q}^T\mathbf{P}_k^{-1}\mathbf{Q}_k \right) \mathbf{F}^T \right\}^{-1} \begin{bmatrix} \mathbf{P}_k^{-1}\mathbf{Q}_k\mathbf{F}^T \\ -\mathbf{I}_{m-k} \end{bmatrix}^T \right.$$
$$\left. \times \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \mathbf{A} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix}^T \right\}.$$

Using (23)-(24), we simplify the above equation as

$$J_{\langle \mathbf{A}, \mathbf{B} \rangle}\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \right) = t_k + \mathrm{tr}\left\{ \left(\mathbf{F}\mathbf{B}_k\mathbf{F}^T \right)^{-1} \mathbf{F}\mathbf{A}_k\mathbf{F}^T \right\}$$
$$= t_k + J_{\langle \mathbf{A}_k, \mathbf{B}_k \rangle}(\mathbf{F}). \quad \blacksquare$$

Lemma A now lets us express the solution of the the modified optimization problem (20)-(21) equivalently as the solution of a regular efficacy maximization (i.e., using Theorem 1), but for a modified matrix pencil. Hence, Lemma A yields the following corollary of Theorem 1.

Corollary 1.1. $\mathbf{F}_{\langle \mathbf{A}, \mathbf{B} \rangle}^{(k)*} = \mathbf{F}_{\langle \mathbf{A}_k, \mathbf{B}_k \rangle}^*$ *and* $J_{\langle \mathbf{A}, \mathbf{B} \rangle}^{(k)*} = t_k + J_{\langle \mathbf{A}_k, \mathbf{B}_k \rangle}^*$.

Proof: In (22), t_k does not depend on $\mathbf{F}$. Therefore, the equalities hold. $\blacksquare$

Remark 2. $J_{\langle \mathbf{A}, \mathbf{B} \rangle}^{(0)*} = J_{\langle \mathbf{A}, \mathbf{B} \rangle}^*$ *and* $J_{\langle \mathbf{A}, \mathbf{B} \rangle}^{(m)*} = t_m$.

C. Nested bounds

Now, using the results of Section V-B, we are ready to develop upper bounds to the original optimization problem (10) defined in Section III. These nested upper bounds are obtained under the assumption that the optimal solution to (10) is calculable for some $k \le m$. The nested upper bounds $\bar{J}_k$ are defined as follows.

Definition.

$$\bar{J}_k \triangleq \max_{\mathbf{C} \in C^{[k\times n]}} \left\{ \max_{\substack{\mathbf{F} \in \mathcal{F}^{[(m-k)\times n]} \\ \mathbf{F}\mathbf{C}^T = \mathbf{0}}} J_{\langle \Sigma_{\underline{X}}^2, \Sigma_{\underline{x}}+\sigma^2\mathbf{I} \rangle}\left(\begin{bmatrix} \mathbf{C} \\ \mathbf{F} \end{bmatrix} \right) \right\} \tag{28}$$

$\blacksquare$

From Remark 2, we clearly see that $\bar{J}_0 \ge J(\mathbf{C}^*)$. We next show that $\bar{J}_k \ge J(\mathbf{C}^*)$ for any $k \le m$.

Theorem 2. *For any $k \le m$, we have $\bar{J}_k \ge J(\mathbf{C}^*)$.*

Proof:
$$J(\mathbf{C}^*) = \max_{\mathbf{C} \in C^{[m\times n]}} J(\mathbf{C})$$
$$= \max_{\mathbf{C}_1 \in C^{[k\times n]}} \max_{\substack{\mathbf{C}_2 \in C^{[(m-k)\times n]} \\ \mathbf{C}_2\mathbf{C}_1^T = \mathbf{0}}} J\left(\begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} \right)$$
$$\le \max_{\mathbf{C}_1 \in C^{[k\times n]}} \max_{\substack{\mathbf{F} \in \mathcal{F}^{[(m-k)\times n]} \\ \mathbf{F}\mathbf{C}_1^T = \mathbf{0}}} J\left(\begin{bmatrix} \mathbf{C}_1 \\ \mathbf{F} \end{bmatrix} \right) = \bar{J}_k, \tag{29}$$

where the inequality follows from $C^{[(m-k)\times n]} \subset \mathcal{F}^{[(m-k)\times n]}$. $\blacksquare$

Remark 3. *For $k=0$, we have $\bar{J}_0 = J_{\langle \Sigma_{\underline{X}}^2, \Sigma_{\underline{x}}+\sigma^2\mathbf{I} \rangle}^* = \sum_{j=1}^{m} \frac{\lambda_j^2}{\lambda_j+\sigma^2}$.*

Remark 4. *When $k = m$, we have $\bar{J}_m = J(\mathbf{C}^*)$.*

We now show that the upper bounds are nested.

Theorem 3. *For any $k \le m-1$, we have $\bar{J}_k \ge \bar{J}_{k+1}$.*

Proof:
$$\bar{J}_{k+1} = \max_{\mathbf{C} \in C^{[(k+1)\times n]}} \max_{\substack{\mathbf{F} \in \mathcal{F}^{[(m-k-1)\times n]} \\ \mathbf{F}\mathbf{C}^T = \mathbf{0}}} J\left(\begin{bmatrix} \mathbf{C} \\ \mathbf{F} \end{bmatrix} \right)$$
$$= \max_{\mathbf{C}_1 \in C^{[k\times n]}} \max_{\substack{\underline{e} \in C^{[1\times n]} \\ \mathbf{C}_1\underline{e}^T = \mathbf{0}}} \max_{\substack{\mathbf{F} \in \mathcal{F}^{[(m-k-1)\times n]} \\ \mathbf{F}[\mathbf{C}_1^T \ \underline{e}^T] = \mathbf{0}}} J\left(\begin{bmatrix} \mathbf{C}_1 \\ \underline{e} \\ \mathbf{F} \end{bmatrix} \right)$$
$$\le \max_{\mathbf{C}_1 \in C^{[k\times n]}} \max_{\substack{\underline{f} \in \mathcal{F}^{[1\times n]} \\ \mathbf{C}_1\underline{f}^T = \mathbf{0}}} \max_{\substack{\mathbf{F} \in \mathcal{F}^{[(m-k-1)\times n]} \\ \mathbf{F}[\mathbf{C}_1^T \ \underline{f}^T] = \mathbf{0}}} J\left(\begin{bmatrix} \mathbf{C}_1 \\ \underline{f} \\ \mathbf{F} \end{bmatrix} \right)$$
$$= \max_{\mathbf{C}_1 \in C^{[k\times n]}} \max_{\substack{\mathbf{F}_1 \in \mathcal{F}^{[(m-k)\times n]} \\ \mathbf{F}_1\mathbf{C}_1^T = \mathbf{0}}} J\left(\begin{bmatrix} \mathbf{C}_1 \\ \mathbf{F}_1 \end{bmatrix} \right) = \bar{J}_k,$$

where the inequality follows from $C^{[1\times n]} \subset \mathcal{F}^{[1\times n]}$. $\blacksquare$

Corollary 3.1. $\sum_{j=1}^{m} \frac{\lambda_j^2}{\lambda_j+\sigma^2} = \bar{J}_0 \ge \bar{J}_1 \ge \cdots \ge \bar{J}_m = J(\mathbf{C}^*)$.

Proof: Combine Remark 3-4 and Theorem 3. $\blacksquare$

We next want to utilize Theorem 1 (more specifically, Corollary 1.1) to efficiently compute the upper bounds $\bar{J}_k$. To that end, we define the matrix pencil $\langle \mathbf{A}_{(\mathbf{C})}, \mathbf{B}_{(\mathbf{C})} \rangle$ as a permutation of the matrix pencil $\langle \Sigma_{\underline{X}}^2, \Sigma_{\underline{X}} + \sigma^2\mathbf{I} \rangle$.

Definition. *Let $\bar{\mathbf{C}}$ denote the **complement** of $\mathbf{C}$, with constraints $\bar{\mathbf{C}} \in C^{[(n-m)\times n]}$ and $\bar{\mathbf{C}}\mathbf{C}^T = \mathbf{0}$.* $\blacksquare$

Definition. *We define*
$$\mathbf{A}_{(\mathbf{C})} = \begin{bmatrix} \mathbf{C} \\ \bar{\mathbf{C}} \end{bmatrix} \Sigma_{\underline{X}}^2 \begin{bmatrix} \mathbf{C} \\ \bar{\mathbf{C}} \end{bmatrix}^T \ and \ \mathbf{B}_{(\mathbf{C})} = \begin{bmatrix} \mathbf{C} \\ \bar{\mathbf{C}} \end{bmatrix} \left(\Sigma_{\underline{X}} + \sigma^2\mathbf{I} \right) \begin{bmatrix} \mathbf{C} \\ \bar{\mathbf{C}} \end{bmatrix}^T. \tag{30}$$
$\blacksquare$

We now reformulate the upper bounds $\bar{J}_k$ so that it easily relates back to Corollary 1.1.

Corollary 3.2. $\bar{J}_k = \max_{\mathbf{C} \in C^{[k\times n]}} J_{\langle \mathbf{A}_{(\mathbf{C})}, \mathbf{B}_{(\mathbf{C})} \rangle}^{(k)*}$.

Proof: Let $\mathbf{F} \in \mathcal{F}^{[(m-k)\times(n-k)]}$ and let $\mathbf{F}_1 = \mathbf{F}\bar{\mathbf{C}}$, for any

$\mathbf{C} \in C^{[k \times n]}$. Then, from (12) and (30), it follows that

$$J_{\langle \mathbf{A}_{(\mathbf{C})}, \mathbf{B}_{(\mathbf{C})} \rangle} \left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \right) = J_{\langle \Sigma_{\underline{X}}^2, \Sigma_{\underline{X}} + \sigma^2 \mathbf{I} \rangle} \left(\begin{bmatrix} \mathbf{C} \\ \mathbf{F}_1 \end{bmatrix} \right). \tag{31}$$

Since $\operatorname{rank}(\mathbf{F}_1) = \operatorname{rank}(\mathbf{F}) = m - k$, and $\mathbf{F}_1 \mathbf{C}^T = \mathbf{0}$, using (28), (21) and (31) we can write

$$\begin{aligned} \bar{J}_k &= \max_{\mathbf{C} \in C^{[k \times n]}} \max_{\substack{\mathbf{F}_1 \in \mathcal{F}^{[(m-k) \times n]} \\ \mathbf{F}_1 \mathbf{C}^T = \mathbf{0}}} J_{\langle \Sigma_{\underline{X}}^2, \Sigma_{\underline{X}} + \sigma^2 \mathbf{I} \rangle} \left(\begin{bmatrix} \mathbf{C} \\ \mathbf{F}_1 \end{bmatrix} \right) \\ &= \max_{\mathbf{C} \in C^{[k \times n]}} \max_{\mathbf{F} \in \mathcal{F}^{[(m-k) \times (n-k)]}} J_{\langle \mathbf{A}_{(\mathbf{C})}, \mathbf{B}_{(\mathbf{C})} \rangle} \left(\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \right) \\ &= \max_{\mathbf{C} \in C^{[k \times n]}} J_{\langle \mathbf{A}_{(\mathbf{C})}, \mathbf{B}_{(\mathbf{C})} \rangle}^{(k)*}. \end{aligned}$$
■

For any $k \le m$, the computation of the upper bound $\bar{J}_k$ requires searching over all $\binom{n}{k}$ matrices $\mathbf{C} \in C^{[k \times n]}$. Therefore, computation of $\bar{J}_k$ has search complexity $O\left(\binom{n}{k} \right)$.

VI. Numerical performance bounds

We performed evaluations on standard IEEE test cases [16][17] to evaluate the performances of the approximate solutions and the family of upper bounds $\bar{J}_k$ (for $k = 0$ to 5). We construct the sample correlation matrix $\Sigma_{\underline{X}}$ by running traditional SE algorithms 1000 times, where each state estimate is extracted from a single scan of SCADA measurements. We used the MATPOWER package [18] as the traditional state estimator, which assumes that the physical network model, based on bus-section/switching device representation, is exact [19]. In the state estimation process, the standard deviations of voltage magnitudes, bus power injections, and line power flow measurements used are 0.01, 0.015, and 0.02, respectively, according to the default setups in [18]. For simplicity, we consider only the voltage magnitudes of the buses as the state variables. Furthermore, we assume that a PMU is always placed at the reference bus [2], and therefore, the reference bus is not considered in our sensor placement algorithms.

A. IEEE 30-bus system

First, we consider the standard IEEE 30-bus test system [16]. The simulation results are shown in Figure 2, where the upper bounds (developed in Section V) are compared to efficacy of the expedient solution $J(\mathbf{C}_E)$ and the efficacy of the greedy solution $J(\mathbf{C}_G)$. Due to the large number of buses, the exhaustive search for the optimal solution $J(\mathbf{C}^*)$ is prohibitively complex. We observe that the expedient algorithm and the greedy algorithm perform very close to each other. The upper bound $J(\mathbf{F}^*) = \bar{J}_0$ is not tight, but the nested bounds $\bar{J}_k$ become tighter as k increases.

B. IEEE 57-bus system

Next, we consider the IEEE 57-bus system [17] and the simulations results are shown in Figure 3. Again, the size of the system prevents us from calculating the optimal efficacy $J(\mathbf{C}^*)$. Similar to the IEEE 30-bus system, the expedient solution and the greedy solutions are indistinguishable. The
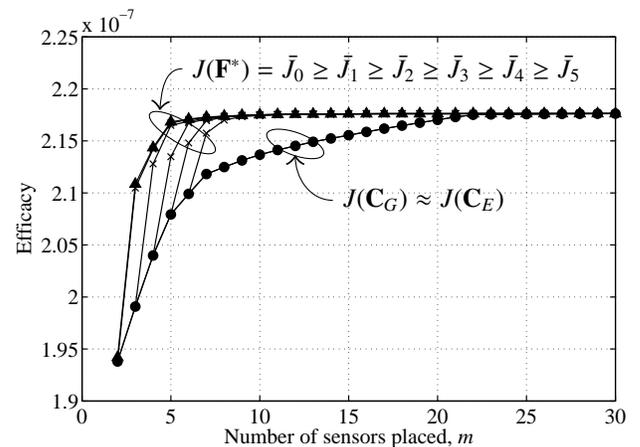
Figure 2. IEEE 30-bus test system: efficacies of approximate solutions and the family of upper bounds $\bar{J}_k$
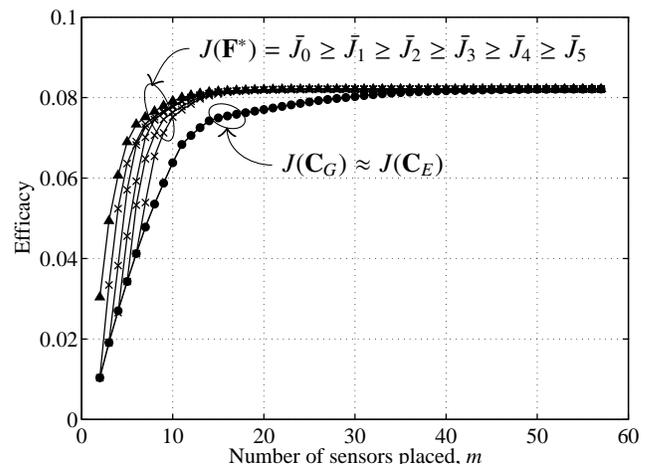
Figure 3. IEEE 57-bus test system: efficacies of approximate solutions and the family of upper bounds $\bar{J}_k$

nested bounds $\bar{J}_k$ also get tighter as k is increased, confirming that the expedient and the greedy solutions are very close to (if not) the optimal solution in this scenario.

VII. Conclusion and Future Work

In this paper, we considered the optimal placement of m PMUs among n bus locations and formulated the optimization problem as an integer programming problem. To understand the performance of the (sub)optimal solutions, we presented a series of nested upper bounds (using generalized eigenvectors and matrix manipulations) that give tighter upper bounds at increasing complexity. These bounds were then numerically evaluated for the IEEE 30- and 57- bus test systems. Some further directions for this research may be finding tighter theoretical bounds when the covariance matrix, $\Sigma_{\underline{X}}$ has some structure (e.g., diagonally block dom-

inant), or when there are additional constraints (e.g., PMUs are not allowed at certain bus locations).

REFERENCES

[1] Y.-F. Huang, S. Werner, J. Huang, N. Kashyap, and V. Gupta, "State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid," IEEE Signal Process. Mag., vol. 29, no. 5, pp. 33–43, Sep. 2012.
[2] A. G. Phadke and J. S. Thorp, Synchronized Phasor Measurements and Their Applications, Springer, 2008.
[3] A. Abur and A. Gómez-Expósito, Power System State Estimation: Theory and Implementation, New York, NY: Marcel Dekker, Inc., 2004.
[4] D. Brueni and L. Heath, "The PMU placement problem," SIAM Journal on Discrete Mathematics, vol. 19, no. 3, pp. 744–761, Dec. 2005.
[5] Q. Li, R. Negi, and M. D. Ilić, "Phasor measurement units placement for power system state estimation: A greedy approach," in 2011 IEEE PES General Meeting, Detroit, MI, July 2011, pp. 1–8.
[6] V. Kekatos, G. B. Giannakis, and B. Wollenberg, "Optimal placement of phasor measurement units via convex relaxation," IEEE Trans. Power Syst., vol. 27, no. 3, pp. 1521–1530, Aug. 2012.
[7] M. Uddin, A. Kuh, A. Kavcic, and T. Tanaka, "Approximate solutions and performance bounds for the sensor placement problem," in Proc. 2012 Third IEEE Int. Conf. Smart Grid Commun. (SmartGridComm), Tainan, Taiwan, Nov. 2012.
[8] S. Chakrabarti, G. K. Venayagamoorthy, and E. Kyriakides, "PMU Placement for power system observability using binary particle swarm optimization," in Proc. Australasian Universities Power Engineering Conf., Sydney, Australia, Dec. 2008, pp. 1–5.
[9] D. Dua, S. Dambhare, R. K. Gajbhiye, and S. A. Soman, "Optimal multistage scheduling of PMU placement: an ILP approach," IEEE Trans. Power Del., vol. 23, no. 4, pp. 1812–1820, Oct. 2008.
[10] N. M. Manousakis, G. N. Korres, and P. S. Georgilakis, "Taxonomy of PMU placement methodologies," IEEE Trans. Power Syst., vol. 27, no. 2, pp. 1070–1077, May 2012.
[11] A. Monticelli, "Electric power system state estimation," Proc. IEEE, vol. 88, no. 2, pp. 262–282, 2000.
[12] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Englewood Cliffs, NJ: PTR Prentice Hall, 1993.
[13] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Introduction to Algorithms, 3rd ed., Cambridge, MA: MIT Press, 2009.
[14] P.-A. Absil, R. Mahony, and R. Sepulchre, Optimization Algorithms on Matrix Manifolds, New York, NY: Cambridge University Press, 2005.
[15] K. M. Abadir and J. R. Magnus, Matrix Algebra, Princeton, NJ: Princeton University Press, 2008.
[16] R. W. Ferrero, S. M. Shahidehpour, and V. C. Ramesh, "Transaction analysis in deregulated power systems using game theory," IEEE Trans. Power Syst., vol. 12, no. 3, pp. 1340–1347, Aug. 1997.
[17] Power system test case archive. University of Washington. [Online]. Available: http://www.ee.washington.edu/research/pstca/ [retrieved: Feb. 2013].
[18] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," IEEE Trans. Power Syst., vol. 26, no. 1, pp. 12–19, Feb. 2011.
[19] F. C. Schweppe and J. Wildes, "Power System Static-State Estimation, Parts I–III," IEEE Trans. Power Apparatus and Systems, vol. PAS-89, no. 1, pp. 120–135, Jan. 1970.

Smart Energy Management System for Home Area Networks

Tan Yi, Wang Yiming, Wang Siying, Rong Hongqian,
Chang Yuan, Zhong Jianlun, Dai Xun, Tiankui Zhang,
Chaowei Wang
Beijing University of Posts and Telecommunications
Beijing, China
jp092816@bupt.edu.cn

Yue Chen, Kok Keong Chai
Queen Mary University of London,
London, United Kingdom
Yue.Chen@eecs.qmul.ac.uk;
Michael.Chai@eecs.qmul.ac.uk

Abstract - **Distributed renewable energy resources systems are becoming widely used in home area networks. These systems can intelligently select energy resources to maintain high efficiency power management at home and energy grids based on the power consumption of home appliances and availability of distributed energy harvesters. This paper demonstrates a smart strategy on energy source selection and automatic power control for decentralized electrical equipment and appliances in a home area network. The proposed smart strategy takes into account the real-time electrical pricing, environmental data from sensors and prediction of energy consumption based on Grey Model Forecasting within home. The strategy is demonstrated in a testbed that adopts uIP stack (IPv6) in CC2530 chip.**

Keywords-Energy Source Selection; Automatic Power Control; Home Area Networks.

I. INTRODUCTION

There are three main factors that characterize energy landscape: the growing energy demand, the need for sustainability and the global pressure to cut CO2 emissions. With the study, by 2030, power consumption is expected to grow from 20,000 TWh today to roughly 33,000 TWh; that's a leap of over 60 per cent [1].

Nevertheless, the traditional grid is only a one-way energy broadcasting network and usually has one energy source for users. This type of energy grid topology could seriously affect consumers' energy utilization if there is any problem in the energy supply, as there is no alternative source for consumers. Therefore, consumers and utility companies are seeking more energy efficient and sustainable solutions.

Wireless Sensor Networks (WSN) [2] have emerged as an exciting technology for a wide range of important applications that acquire and process information from the physical world. Meanwhile, increased penetration of small modular generation technologies interconnected to distribution systems form a new type of power system. The traditional architecture of one-way energy broadcasting network and single national energy source in the traditional grid will be replaced by the emerging smart grid system [3] that combines the emerging WSN application and an intelligent distributed energy resources system. A Smart Grid is an electrical grid that uses computers and other technology to gather and act on information, such as information about the behaviours of suppliers and consumers, in an

automated fashion, to improve the efficiency, reliability, economics, and sustainability of the production and distribution of electricity. The smart grid system can satisfy the requirements of the flexibility in energy generation and selection, efficiency in power consumption and reliability in emergency [4].

A variety of countries and regions have proposed their development plans according to their regional features. In this paper, we briefly introduce the smart grid in USA, Europe, and China.

M. Lowe et al. [5] discuss the grid state in USA. Due to serial large outages in recent years, the industry has paid more attention to the quality and reliability of electricity. The increasing demand for national security and environment protection leads to a higher standard of grid construction and management

In Europe, the government especially addresses the environment protection. The European smart grid focuses more on the assessment of renewable resources, the impact on the wildlife and the real-time monitoring and remote control [6]. The European Technology Platforms (ETP) began to operate in 2005 [7] in formulating and promoting a vision for the development of smart grids for year 2020 in compliance with EU policy.

The grid in China exposes the weakness in dealing with the emergencies. The lack of smart power distribution leads to the regional and seasonal electricity shortage. Faced with these problems, IBM proposed that it can provide a whole scheme solution Architecture for Energy (SAFT) for the power companies in China to use the smart power grid effectively [8].

As mentioned above, smart grids in different regions in the world are trying to achieve the goal of reliability and efficiency of energy demand. Real-time analysis is one of the key technologies for achieving the goal. One of the notable features of smart grid is the rapid response to the environment changes. The supervisory control and data acquisition (SCADA) systems have been used in transmission networks for several decades. They provide control centre operators with real-time data concerning the network state and allow remote manual, automatic or semi-automatic procedures [9]. The real-time analysis is no doubt the key technique for SCADA system. The sensors collect the data of the environment and periodically

transmit the data to near communication devices. The system makes quick analysis of the data and generates the optimal decisions, then it sends the control signals to end devices. The whole procedure is completed in very little amount of time and repeated frequently enough to guarantee the real-time of the smart grid.

Many countries have proposed to implement smart grid technology that focuses on how to validate intelligent control in the process of electrical generation, transmission and distribution, and how to access renewable sources on a large scale such as district, city or country wide. However, there is a limited development in smart energy management for home area
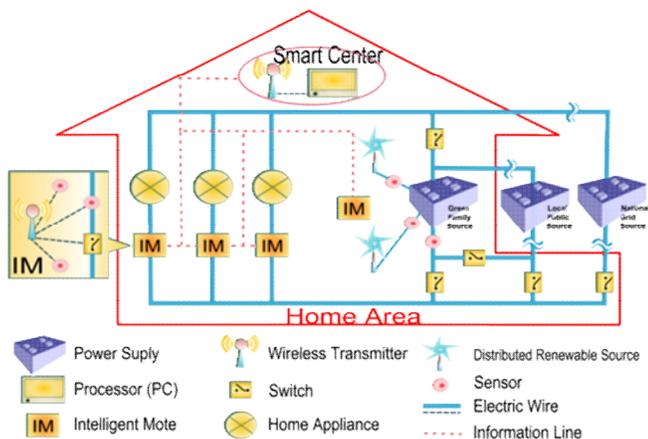
Figure 1. System Architecture of Home Area Network prototype

networks.

This paper aims at demonstrating a home area network to integrate the technology of WSN into the traditional grid, and to combine the family levelled distributed renewable sources into energy grid, meanwhile, to make power consumption more efficient. It also shows a smart strategy based on grey prediction model [10] in energy source selection and intelligent power control by predicting the future energy consumption and pricing variation. Figure 1 shows the architecture of home area network

The rest of this paper is organized as follow: Section II explains the home area network prototype and its functionalities. The hardware implementation and the smart energy management system implementation are explained in Sections III and IV, respectively. Finally, the conclusion is drawn in Section V.

II. HOME AREA NETWORK PTOTOTYPE

A. Architecture of home area network protoype

Figure 2 shows the design of a home area network. The intelligent motes (IMs) are integrated into the electrical switches, wireless transmitters and variety of sensors. They are used to collect environment information, transmit sensing and control data.
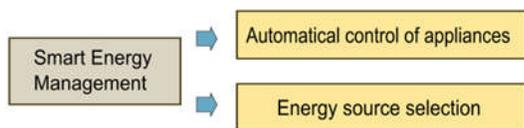
Figure 2. Smart Energy Management

There are current sensors, temperature sensors, infrared sensors, optical sensors and carbon oxide sensors in this prototype. Each sensor and mote performs according to specific condition. The prototype divides into three parts:

a) Smart Center

Smart Center is the main controller with a wireless transceiver. It can synthesize and analyze the information received from all IMs; send back data to distributed IMs, information it sends such as the real time price response to the user terminal and the control signal to the wireless switch (control the access of different loads and power sources). It also displays data via a user friendly interface.

b) Energy Supply Terminal

The household has three energy resources: a green family source, which stores electricity via distributed renewable energy harvesters for its own use and sell it to outside; a local public source, which is the neighbourhood energy storage and the neighbourhood electricity seller; and a national grid source which consistently provides power to neighbourhood household. The IM connected to green family source reports the household resource's condition to the smart center regularly and in the charge of energy source selection by controlling the five electrical switches.

c) Home Appliance Terminal

Each home appliance is connected with an IM. The IM collects sensor information from its appliance and forwards it to smart centre or other IMs; it also controls its appliance according to the suggested strategy.

B. Main Functionalities of the prototype

a) Data Transmission and Prediction

Energy consumption data collected by IM is sent to and stored in smart center which applies Grey Model (GM) prediction to make control decision.

b) Energy Resource Selection

Based on energy consumption prediction, condition of green family resource and the real time electrical pricing, the smart center decides the energy resources to run the household appliances and electrical equipment, and advises whether the surplus energy from the household resources can be resold to the local public resource.

c) Intelligent Power Control

Each IM can make automatic response in consideration of the characteristics of its connected home appliances and the environmental condition.

d) Remote Comtrol through Mobile Devices

Authorized users can view the household electrical condition on mobile phone in real time, to read the energy consumption and its cost. It also controls the strategy of each home appliance.

C. Use Case

a) Real Time Charge Response

Each IM connected to the users will transmit its power usage status to the smart meter by wireless communication, and with the aid of users' historical records of power usage (stored in a database), the Smart Center could respond to the power usage

immediately based on a given algorithm and feedback the real time charge to users in time.

b) Automatic Source Access Selection

With adding the distributed renewable energy harvester, each house has options to choose source. The determination for power choice will be realized in consideration of a variety of factors:

- The future prediction and current energy consumption of user terminal;
- The energy generation prediction of the distributed energy harvester;
- The real-time price variation.

Generally speaking, the priority of energy source access is depended on the contrast of the current power charge of main grid source and the distributed source, the lower price, the higher access priority.

c) Emergency Control to User Terminal

By wireless communication, the smart meter can actively restrict some part of users' power need in some power shortage condition. For instance: if there is earthquake happened, the main grid is partly destroyed and cannot provide power to this region in time, plus the limited ability of renewable energy source, it cannot provide enough power to all the users simultaneously, and by remote controlling wireless switches, the smart meter can restrict some factories' power usage and focus the power supply to the some more important users immediately, like hospital, transport & communication station and so on.

III. HARDWARE IMPLEMENTATION

A. Intelligent Mote

Intelligent motes are the primary important components for achieving the distributed strategy and the execution of the command from the centre. They are connected to appliances or the energy supply and switches. Each of them consists of CC2530F256 and the peripheral circuit containing necessary sensors.

In the mote, sampling resistor circuit is used to sensing the circuit. And the output is converted to digital numbers by TM7709 which is an analogue to digital converter (ADC). The voltage value is measured in a similar way. Then the power consumed by the appliance is computed by multiplying the value of the current and the value of the voltage. The energy consumption data is sent to the centre every 3 seconds to help the smart strategy of the centre. The hardware of our motes is shown in Fig 3.
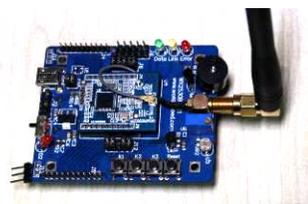
Figure 3. Hardware of Mote

B. Smart Appliances

The circuit in the prototype is a direct current circuit whose voltage is below 5V. There are three appliances (Figure 4) in the prototype. A led lamp panel simulates the lighting in the house. A LED simulated the washing machine. There is also a fan which can response to the environmental parameter smartly. They send their energy consumption data to the centre every 3 seconds.

Figure 4. Smart Appliances and Information Network

a) Lighting

The lighting can decide whether to switch on or off based on the data sensed from the light sensor and the infrared sensor. Meanwhile, it send its working state (the light intensity) when it is on to the centre every 3 seconds.

b) Washing Machine

It receives the data of electricity price from the centre every 3 seconds. Customer can schedule the completion time of a washing task, set a desired threshold price. In a period before the latest starting time corresponding to the completion time, the price is regularly checked and when the price is lower than the threshold, the washing machine starts washing. It reports to the centre when it starts working.

c) Fan

The fan can decide whether to switch on or off and the speed based on the data from the infrared sensor and the temperature data sent by the centre. Similar to the lighting, it sends its working state (the speed) when it is on to the centre every 3 seconds.

C. Energy Supplies
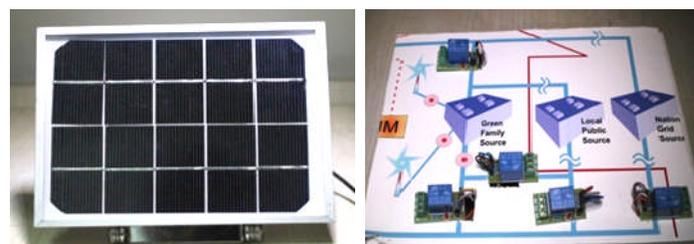
Figure 5. Solar Panel & Source Switches

In our prototype, a solar panel (Figure 5) is used as renewable energy source. An intelligent mote is connected to it, senses the value of energy generated and reports to the centre. Moreover, the switches which control the power access are connected to the mote. The centre sends commands of energy supply selection to

this mote. Then the mote chooses the corresponding energy supply using the switches.

D. Information Network.

Two-ways communication allows the information change between the intelligent motes and the centre. Star topology network is adopted in our prototype. The intelligent motes are the children nodes and the mote connected to the centre PC is the parent node. Every node has an ipv6 address and they communicate based on 6LowPAN. Contiki is the operating system used in the mote. The standard protocol in the lower layer is 802.15.4. And we adopt the uIP stack which is a lightweight open source TCP/IP stack in the Contiki system.

IV. SMART ENERGY MANAGEMENT SYSTEM

A. Software Design Pattern

Figure 6 shows the design pattern of software architecture.
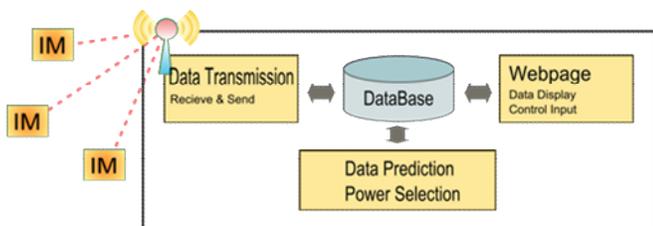
Figure 6. Software Architecture

All data including power consumption, working state from intelligent mote will be received and stored into a database. The software center will access the database every 3 seconds, making prediction and decision for power source selection, all these result data will be fed back to IM, meanwhile it will be stored into the database. Webpage can access the database, showing wanted data on the mobile terminals, which can help users check the condition for their appliance and energy whenever they want and wherever they are. They can also interact with the webpage and all their command will be sent to the corresponding mote and also stored into the database, and could be processed by the software centre for the next 3 seconds cycle.

B. Data Flow

Figure 7 shows the data flow chart of the proposed smart energy management.

Figure 7. Data Flow Chart

Smart Center is designed for data processing and decision making. It will receive 3 kinds of data from outside environment every 3 seconds: 1) power comsumption for each appliance; 2) power generation from home harvester; 3) Real-time electricity price of other power supply. And based on source 1) and 2) ,the center can make prediction for future demand and generation. When all data is ready, the center will make decision on souce selection. If the decision alters, it will send the change to the

mote which controls the sower source, and update otherwise the center will process the data for the next 3 seconds.

C. Prediction Algorithm

The prediction takes advantage of the modelling mechanism from the traditional GM (1, 1) model [11]. Every 3 seconds, the system forecasts a series of future data by using previous collected energy consumption data as original series.

Via pre-processing, 8 individual points of average power consumption which is based on the recent 80 points. It is donated as $X^{(0)}\{X^{(0)}(1), \dots, X^{(0)}(8)\}$. Taking them as raw series to get the one-accumulation series $X^{(1)}\{X^{(1)}(1), \dots, X^{(1)}(8)\}$, where $X^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \ (k = 1,2,\dots,8)$. (1)

$X^{(1)}$ satisfies the following grey differential equation, of which the winterization form is $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$, where a is called developing coefficient, and b is called grey input. They can be retrieved by the following equations:

$$z^{(1)}(k+1) = \frac{1}{2}\left[x^{(1)}(k) + x^{(1)}(k+1)\right] \tag{2}$$

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(8) & 1 \end{bmatrix} \qquad Y_n = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(8)]^T$$

$$\alpha = \begin{pmatrix} a \\ b \end{pmatrix} = (B^T B)^{-1} B^T Y$$

(3-5)

Then deduce the subsequent series $x^{\hat{}(1)}$ by using the a, b get above, $\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a}$. (6)

Finally, implement inverse accumulation on $x^{\hat{}(1)}(k)$ will get the prediction series: $\hat{x}^{(0)}(k+1) = (\hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k))$
$= (1 - e^a)(x^{(0)}(1) - \frac{b}{a})e^{-ak}$, where $k = 1, 2, 3 \dots 8$. (7)

Figure 8 illustrates a sample of predicted results based on the GM.

Prediction

	1	2	3	4	5	6	7	8	9	10	11
real data	4491	4891	5360	7602	6446	5965	4928	5512			
prediction									5800	5858	5918
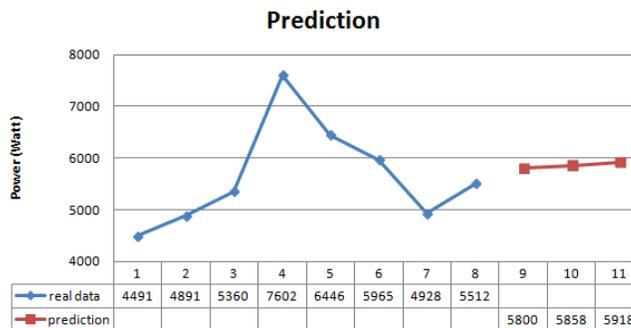
Figure 8. Sample of predicted results

D. Energy Source Selection

Power selection is an important part in the prototype software design. Every 3 seconds, the center will go through a decision process. Based on the overall power consumption, and energy generation, the center will first make prediction for demand and supply. On stage 1, the center will tell whether family power supply can meet the demand for short period (30s in the prototype). If yes, it will go to stage 2, otherwise it will use external power supplies. On stage 2, the center will further judge whether the home power supply can meets demand for a longer time, if yes, it will go to stage 3, otherwise it will use external

supplies when their price are lower than the set price for home supply and family can store power in case of future lack of electricity when external price is high. On stage 3, the center will make decision whether to make money by selling electricity to other sources, so that the family can make extra profits by their green energy harvester.

E. User Interface

Figure 9 shows the user interface of the smart control for smart energy management system.

Figure 9. User Interface

The interface is implemented with PHP Language. We can remote view the state and control operation of the family appliances.

V. CONCLUSION

Overall, this smart energy management system for home area network focuses on optimizing the energy efficiency and advocating the penetration of green energy into household. This demo implements the idea by establishing wireless sensor communication network and proposing an algorithm for energy resource selection and intelligent power control in response with environmental factors, real time consumption demand and pricing information. In the future, this algorithm can be further developed by allocating part of center's prediction work to distributed nodes, to make forecasting faster with lower energy

consumption, and at the same time, to improve the reliability of the control system.

VI. ACKNOWLEDGEMENTS

VII. REFERENCE

[1] Siemens Energy Corporation, "A smart approach to the future", http://www.ngpowereu.com/article/A-smart-approach-to-the-future/.

[2] Y. Jia, Z. Yang and B. Cao; "Research on Distributed Renewable Energy Power System"; Power Electronics, 2005.

[3] Lowfield Heath, "Renewable Energy Commercialization: Renewable Energy in Australia, Dye-Sensitized Solar Cell" ISBN 978-1-157-71383-8

[4] Energy Future Coalition, "Challenge and Opportunity: Charting a New Energy Future," Appendix A: Working Group Reports, Report of the Smart Grid Working Group.

[5] M. Lowe, H. Fan and G. Gereffi, "U.S. Smart Grid，Finding new ways to cut carbon and create jobs", Center on Globalization, Governance & Competitiveness Duke University, 2011.

[6] European Commission, "Vision and Strategy for Europe's Electricity Networks of the Future". Smart Grids European Technology Platform, 2006.

[7] V. Salehi, A. Mazloomzadeh and O. Mohammed, "Real-time analysis for developed laboratory-based smart micro grid", IEEE Power and Energy Society General Meeting, 2011.

[8] J. Lu, D. Xie and Q. Ai, "Research on smart grid in China", Transmission & Distribution Conference & Exposition: Asia and Pacific, 2009.

[9] D. Zhang, "The improvement of the self-healing smart grid compared to traditional grid", Yunan Electric Power, Vol. 38, No.4, Aug. 2010.

[10] S. Liu and J. Deng, "The Range Suitable for GM (1,1)", Systems Engineering Theory & Practice, 2000.

[11] X. Wang, H. Wang and L. Hou, "Electricity Demand Forecasting Based on Three-point Gaussian Quadrature and Its Application in Smart Grid" International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), 2010.

Profiling Power Consumption on Mobile Devices

Luca Ardito, Giuseppe Procaccianti, Marco Torchiano, Giuseppe Migliore

Dipartimento di Automatica ed Informatica

Politecnico di Torino

Torino, Italy

E-mail: name.surname@polito.it

Abstract—The proliferation of mobile devices, and the migration of the information access paradigm to mobile platforms, motivate studies of power consumption behaviors with the purpose of increasing the device battery life. The aim of this work is to profile the power consumption of a Samsung Galaxy I7500 and a Samsung Nexus S, in order to understand how such feature has evolved over the years. We performed two experiments: the first one measures consumption for a set of usage scenarios, which represent common daily user activities, while the second one analyzes a context-aware application with a known source code. The first experiment shows that the most recent device in terms of OS and hardware components shows significantly lower consumption than the least recent one. The second experiment shows that the impact of different configurations of the same application causes a different power consumption behavior on both smartphones. Our results show that hardware improvements and energy-aware software applications greatly impact the energy efficiency of mobile devices.

Keywords-Software; Energy Aware; Energy Profiling; Power Consumption; Android;

I. Introduction

Modern handheld devices are increasing their diffusion sharply. This means that, inevitably, the paradigm for accessing information and Internet services is migrating to mobile. The computational capabilities of this class of devices are growing rapidly, while their size is decreasing and they are increasingly becoming part of our daily life by providing us a wide range of functionality. Many features of modern devices like high processor speed, more efficient displays, more powerful data storage, Wi-Fi/GPRS/UMTS network adapters, advanced 3D graphics, etc. enable users to connect the now pervasive network, allowing access to personal data, as well as public resources anytime, anywhere. However, the use of these features significantly affects the energy consumption of the portable device and the advances in the development of battery technologies cannot keep pace with the rapid growth of energy demand.

This issue moved technology producers, information systems managers, and researchers to deal with energy consumption reduction [1]. For this reason, research has increasingly focused on improving the energy efficiency of hardware, but the literature still lacks in quantifying accurately the energy impact of software. Software does not consume energy directly, however it has a direct influence on the energy consumption of the hardware underneath. The battery is thus apparently a key parameter to control, in order to manage the power budget of the mobile device. It is essential to have precise figures of the current energy consumption of mobile devices to understand how to reduce their power consumption and how to design future energy efficient equipment. Today these figures are incomplete and not precise. The aim of this work is profiling the energy consumption behavior of a Samsung Galaxy I7500 and of a Samsung Nexus S in order to understand how the typical energy consumption of smartphones varied over the years.

We designed this work in two parts: the first one aims at identifying different usage scenarios corresponding to "high-level features", which do not take into account what is underneath. The second part profiled the power consumption of a known source code, which provides context-aware functionalities.

The main contributions of this paper are:
- A profiling test bench, which allows executing defined scenarios on mobile devices and profiling the related power consumption through external measurement hardware;
- A comparison between power consumption of two different generations of Android OS-based mobile devices validated by statistical analysis of gathered data.

The remainder of this paper is structured as follows: Section II introduces the related work, Section III describes the context of our work, including instrumentation and research questions, Section IV presents results while Section V discusses them and, finally, Section VI provides conclusions and future works.

II. Related Work

The energy profiling of mobile devices is an active research stream, especially as regards mobile and embedded devices. The concept of energy-awareness is based upon a complete knowledge on how and where energy is consumed on a device. In [2], authors present a detailed analysis of power consumption in a mobile device, focusing on the hardware subsystems, through common and realistic usage scenarios. Results show that the GSM module and the display are the most power-consuming components: for example, a GSM phone call on OpenMoko Neo Freerunner, HTC Dream G1 and Google Nexus One consumes 1135 mW, 822 mW and 846 mW respectively.

Usually, an accurate power consumption analysis of mobile or embedded devices is component-based. However,

instantaneous information about discharge current and remaining battery capacity is not always available, because most devices do not have built-in sensors to collect these data. In [3], a technique called PowerBooter is proposed to build a battery-based model automatically. Authors motivate this decision by considering that different mobile devices of the same category show different power consumption, and a specific power consumption model for each device is difficult to obtain. Thus, instead of using external metering instrumentation to detect power consumption, only the internal battery voltage sensor is used, which is found across many modern smartphones.

From a software engineering point of view, most contributions are devoted in developing frameworks and tools for energy metering and profiling. Also in [3], authors propose an on-line power estimation tool called PowerTutor. It implements the PowerBooter model in order to profile power consumption of applications, basing upon their component usage. Another example, which makes use of external metering devices, is ANEPROF [4], which authors define as a real-measurement-based energy profiler able to reach function-level granularity. It is developed for Android OS based devices, thus it is aimed at profiling Java applications. It is based on JVM event profiling, using software probes to record runtime events and system calls. Authors had to address several design issues, such as overhead control and proper time synchronization. Power consumption profiling is made through correlation of real-time power measurements done by an external DAQ, connected to a ARM Computer-on-Module running Android 2.0. Authors also provide profiling data of four popular applications (Android Browser, Gmail, Facebook, YouTube). The accuracy of ANEPROF depends on the hardware meter used. Its CPU overhead is stated to be less than 5%. Finally, SEMO [5] is a smart energy monitoring system, developed for Android, which provides also application-level consumption monitoring. This system is composed of three components: an inspector, which monitors the information on the battery, warning users when the battery reaches a critical condition; a recorder, which basically logs the actual charge of the battery and the running applications, and an analyzer, which calculates the energy consumption rate for each application and ranks them according to it.

As we have shown in this section, several efforts have been made as regards energy profiling in mobile devices. However, these works differ greatly in terms of methodologies and formalisms used. Palit et al. in [6] propose an interesting framework for performing experiments to measure the energy cost of software applications on smartphones.

They define the concept of user-level test case γ_i as a pair $<$ *input; output* $>$ where the input is composed of an application setting α_i and a device configuration β_i, and the output is the energy cost θ_i, expressed as a custom metric depending on battery capacity and amount of current consumed. Formally: $\gamma_i = < \alpha_i, \beta_i; \theta_i >$

Authors describe also a typical workbench for experimentation, which is very similar to the one we used in this work. In this contribution, we will follow the methodology and approach used in our previous work [7], where we performed an analog experiment aimed at assessing the software impact over power consumption in Desktop computer systems.

III. STUDY DESIGN

The aim of our research is to compare the impact of software usage on power consumption in two different Android OS-based mobile phones. For this purpose, we performed two experiments: the first one, called "training tools", fixes a set of high level features in order to compare the power consumption between the two devices, while the second one, called "gLCB", executes different profiles of a Context-Aware application and compares its power consumption on the two devices according to the selected profiles.

A. Variable selection

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Experiment 1: Training Tools. We aim at quantifying, in two different models of smartphone, the power consumption of hardware components, when performing daily activities for a common user. We selected two independent variables: the smartphone model (M) and the specific scenario (S). Each has been executed 30 times, with a fixed duration of 4 minutes per scenario. Our dependent variable is the consumed power (P). The scenarios are:

S0: Standby. This scenario provides the baseline for our analysis. During this scenario, there are no user applications in execution, and 2G and 3G connections are enabled.

S1: Phone call over 2G network. This scenario executes a phone call to a prefixed number, for a total duration of 4 minutes.

S2: Phone call over 3G network. Same as above, except that the call is made using the UMTS network.

S3: File download through Wi-Fi connection. In this scenario, the scheduled task launches a new thread, which downloads a remote file, the Ubuntu 11.10 disk image, up until the scheduled timeout (4 minutes).

S4: File download through 2G (EDGE/GPRS) connection. Same as above, except that the downloaded file is the Android SDK, which is smaller in size.

S5: File download through 3G (UMTS) connection. Same as above, except that the UMTS network is used.

S6: Localization activity through GPS. This scenario manages position updates. The task simply registers on location updates and reads the new values of latitude and longitude, up to the 4 minutes timeout.

S7: Scan for Bluetooth devices. In this scenario, a scan for Bluetooth devices is performed. The scan process lasts, according to specifications, 12 seconds in average. At the end of the scan procedure, the task simply restarts, up until the prefixed duration.

S8: CPU-intensive activity. The aim of this scenario is maintaining a high CPU workload while gathering power consumption data. For this purpose, repeated cryptography

operations are performed, with a pool of 20 threads, each of them iterating the procedure 10 times.

S9: Playback of an audio file. This scenario plays an mp3 compressed audio file, 4.78 MB in size, played in loop up until the scheduled timeout.

S10: Active display with 50% Brightness. The aim of this scenario is assessing the impact of the active display over power consumption. This scenario is similar to S0, the only difference being that all radios (2G, 3G, Wi-Fi) and the SIM card were disabled.

Experiment 2: gLCB. We analyze the energetic behavior of an application, called gLCB, the Android porting of a Context-Awareness application developed by Telecom Italia Lab [9]. Basically, its purpose is to retrieve diverse context information (such as geographical location, Wi-Fi hotspots, Bluetooth devices, etc.) from a portable device, in order to send it to a remote Context Provider for the implementation of Context-Aware services to the end user.

gLCB is based on an event mechanism, that triggers the data upload only when a context change is detected. Depending on the usage profile chosen for the application, which can be one of the following: *VERY LOW, LOW, NORMAL, HIGH, AUTO, CUSTOM*, the data retrieving and upload ratio are adjusted, thus affecting the energy behavior of the application. In our experiment, each profile was set through a server application and data was collected during execution sessions of the gLCB application of the fixed duration of 60 minutes for each profile.

B. Hypothesis Formulation

We define our goal through the Goal-Question-Metric (GQM) approach. [10]. For Experiment 1 the goal is: "Analyze usage scenarios of two mobile devices for the purpose of assessing differences with respect to power consumption from the viewpoint of the System User in the context of mobile applications" while for Experiment 2 the goal is: "Analyze usage profiles of gLCB source code for the purpose of assessing differences with respect to power consumption from the viewpoint of the System User in the context of mobile applications".

1) Experiment 1: Training Tools:
- *Research Question 1.1: Do usage scenarios have the same power consumption?*
  $H_{0,1} : P_{0,1} = P_{1,1} = ... = P_{10,1}$
  $H_{0,2} : P_{0,2} = P_{1,2} = ... = P_{10,2}$
- *Research Question 1.2: Is the energy consumption the same among the devices?*
  $H_0: P_{i,1} = P_{i,2}$, $i \in [0, 10]$
  $H_A: P_{i,1} \neq P_{i,2}$, $i \in [0, 10]$

2) Experiment 2: gLCB:
- *Research Question 2.1: Does gLCB cause a variation of the devices power consumption?*
  $H_{0,1}: P1_{\text{with gLCB}} = P1_{\text{without gLCB}}$
  $H_{0,2}: P2_{\text{with gLCB}} = P2_{\text{without gLCB}}$
- *Research Question 2.2: Are there statistical differences between different user profiles?*
  $H_{0,1}: P_{1 \text{ high}} = P_{1 \text{ normal}} = P_{1 \text{ low}} = P_{1 \text{ verylow}}$

$H_{0,2}: P_{2 \text{ high}} = P_{2 \text{ normal}} = P_{2 \text{ low}} = P_{2 \text{ verylow}}$
- *Research Question 2.3: Are there statistical differences between the behaviors of gLCB in different devices?*
  $H_0: P_{i,1} = P_{i,2}$, $i \in$ (high, normal, low, verylow)

The metric is Power consumption.

C. Instrumentation and Experiment Design

The selected usage scenarios have been implemented in Java code using the Android SDK. In order to obtain a statistically relevant data set, each scenario has a fixed execution time of 4 minutes, and each execution was repeated 30 times. This procedure was equally applied on each smartphone.

Hardware Instrumentation. The experiments were performed on two different models of smartphones: the "Galaxy i7500", first announced in April, 2009, which is the first model produced by Samsung based on Android OS; and the "Nexus S", first announced in December, 2010, produced by Google and Samsung. Their technical specifications are listed in the producer website. The power consumption data was acquired through a power metering architecture. The battery was removed from the devices, in order to avoid bias due to discharge and subsequent OS power saving procedures. The battery terminals were directly connected to a DC power supply, providing 5 V steadily. This value was chosen after different tests, that showed how lower values were not able to maintain the device operational during the most power consuming scenarios, because of the voltage drop on the shunt resistance. The DC power supply used is the TPS-2000D produced by Topward Electric Instruments Co. A Data Acquisition Board (DAQ), the DAQLite produced by Eagle Technology, was used to acquire the power consumption data. The DAQ was set to a sampling frequency of 350Hz, in order to produce an amount of data statistically relevant, but not prohibitive for subsequent computation.

Software Setup. In order to automate scenario execution in our experiments, a supporting software environment was developed, composed of two Android applications, a server side application and macro scripts, to be executed by the tool AutoHotKey3. The developed Android application allows enabling or disabling components, such as Bluetooth, GPS or Wi-Fi interface, in order to avoid bias during scenarios that do not use them. For our second experiment, another Android application has been developed to control the execution of gLCB, specifying an execution time and a usage profile. With this solution, we assessed how the execution of different profiles of the application affected the power consumption of the device. These applications communicate with a server machine, which is then connected to the DAQ via USB. The server application then launches a AutoHotKey script that performs the needed operations for data acquisition and logging.

D. Instrumentation and Experiment Design

The goal of data analysis is to apply appropriate statistical tests to reject the null hypothesis. As we expected, the collected power consumption values, for both smartphones, do not follow normal distribution. This was verified by means of the Shapiro-Wilk test, with a resulting p-value lower than 0.05. This is true for our first experiment, "Training tools" as well as for the second one, "gLCB". Thus, in order to verify our hypotheses, we used non-parametric versions of the Kruskal-Wallis and Wilcoxon-Mann-Whitney tests [8], to assess the statistical independence between the different scenarios and profiles evaluated during our experiments. Again, we will draw conclusions from our tests based on a significance level $\alpha = 0{:}05$, that is we accept a 5% risk of type I error – i.e. rejecting the null hypothesis when it is actually true.

E. Threats to validity

We will classify threats of experiment validity in two categories: internal threats, derived from our treatments and instrumentation, and external threats that regard the generalization of our work. A possible internal threat concerns the sampling frequency adopted by the DAQ, namely 350 Hz. We chose this frequency value for practical reasons, in order not to obtain a huge amount of data, which could not be computed in a reasonable time by our servers. However, this frequency, compared to the operational frequencies of the selected smartphones, could be seen as quite low. A more significant threat comes from the usage,

in some of our scenarios, of different communication networks, which are characterized by an unpredictable behavior. This behavior may add bias to our measurements, introducing high data variability. For example, as regards the cellular network, power consumption could be affected by the following mechanism: the base station to which the mobile device is connected detects the signal power, and if the SINR is below or above a specific threshold it may negotiate a signal power increase or reduction to the device antenna. Finally, although is it not possible to generalize our results, because we performed our experiments on two specific models of smartphones, it is however possible to consider them representatives of category of devices with similar specifications.

IV. RESULTS

A. Preliminary Data Analysis

We present in Tables I, II, III, IV the following descriptive statistics about collected data. Tables report in this order: median (milliWatts), mean (milliWatts), standard deviation (σ), and variation coefficient (the standard deviation divided by the mean). Tables II, III contain descriptive statistics for each scenario of our "Training Tools" experiment, while Tables IV, V contain descriptive statistics for each profile of our "gLCB" experiment.

TABLE I EXPERIMENT "TRAINING TOOLS": SCENARIOS STATISTICS – SAMSUNG GALAXY I7500

Scenario	Median (mW)	Mean (mW)	Std.Dev.	Var.Co
2G Standby	8.663	17.840	65.763	3.686
3G Standby	8.663	27.248	97.592	3.581
2G Call	658.618	746.447	371.118	0.497
3G Call	957.803	988.069	97.313	0.098
WiFi Download	628.724	646.604	61.403	0.094
2G Download	669.467	784.099	742.696	0.947
3G Download	955.175	947.515	181.155	0.191
GPS	450.189	484.753	79.748	0.164
Bluetooth Scan	251.526	273.960	78.018	0.284
CPU-Intensive	606.923	608.708	38.442	0.063
Mp3 Audio	324.720	374.971	142.073	0.378
Display	386.252	408.754	81.598	0.199

TABLE II EXPERIMENT "TRAINING TOOLS": SCENARIOS STATISTICS – SAMSUNG NEXUS S

Scenario	Median (mW)	Mean (mW)	Std.Dev.	Var.Co
2G Standby	8.663	26.830	55.572	2.071
3G Standby	8.663	18.958	48.708	2.569
2G Call	379.488	543.487	565.230	1.040
3G Call	846.688	878.850	126.708	0.144
WiFi Download	455.733	513.046	166.444	0.324
2G Download	605.874	722.422	854.086	1.182
3G Download	965.368	931.798	208.819	0.224
GPS	296.626	300.444	20.375	0.067
Bluetooth Scan	217.571	227.051	42.882	0.188
CPU-Intensive	886.552	877.747	54.055	0.061
Mp3 Audio	155.035	164.709	26.666	0.161
Display	598.734	708.075	177.169	0.250
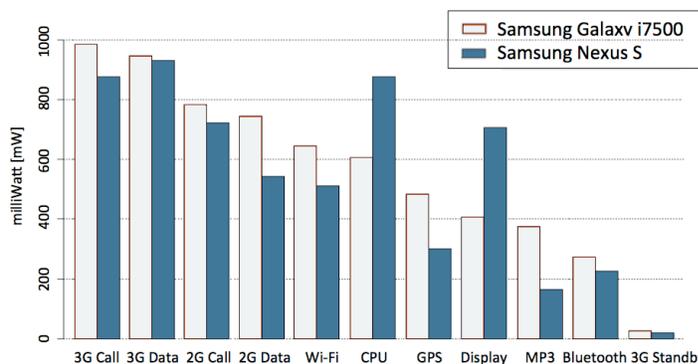
Figure 1 Instant Power Consumption (avg) comparison

TABLE III EXPERIMENT "GLCB": PROFILE STATISTICS - SAMSUNG GALAXY I7500

Scenario	Median (mW)	Mean (mW)	Std.Dev.	Var.Co
Very Low	396.474	521.565	213.243	0.408
Low	396.474	544.451	226.976	0.416
Normal	455.869	594.340	248.253	0.417
High	514.952	617.016	253.448	0.410

TABLE IV EXPERIMENT "GLCB": PROFILE STATISTICS - SAMSUNG NEXUS S

Scenario	Median (mW)	Mean (mW)	Std.Dev.	Var.Co
Very Low	420.271	536.863	225.055	0.419
Low	435.116	556.254	244.286	0.439
Normal	690.392	834.233	315.046	0.377
High	808.711	876.160	334.903	0.382

B. Hypothesis Testing

In this section we provide the results of hypothesis testing for our research questions. All p-values have been verified to be lower than the chosen significance level $\alpha = 0{:}05$.

- RQ 1.1: *Do usage scenarios have the same power consumption?*
  - $H_{0,1} : P_{0,1} \neq P_{1,1} \neq ... \neq P_{10,1}$
    Our values range from an average of 17.8 mW for Scenario S0 to an average of 988 mW for Scenario S2. The Kruskal-Wallis test for the hypothesis resulted in a p-value lower than 2.2e-16. Thus, we reject the null hypothesis.
  - $H_{0,2} : P_{0,2} \neq P_{1,2} = ... \neq P_{10,2}$
    Our values range from an average of 18 mW for Scenario S0 to an average of 931.8 mW for Scenario S5. The Kruskal-Wallis test for the hypothesis resulted in a p-value lower than 2.2e-16. Thus, we reject the null hypothesis.
- *RQ 1.2: Is the energy consumption the same among the devices?*
  - H_0: $P_{i,1} \neq P_{i,2}$, $i \in [0, 10]$
    The Mann-Whitney test resulted in a p-value lower than 0.001 for each scenario. Thus, we reject the null hypothesis.
- RQ 2.1: *Does gLCB cause a variation of the devices power consumption?*
  - $H_{1,1}$: $P_{1 \text{ with gLCB}} \neq P_{1 \text{ without gLCB}}$
    The Mann-Whitney test resulted in a p-value lower than 2.4e-09 for each profile compared to the standby consumption. Thus, we reject the null hypothesis.
  - $H_{1,2}$: $P_{2 \text{ with gLCB}} \neq P_{2 \text{ without gLCB}}$
    The Mann-Whitney test resulted in a p-value lower than 1.5e-09 for each profile compared to the standby consumption. Thus, we reject the null hypothesis.
- *RQ 2.2: Are there statistical differences between different user profiles?*
  - $H_{2,1}$: $P_{1 \text{ high}} \neq P_{1 \text{ normal}} \neq P_{1 \text{ low}} \neq P_{1 \text{ verylow}}$
    Our values range from an average of 521.5 mW for Very Low profile to an average of 617 mW for High profile. The Kruskal-Wallis test for the hypothesis resulted in a p-value lower than 1.146e-15. Thus, we reject the null hypothesis.
  - $H_{2,2}$: $P_{2 \text{ high}} \neq P_{2 \text{ normal}} \neq P_{2 \text{ low}} \neq P_{2 \text{ verylow}}$
    Our values range from an average of 536.8 mW for Very Low profile to an average of 876 mW for High profile. The Kruskal-Wallis test for the hypothesis resulted in a p-value lower than 3.433e-15. Thus, we reject the null hypothesis.
- *RQ 2.3: Are there statistical differences between the behaviors of gLCB in different devices?*
  - $H_{2,1}$: $P_{1 \text{ high}} \neq P_{1 \text{ normal}} \neq P_{1 \text{ low}} \neq P_{1 \text{ verylow}}$
    Our values range from an average of 521.5 mW
    The Mann-Whitney test resulted in a p-value lower than 2e-10 for each profile. Thus, we reject the null hypothesis.

The bar plot in Figure 1 shows the average power consumption values in mW for each scenario, on both smartphone models.

As we expected, we may notice that the standby values are the lowest, below 27 mW in the worst case. From the graph, it is evident how the most recent smartphone, the Samsung Nexus S, consumes a significantly lower amount of power in each scenario, exception given by the CPU Intensive and the Active Display scenarios. The percentage variations between the two smartphones, reported in Table VI, spread from a minimum -56,08% in the Mp3 Audio scenario, to a +73,23% in the Active Display scenario.

As regards our gLCB experiment, the bar plot in Figure 2 shows the average power consumption values in mW for each profile, on both smartphone models. It is possible to notice that, between the two, the most recent smartphone consumes less in verylow and low profiles.

V. DISCUSSION

A. Experiment 1: "Training Tools"

From the results obtained from our first experiment, we can conclude that the most power consuming user activities, among the ones we selected, on both the smartphone models used, are those that use the radio module, namely the phone
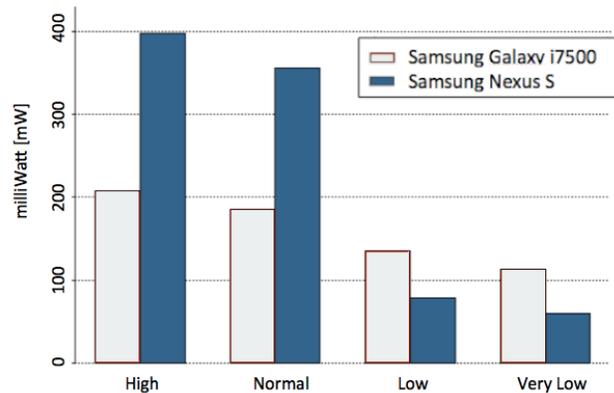
Figure 2 Instant Power Consumption (avg) of gLCB energy profiles

TABLE V COMPARISON IN MILLIWATTS OF SCENARIOS POWER CONSUMPTION OF THE TWO SMARTPHONES

Scenario	Galaxy	Nexus S	Galaxy VS Nexus S
2G Display	408.750	708.080	+73,23%(299.330 mW)
CPU Intensive	608.710	877.750	+44,19% (269.040 mW)
3G Download	947.510	931.800	-1,65% (15.710 mW)
2G Download	784.100	722.420	-7,86% (61.680 mW)
3G Call	988.060	878.850	-11,05% (109.210 mW)
Bluetooth Scan	273.960	227.050	-17,12% (46.910 mW)
Wi-Fi	646.600	513.050	-20,65% (133.550 mW)
2G Call	746.450	543.490	-27,19% (202.960 mW)
GPS	484.750	300.450	-38,02% (184.300 mW)
Mp3 Audio	374.980	164.700	-56,08% (210.280 mW)

TABLE VI COMPARISON IN MILLIWATTS OF PROFILE POWER CONSUMPTION (WITHOUT DISPLAY OVERHEAD)

Profile	Galaxy	Nexus S	Galaxy VS Nexus S
Very Low	112.811	59.185	-47,54 % (53.626 mW)
Low	135.697	78.576	-42,09 % (57.121 mW)
Normal	185.586	356.555	+92,12 % (170.969 mW)
High	208.262	398.482	+91,34 % (190.22 mW)

call and data transfer on both 2G (EDGE/GPRS) and 3G (UMTS) networks. This finding is coherent with [2], in which the most power-consuming scenario is indeed the phone call. Also in [6], examining some of the network related scenarios, it emerges that the battery lasts longer in all cases if the WiFi network interface is used, rather than 2G or 3G. From our results, it is worth to notice how 3G networks causes a sensible increase in power consumption with respect to 2G, in both voice and data communications.

Moreover, as regards the power consumption difference between the two smartphone models, we notice that the most recent model has in general a lower power consumption, the only exception being the CPU-intensive and active display scenarios: this can be justified, considering the increase in CPU frequency (1 GHz compared to 528 MHz) and in display dimensions (4" compared to 3.2"), which characterize the most recent model.

B. Experiment 2: "gLCB"

As regards our gLCB experiment, it is immediate to notice that in every profile the Samsung Nexus S consumes a higher amount of power. This is likely because during this experiment, the display of the smartphone was active, in order to verify the correct execution of the application, and the switching between profiles and execution sessions. This was done on purpose, in order not to introduce random bias due to occasional checking of the application behavior. Instead, we may subtract the display overhead from the power consumption values, and this is valid because we know that, from the previous experiment, the display consumption is characterized by low variance and dispersion values. The recalculated values, without the display overhead, are shown in Table VII. It is interesting to notice how, from these results, emerges that the Nexus S has actually lower power consumption values than the Galaxy in profiles verylow and low, namely 47,5% and 42,1% respectively. The other profiles show significantly higher power consumption. Given that the step-up from low to normal profile is characterized by the activation of WiFi, Bluetooth and GPS components, during normal and high profiles a higher computational load is expected. Thus, we may conclude that the power consumption increase is due to the CPU activating more frequently than in the other two profiles, also because we know, from the results of the previous experiment, that the CPU has a higher impact on the Nexus S smartphone. These results show that the impact of the gLCB application in terms of power consumption gradually reduces, by adopting lower energy profiles, on both smartphones.

VI. CONCLUSIONS AND FUTURE WORKS

From the analysis of the results provided by our experiments, we can conclude that the most recent device, in terms of OS and hardware components, shows significantly lower power consumptions than the least recent one, except for the CPU-intensive and active display cases. We showed that differences of energy-awareness, based upon a complete knowledge on execution profiles of the same application, can significantly affect the power consumption of a device.

This finding shows that energy-aware software applications can improve the energy efficiency of mobile devices, while providing the same functionalities.

As regards future works, it would be interesting to profile the energy consumption of other usage scenarios, for example those who require a higher interaction between the user and the device. Moreover, because of our experiment design, the smartphones were constrained to a single physical location; it could be interesting profiling the power consumption of a moving user, in order to get closer to the real case and evaluate with more precision the contribute of the subsequent handoffs between different cells in mobile networks. Another interesting point of view could be analyzing the power consumption of different generations of smartphones running the same version of the Android OS, in order to isolate the only impact of hardware changes.

REFERENCES

[1] A. Berl, E. Gelenbe, M. Di Girolamo, G. Giuliani, H. De Meer, M. Q. Dang, and K. Pentikousis, "Energy-efficient cloud computing," The Computer Journal, vol. 53, no. 7, pp. 1045–1051, 2009.

[2] A. Carroll and G. Heiser, "An analysis of power consumption In a smartphone," in Usenix technical conference, Boston, MA, USA,, pp. 1–14, 2010.

[3] L. Zhang, B. Tiwana, Z. Qian, Z. Wang, R. P. Dick, Z. M. Mao, and L. Yang, "Accurate online power estimation and automatic battery behavior based power model generation for smartphones," in 8th IEEE International conference on HW/SW codesign and system synthesis, pp. 105–114, 2010.

[4] Y.-F. Chung, C.-Y. Lin, and C.-T. King, "Aneprof: Energy profiling for android java virtual machine and applications," Parallel and Distributed Systems, International Conference on, vol. 0, pp. 372–379, 2011.

[5] F. Ding, F. Xia, W. Zhang, X. Zhao, and C. Ma, "Monitoring energy consumption of smartphones," CoRR, vol. abs/1201.0218, pp. 610-613, 2012.

[6] R. Palit, R. Arya, K. Naik, and A. Singh, "Selection and execution of user level test cases for energy cost evaluation of smartphones," in Proceedings of the 6th International Workshop on Automation of Software Test, pp. 84–90, 2011.

[7] G. Procaccianti, L. Ardito, A. Vetro', and M. Morisio, "Profiling power consumption on desktop computer systems," LECTURE NOTES IN COMPUTER SCIENCE, vol. 6868, pp. 110–123, 2011.

[8] L. Sachs, "Applied Statistics--A Handbook of Techniques", Springer-Verlag, 1984.

[9] L.Ardito, M.Torchiano, M. Marengo, P. Falcarin, gLCB: An Energy Aware Context Broker, Sustainable Computing: Informatics andSystems, ISSN 2210-5379, pp. 1-9, 2013.

[10] R. Van Solingen and E. Berghout, Goal/Question/Metric Method. McGraw-Hill Inc.,US, January 1999.

Dynamic Classification of Repetitive Jobs In Linux For Energy-Aware Scheduling: A Feasibility Study

Shane Case, Kanad Ghose
SUNY Binghamton
Binghamton, NY, USA
{shane, ghose}@cs.binghamton.edu

Abstract—The workload offered to a typical server consists of many repeated tasks. We present and evaluate a feasibility study that shows how repetitive server workloads can be exploited to enhance the server and CPU energy savings realized by state-of-the-art linux power governors. To minimize dramatic modifications to the web server and the core kernel scheduler we exploit the forensic logging capabilities of the Apache server to collect workload specific information and to schedule requests. We use a daemon to collect the classification statistics and control the dynamic voltage frequency scaling (DVFS) setting of the kernel to batch schedule requests with similar characteristics and thus amortize the energy and performance overhead of making changes to the DVFS settings. Our experimental results show that an energy savings of up to ten percent can be realized for the server on the workloads generated by the SPECweb2005 benchmarks.

Keywords-Energy Conservation; Power Savings; Process Management;Web Server; Performance Experiment; Scheduling; Measurement, Power Measurements, Energy Reduction

I. Introduction

Data centers are an essential part of a modern nation's infrastructure. Historically, the US has led the rest of the world in the development, deployment and widespread use of data centers in supporting a plethora of services and activities. A 2007 report places the energy expenditures associated with US data centers in 2006 at upwards of 1.5% of the total electricity expenditures of the nation [1]. These figures are likely to grow significantly as data centers are deployed widely to support a growing and wider variety of cyberservices. The hardware devices used in a data center continue to be more energy efficient but because of the annual increase in utility costs (8% to 11% annually), the cooling costs remain relatively stagnant [2]. This situation is therefore a crisis in the making and has the potential of impeding the growth rate of data centers and their increasing use in everyday life for the betterment of humanity.

The primary technique used in Linux servers is a power-performance governor that exercises the *DVFS* logic within the processor. At a low level of core utilization, the CPU clock frequency can be slowed down and the core supply voltage is simultaneously reduced. Just reducing the clock frequency does not result in a savings of the total energy consumption of the CPU, as the application still needs the same number of clock cycles to complete. Thus, the energy

expended per clock cycle is lowered by reducing the core supply voltage. The *DVFS* mechanism in a modern processor has a time overhead as well as energy overhead, as the CPUs voltage and frequency cannot be changed instantaneously.

The technique presented in this paper examines the feasibility of classifying repetitive server jobs into classes (two such classes are used in the present study), with each class requiring a specific *DVFS* setting. One would expect such repetitive jobs to be abundant in a typical server workload: the same scripts or transactions are often executed repeatedly. For example, in a server supporting banking transactions, the same scripts are executed on queries that determine account balances, for queries that post a charge to an account etc.

In the proposed technique, as new server jobs come in, they are classified into one of the two possible classes based on their observed execution characteristics. An already classified job, identified using the address of the executable, is batch scheduled on a core that is set at the matching *DVFS* step, to amortize the overhead of changing the *DVFS* settings. Appropriate timeout mechanisms are used to avoid undue prolongation of the service times for jobs. Energy savings of 5% to 10% over Linux on-demand governor are realized in running the SPECweb benchmarks in a prototype implementation running the Apache server that supports job classification (in conjunction with a daemon) and passes on the classification to the kernel. This 5% to 10% energy savings at the level of an individual server translates to an energy savings of 10% to 20% for an entire data center, where the utility costs of operating the cooling equipment equals the utility costs of operating the IT equipment [1] [3] [4].

The rest of the paper describes related work in Section 2, some background information on power governors and the Apache web server in Section 3. In Section 4, we describe the proposed scheme and some relevant implementation details, Section 5 presents the experimental setup and the measured results. Section 6 represents our conclusions.

II. Related Work

There is a plethora of work that has examined data center job scheduling strategies in general using simulators, synthetic job characteristics or actual server workload

trace data. Most of this body of work, exemplified by [5] focuses on scheduling jobs to servers. We address the problem of scheduling jobs in an energy-efficient manner within a server. However, very few server-local energy-aware scheduling techniques have targeted Linux or have actually described real implementations, validated with actual power measurements, and these are the work that we focus on in this section. In particular, these techniques do not explicitly exploit the characteristics of repetitive workloads.

Another way of managing the energy consumption of jobs allocated to a server is to limit the energy expended by the virtual machines running on a physical host. For example, the work of [6] describes an approach for allocating energy budgets to virtual machines. Such VM energy budgets are not easy to implement, as energy expended by a VM is not easy to track and control; energy dissipation in many related components are ignored in simplifications that are used. Allocation of jobs is also studied in [7], but this study focuses primarily on performance increases in a cluster environment, not on energy savings for a single server.

In general, emerging solutions have a number of potential limitations:

- The energy and performance overhead associated with job rescheduling and VM management and server-local scheduling overhead are ignored. The communication infrastructures within a data center are heavily utilized and are prone to congestion, resulting in significant added energy dissipation if jobs are rescheduled.
- A simple rescheduling of the jobs may not make the most energy-efficient use of the servers and racks - the operating configurations of such servers have to be continuously adapted to fit the characteristics of the workload.

The work of [8] classifies workload for long-lived connection-oriented services and shows, using trace data, how the information collected can be exploited for server provisioning and scheduling to save energy, minimizing the loss in live sessions in the process. A similar study is clearly needed to validate the repetitive nature of server workloads, as exploited in the proposed technique, but that study is beyond the scope of this paper. Another approach for exploiting workload characteristics in server provisioning and scheduling is presented in [9]. Global system monitoring facilities, such as IBM's Tivoli [10], [11] can collect energy-performance characteristics of jobs, including repetitive jobs, and can, in theory, use that information to schedule jobs globally and locally within a server. However, any automated scheduling facility that uses the collected information for local scheduling is missing at this point. The proposed technique not only collects the job's classification data but also integrates it with scheduling.

In [12], a framework called Koala for a model for predicting the energy and performance characteristics of a program for different *DVFS* settings and makes use of the predictions for energy-aware job scheduling. Koala requires a recalibration of the power model on any changes in the physical host configuration. In contrast, we use higher level job classification parameters to classify repetitive jobs and thus our approach does not depend on host-specific attributes. Koala can be extended to account for any *DVFS* overhead, in terms of performance and energy, repetitive jobs can in fact use the Koala framework as well.

A job classification scheme that characterizes map-reduce type jobs into compute-bound and I/O bound is introduced in [13], using a single instance of a map task and a single instance of a reduce task for classifying the set of map and reduce tasks that make up a single workload. However, results, experimental or otherwise nor any quantitative analysis is presented to assess the benefits of the proposed scheduling strategy.

In [14], the authors present a methodology for characterizing and classifying known background workload for Google's back end server in terms of their CPU and memory demand. The authors of [14] also describe very briefly the use of this classification in background job scheduling. The integration of the scheduling technique into an existing framework and its benefits are unclear.

III. BACKGROUND

A. CPU Frequency Governor

In the Linux Kernel, the most commonly used CPU frequency governor is the *Ondemand* governor. This governor gradually increases or decreases the frequency according to the current system load. The system load is measured from the scheduler, not the "loadavg" present in /proc. System load is sampled at 1000 times the frequency latency change value. Scaling is performed when the system load exceeds 95 percent and the CPU frequency is changed to the highest value. When system load decreases, the frequency is gradually stepped down.

B. The Apache Web Server

The main focus of this study will be on the pre-forking architecture of the Apache Web server [15]. Our main reason for focusing on the Apache web server are as follows. First, using the Apache web server's request scheduling capabilities and the ability to use its forensic logging mechanism to capture request specific statistics, removes the need to design any modified request scheduling and statistics collection features. Second, our approach permits us to accomadate independent changes made to both Apache and the core kernel very easily.

The performing architecture consists of the Apache parent process creating multiple processes called children to handle the incoming requests on the web server. Upon startup, the parent process will fork off a configurable number of children that will handle tasks assigned to them by the parent
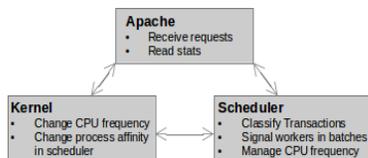
Figure 1.   Software Components in the Proposed Technique

process. When an HTTP request is received, each element of the requested document is a transaction from the view of the web server. If the ForensicLog module is enabled, each transaction will be double logged by the web server, once upon reception of the request and once upon its completion.

IV. THE PROPOSED TECHNIQUE

In order to exploit the repetitive nature of transactions serviced by the web server, two components are required. The first component is a modified ForensicLog module that is loaded into Apache web server at runtime, and the second is a system daemon that will process the data gathered by the logging mechanism.

The system daemon, or power manager must be an extremely lightweight process while executing in order to avoid skewing the statistics of the hardware that is being monitored, and consuming overhead power. The overhead of any type of busy waiting, such as spin locks or sleeping, is not tolerable, as a process utilizing either of these will consume more power in busy waiting. Busy waiting does not render the CPU as idle, and will show up as CPU usage, which can affect the decision of the power manager to enter into a sleep state. Therefore, the synchronization method used to communicate between Apache web server and our power manager will be signals. The benefit of signals is twofold, as signals are a lightweight inter-process communication method, and adding signals to the Apache interface will require very little code change. The bulk of the code change will be required in the power manager itself, by simply adding a signal mask to the power management process. The power manager will then block until a signal is received from Apache.

There will only be two signals required to achieve our goal of communication. In this case the signals SIGUSR1 and SIGUSR2 will be used. SIGUSR1 will serve as the notification that a transaction has been received by Apache and processing will begin. SIGUSR2 will serve as notification that the transaction has been completed. The power manager can also read an integer value that is passed along with the signal which will serve to identify the current entry in the shared memory segment that it will be profiling. On the power manager side, SIGUSR1 will be the notification to begin profiling and search our statistic table to see if the occurring transaction has already been previously profiled. If the transaction has already been profiled, profiling can be

terminated, squashing any statistics that have been gathered. The power manager can then apply the appropriate power profile, according to the transaction categorization in the statistic table. If the transaction has not been previously profiled, the counters have already begun before the search was conducted, and can then be stopped when SIGUSR2 is received. Once the statistics have been gathered, the numbers must be analyzed and a category assigned to that transaction before the transaction may be added to the statistic table.

A. Job Classification using the ForensicLog

When the web server initially receives a transaction, an entry is made in the ForensicLog via the *log_before()* function in ForensicLog. It is at this point that we modify the Log mechanism to not only continue its original logging duties, but to classify the task. At the time of the log entry, the parent Apache process has received the header of the server request and delegated processing to the active listening child. The log entry is made in the log file by the child process that is processing the request. It is at this point the ForensicLog is modified to monitor the current PID, which is the active child.

The statistics that are monitored for the purpose of system resource usage discovery are found in /proc/PID/schedstat, where PID is the currently running child's PIDİn these statistics, the CPU time spent and the I/O wait time spent is discovered about the currently running transaction. These metrics are reported on a per PID basis by the kernel's CFS scheduler. It is these two metrics that are used to profile a particular transaction. These metrics will be stored in a shared memory segment for communication to the classification daemon. This memory segment is locked using a mutex while the child process is acquiring the data for its transaction. When the metrics are gathered, the shared memory segment is unlocked, and the classification daemon is sent a *SIGUSR1* signal containing the index of the transaction of the signaling child process in the statistics table (Fig. 2). The child process will then enter a sleep state for a short period of time, via *sigtimedwait()*. The child sets its maximum timeout value to 2 seconds. The reason for this maximum timeout value is explained in the following section. Experiments were performed on different values and showed that the lower the timeout value, the lower the power savings percentage.

Upon either expiration of either the child's sigtimedwait() timeout value, or scheduling of the child by the classification daemon, the child will then process its assigned transaction. Upon completion of its transaction, the child will return to the ForensicLog module to log the transaction completion via the *log_after()* function. This function is altered to repeat what was done in the *log_begin()* function, except it will calculate its actual system resource metrics by using the values logged at the beginning of the transaction as a zero value. The CSD is again signalled, this time using a

SIGUSR2 value. The child process is then returned to the idle wait queue by the internal Apache scheduling mechanism.

B. Classification and Scheduling Daemon (CSD)

The CSD serves as the scheduler for the Web Server by scheduling child processes with a similar workload back to back. Their purposes of scheduling the children are to execute transactions of similar workload in batches, rather than as the transactions are received. Upon startup of the CSD, it will change the system power management policy to be in its lowest active power state.

The CSD is signalled upon every transaction beginning and ending. When a transaction is encountered that is not in the statistics table, the transaction is immediately signalled to complete. The signal that is received from the Apache child is the same signal that is returned. An entry for the unknown entry can then be established upon transaction completion.

The power manager will implement two work queues for transactions that have already been profiled. The queue will consist of PIDs to be signalled for transaction completion. One queue will be for transactions that have been categorized as CPU bound, and another queue will be for I/O bound transactions. The method used to implement these queues will require that a transaction be looked up in the statistics table to find its categorization. When a transaction has been found to be in the table, then that httpd worker child will be placed in a blocked state. The PID of the child worker will be entered into respective queue based on its classification type. While the child transaction is blocked, the power state of the system will still be at its lowest point. The queues will be emptied when a user defined number of transactions have entered a queue. In the worst case, a queue will not exceed the interval for an extended period of time, this is where the time-out value on the child waiting for a signal to continue comes in.

The queues containing the PIDs of blocked child processes are emptied when they reach a certain number of PIDs contained in each queue. This value by default was set at 3, this is a configurable number. Experiments performed with different values as a queue empty value showed that a smaller number would decrease the amount of power savings.

A different power scheme is employed based on the queue that is emptied. When the queue containing tasks primarily performing I/O based tasks, the CPU frequency is set to its lowest value. There is no loss in performance here, as the processor is making requests to the I/O subsystem and waiting for data to be returned. When the queue containing primarily CPU bound transactions is emptied, the CPU frequency is stepped up based on how many transactions are in the queue. If after initially emptying the queue, the queue becomes full again before the frequency is lowered, the frequency is stepped up again. This frequency stepping value is hardware dependent.

C. CPU Frequency Stepping

CPU frequency stepping is performed by the CSD using ioctls. To enable this ability, a new CPU frequency governor was created. The frequency governor is almost identical to the existing userspace frequency governor. The governor used by the CSD is implemented as a device driver that has ioctl() ability. The ioctls passed by the CSD contain the CPU number and a flag of 1 or 0, 1 being to increase the frequency of the specified CPU, 0 being to decrease the frequency. Individual CPUs cannot change their clock frequency in the hardware in our web server. To exploit CPU stepping, we migrated child processes to different CPUs by setting their scheduler affinity in the kernel scheduler.

In addition to handling the scheduling of the Apache child processes and the CPU frequency, the CSD will manage the CPU(s) that a child process will execute on. When the hardware that a Web Server is running on has multiple CPU sockets, the sockets can be configured to handle a specific task type. This hardware setup allows a CPU bound task to be scheduled on one socket and an I/O bound task to be scheduled on another socket. We will call this process *socket pinning*. In this case, each socket, can have different managed CPU frequencies.

V. RESULTS

We compare the results of our proposed Apache scheduler with an unmodified Apache Web Server. We use SPECweb2005, which contains three workload Banking, E-commerce, and Support, to compare the energy dissipation of the base web server compared to a web server with our modified scheduler. All of the SPECweb workloads consist of executing PHP scripts to perform specific transactions based on the workload. The data that is manipulated, whether it be text or images, is generated randomly based upon the number of concurrent connections. For our experiments, the number of clients is the same across all workloads. We used 5 clients, each generating 150 concurrent connections to place the web server under a load.

A. SPECweb Workloads

1) Banking: The banking workload of the SPECweb2005 benchmark is designed to simulate an online banking website. This workload primarily uses SSL connections to perform tasks. The makeup of the requests made on the server consist of images of checks and data to be entered into a bank ledger.

2) Ecommerce: The ecommerce workload is designed to simulate an online storefront, specifically a customizable computer system storefront. The workload will consist of a mix of SSL and non-SSL connections, due to the nature of the workload. Clients will benchmark the web server based on three phases, browsing the website, customizing a product, and ordering a product.

TABLE I
OVERALL SERVER ENERGY CONSUMPTIONS, TRANSACTIONS COMPLETED, AND ENERGY PER TRANSACTION

Power Management Technique	Parameter	Banking	Ecommerce	Support
Performance	Energy (J)	200285	193460	187830
	Total Transactions	40961	25285	26447
	Energy(J) per Transaction	4.889651132	7.651176587	7.102128786
Ondemand	Energy (J)	189637	190813	187713
	Total Transactions	40855	25471	26845
	Energy (J) per Transaction	4.641708481	7.491382356	6.992475321
Proposed Technique	Energy (J)	180515	184252	186730
	Total Transactions	40840	25261	26474
	Energy (J) per Transaction	4.420053869	7.293931357	7.053335348

TABLE II
CPU CORE ENERGY CONSUMPTIONS, TRANSACTIONS COMPLETED, AND ENERGY PER TRANSACTION

Power Management Technique	Parameter	Banking	Ecommerce	Support
Performance	Energy (J)	60218.6	57146.3	50328.8
	Total Transactions	40961	25285	26447
	Energy(J) per Transaction	1.470144772	2.260087008	1.903006
Ondemand	Energy (J)	56176	55402.2	50917
	Total Transactions	40855	25471	26845
	Energy (J) per Transaction	1.375009179	2.175108947	1.896703
Proposed Technique	Energy (J)	52822.3	49218.7	49771.8
	Total Transactions	40840	25261	26474
	Energy (J) per Transaction	1.29339618	1.948406635	1.880026

3) Support: The support workload is designed to simulate an online vendor's support website, whether it be a computer system vendor or a software package vendor. The workload will consist of searching and browsing a website for available downloads. The client will then download a file from the web server.

B. Experimental Setup

The hardware in our experiments consists of dual quad core Xeon E5520 CPUs with a max clock frequency of 2.27 GHz and 12GB RAMThe power supply is an 850W, 80 plus certified unit, with four 12V rails. The network interface card is an Intel 82574L gigabit Ethernet card using the e1000e driver. For CPU core energy measurements, we used two Fluke Y8100 current probes to collect and log data using a National Instruments Data Acquisition Unit. The overall server power consumption was measured using a Summit Technologies PS2500 Data Logging AC power meter measuring the AC power consumption for the server.

C. Overall Energy Consumption Characteristics

Table I shows the overall energy consumption, the number of transactions completed, energy consumed per completed transaction, averaged across multiple runs for a 30 minute duration for the Banking, E-commerce, and Support components of the SPECweb2005 benchmark. It is seen from Table I that the number of transactions processed for each of these benchmark components is roughly same across all of the techniques used (the performance governor, the ondemand governor, and the proposed technique). The energy per completed transaction is therefore a good indicator of the energy savings achieved by the proposed scheme. For the banking benchmark the energy per completed transaction in the proposed scheme is 4.42 Joules vs. 4.64 Joules using the ondemand governor and 4.90 Joules for the performance governor. Our scheme thus realizes roughly 9.8% energy savings on a per transaction basis compared to the performance governor and a 4.74% savings over the ondemand governor. The energy savings per transactions for the ecommerce benchmark is somewhat lower and there is hardly any energy savings on a per transaction basis for the support benchmark component. The reasons for the lower energy savings, particularly for support stem from the fact that these two components have a significantly higher fraction of I/O compared to both of the other components and both governors appear to be doing a good job in managing the *DVFS* settings for the I/O intensive benchmarks.

D. CPU Energy Consumptions Characteristics

Table II shows the energy consumption characteristics for the CPU cores for the three benchmark components of the SPECweb2005 benchmarks for each of the three power management schemes studied. The core power consumptions and the server power consumptions pretty much track each other but the percentage of energy savings realized by our scheme over the two governors are higher when one just looks at the CPU core energy consumption. This is simply because of energy consumptions elsewhere in the server and energy conversion inefficiencies within the power supply and voltage regulators on the motherboard. These other components dissipate energy none proportionately with the CPU cores.
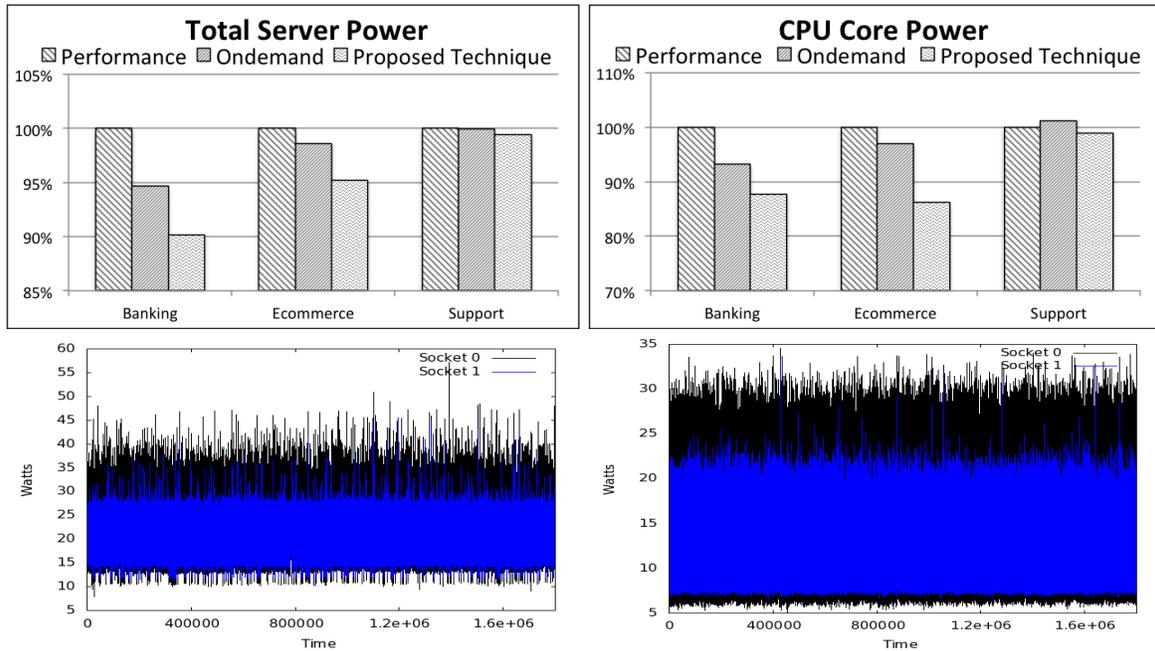
Figure 2.   Reduction in Energy Consumption (a) AC Total Server Power (b) DC CPU Core Power (c) DC CPU Power Ondemand for SPECWeb2005 Banking (d) DC CPU Power for Proposed Technique for SPECWeb2005 Banking

E. Total Server Energy Consumption

In Fig. 2(a), the total energy consumption of the entire server is displayed. We expect these results to look similar to the results from the CPU Core Power experiments. There is some overhead involved with total box power however, in that the total box power is taking into account the power consumed by all components of the server plus any power lost in the power supply alone.

The left most bar is the performance governor, which is keeping the CPU frequency at it's highest setting, and therefore should consume the most power in each of the test cases. The banking benchmark displays a good mix of both I/O and CPU bound transactions, this is verified by the Ondemand governor being about midway between the performance governor and the scheduler scheme at 94.69% power of the performance governor. The CSD must keep both sockets active due to the types of transactions being performed being almost equally distributed between CPU bound and I/O bound.

The next workload, e-commerce, has a workload that executes more CPU bound transactions. The nature of the workload involves searching the product lines by the client process, and switching between SSL and non-SSL connections for checkout, which generates CPU bound transactions. This is shown as the amount of power savings between the performance governor and the ondemand governor is less than half the power savings shown between the ondemand governor and the scheduler. Since the amount of transactions

that are CPU bound is greater than I/O bound, the scheduler is setting a majority of the child processes CPU affinity values to one socket, leaving the remaining socket idle. In this case, batch processing can be exploited by maximizing the efficiency of the processor when it is placed in the high frequency, and leaving the CPU in a lower frequency for a longer period than the ondemand governor does.

The last workload, support, is mostly I/O bound transactions. In this workload, the clients are primarily downloading files from the web server and this is reflected in the total power, and in the power savings. When the workload executes I/O bound transactions, the CPU will spend most of its time waiting for I/O operations to complete. As the CPU spends most of it's time in a waiting state, the load on the CPU is minimal and therefore the power consumed is minimal. The difference between the performance governor and the ondemand governor is almost non existent at 0.06%. The scheduler will still see a small power savings due to socket pinning in this case, as the CPU bound socket will be left idle during I/O bound transactions, and since the frequency of the I/O bound socket is kept to a minimum, we observe a small amount of power savings.

F. CPU Core Power

Fig. 2(b) shows the percentage of CPU core energy savings realized by our scheme against the performance governor (base case) and the ondemand governor. The percentage of energy savings realized by our scheme against the two governors is higher compared to the percentage

savings realized on the total server energy consumption. This is again because other components within the server do not dissipate energy proportional to the CPU energy consumption. It is also worth noting that for the support benchmark the ondemand governor actually leads to more energy energy consumption compared to the performance governor as the performance governor does a better job at handling I/O intensive workloads.

In Fig. 2(c) and (d) the power consumption of the CPU is examined on a per socket basis for the banking workload using the ondemand governor in (c) and the proposed technique in (d). When the ondemand governor is used the power consumption can range from 10 W to 50 W, compared to the proposed technique's range of 6 W to 34 W. This shows that batch scheduling transactions is effective at balancing CPU usage and idle time, lowering the range of power dissiptaion of the CPU.

VI. CONCLUSION

The ondemand and performance governors used in the Linux kernel for managing the *DVFS* settings have been refined over the years and do a very good job at reducing the power consumptions of servers. The goal of this effort was to exploit the repetitive characteristics of server workloads and exploit these characteristics in managing the *DVFS* settings of the processing cores, going beyond what is realized by the state of the art governors. Our experimental results clearly demonstrate that the proposed techniques realize additional energy savings beyond what is realized by the governors, up to about 10% for the Banking component of the SPECweb2005 benchmark. The energy savings realized over the governors over the other two specweb benchmark components are somewhat lower because the existing governors do a very good job at managing the *DVFS* settings for long periods of I/O where the CPU remains fairly idle.

The approach described in this paper minimized the need for modifications to the Apache server and the core kernel scheduler by using classification information collected by the Apache server and returning it to its job scheduling component and by using a separate daemon to collect the classification information and exercise the *DVFS* interface in the core kernel. Our ongoing work will implement the classification and scheduling components entirely within the core kernel, to realize a higher level of energy savings.

REFERENCES

[1] "US Environmental Protection Agency Energy Star Program, Final Report Congress on Server and Data Center Energy Efficiency", Public Law 109-431, August 2007.

[2] J. G. Koomey, "Estimating Total Power Consumption By Servers in the U.S. and the World", Analytics Press. Feb. 2007.

[3] L. A. Barroso and U. Holzle, "The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines", Morgan-Claypool Publishers, 2009 (ISBN No. 9781598295566).

[4] J. Hamilton, "Where Does the Power Go In A Data Center?", keynote presentation at SIGMETRICS/Performance 2009.

[5] Q. Tang, S. K. S. Gupta, and G. Varsamopoulos, "Energy-Efficient, Thermal-Aware Task Scheduling for Homogeneous, High Performance Computing Data Centers: A Cyber-Physical Approach", in IEEE Trans. On Parallel and Distributed Systems, vol. 19, no. 11), pp. 1458-1472, 2008.

[6] R. Nathuji and K. Schwan, "VPM Tokens: Virtual Machine-Aware Power Budgeting in Datacenters", in Proc. of the ACM/IEEE International Symposium on High Performance Distributed Computing (HPDC), pp. 119-128, 2008.

[7] Z. Liu, "Load Balancing and Task Partitioning Strategy for Mixed Nature of Tasks", 2011 International Conference on Computer and Software Modeling, pp. 48-52, 2011

[8] Chen et al, "Energy-aware server provisioning and load dispatching for connection-intensive internet services", in Proc. 5th USENIX Symposium on Networked Systems Design and Implementation, pp. 337-350, 2008.

[9] S. Govindan, J. Choi, B. Urgaonkar, A. Sivasubramaniam, and A. Baldini, "Statistical profiling-based techniques for effective power provisioning in data centers", in Proc. EuroSys '09, pp. 317-330, 2009.

[10] IBM Corporation, "IBM Tivoli Usage Accounting Manager V7.1 Handbook", IBM Redbook, March 2008.

[11] IBM Corporation, "Value Proposition for IBM Systems Director: Challenges of Operational Management for Enterprise Server Installations", IBM ITG Group, Mangement Brief (34 pages), Nov. 2008.

[12] D.C. Snowdon, E. Le Sueur, S. Petters, and G. Heiser, "Koala: a platform for OS-level power management", Proceedings of the 4th ACM European conference on Computer systems, pp. 289-302, 2009.

[13] P. Visalakshi and T.U. Karthik, "MapReduce Scheduler Using Classifiers for Heterogeneous Workloads", IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.4, pp.68-73, April 2011,.

[14] A. K. Mishra, J. Hellerstein, W. Cirne, and C. Das, "Towards Characterizing Cloud Backend Workloads: Insights from Google Compute Clusters", ACM SIGMETRICS Performance Evaluation Review on Industrial Research (SIGMETRICS PER, 2010), Volume 37 Issue 4, pp. 34-41, March 2010

[15] B. Grone, A. Knopfel, R. Kugel, and O. Schmidt, "The Apache Modeling Project", available at http://www.fmc-modeling.org/projects/apache 2008

An Integrated Framework for Power-Performance Analysis of Parallel Scientific Workloads

Sergio Barrachina, Maria Barreda, Sandra Catalán, Manuel F. Dolz,
Germán Fabregat, Rafael Mayo, Enrique S. Quintana-Ortí
Depto. de Ingeniería y Ciencia de Computadores
Universitat Jaume I
Castellón, Spain
Emails: {barrachi,mvaya,catalans,dolzm,fabregat,mayo,quintana}@uji.es

Abstract—**The path towards Exascale systems will require to energetically address power consumption of future high performance computing (HPC) workloads which, in turn, urges for a better understanding of power usage. In this paper we, present an evolved framework to trace and analyze the power and energy consumption made by parallel scientific applications. The framework includes i) a flexible and extensible design that enables easy integration of different types of power measurement devices and addition of new functionality; ii) a new module that records information on processor states related to power consumption; and iii) an improved power measurement device to monitor internal direct current (DC) power consumption. This environment is thus revealed as a powerful yet easy-to-use tool to investigate and progress on the development of energy-efficient HPC applications.**

Keywords-power consumption; high performance computing; performance analyzers; scientific applications;

I. INTRODUCTION

Power consumption has traditionally been a strong constraint for mobile devices due to its impact on battery life. In recent years, the need to reduce energy expenses has also reached the market of desktop and server platforms, which are embracing "greener technologies"; and even in the HPC arena, the power wall is now recognized as a crucial challenge that the community will have to face [1], [2], [3]. Clear signs of this trend are varied, ranging from the energy efficiency regulatory requirements set by the US Environmental Protection Agency to the biannual elaboration of the Green500 list [4] and the ongoing standardization effort of this ranking.

While system power management (especially that related to the processor) has experienced considerable advances during this past period, application software has not benefited from the same degree of attention, in spite of the power harm that an ill-behaving software ingredient can infer. Indeed, tracing the use of power made by scientific applications and workloads is key to detecting energy bottlenecks and understanding power distribution. However, as of today, the number of fully integrated tools for this purpose is insufficient to satisfy a rapidly increasing demand. In this paper, we revisit the tools in [5] for power-performance analysis of parallel scientific applications, and introduce a

new framework featuring stronger integration and modular design as well as several significant extensions. In particular, the paper includes the following contributions:

- We present a new release of our package `pmlib` for power measurement, portray the users' view with a detailed example, and expose the modular implementation of this software.
- We describe a new appliance to `pmlib` which collects information on processor energy states, like the C-states and P-states, offering information complementary to that in the performance and power traces.
- We enlarge the number of standard and *ad hoc* power measurement devices that can interface with `pmlib`, and review the hardware implementation of one of our own powermeters.

These new additions increase the practical utility of the framework, while maintaining its accessibility.

The rest of the paper is structured as follows. In Section II, we provide a brief overview of the framework architecture and its use. In Section III we enumerate a collection of powermeters that can interact with our framework, describing in more detail the internals of our own power device. In Section IV we review `pmlib`, the pivotal component of the framework. Finally, the paper is closed with a discussion of related work, in Section V, and a list of concluding remarks and future work in Section VI.

II. OVERVIEW OF THE POWER ANALYSIS FRAMEWORK

Figure 1 offers a graphical representation of the framework for power-performance tracing and analysis. The starting point is a concurrent *scientific application*, instrumented with the `pmlib` software, that runs on a parallel *target platform* (e.g., a cluster, a multicore architecture, or a hybrid computer equipped with one or several graphics processor units (GPUs)) yielding a certain power consumption. Attached to the target platform there is one or several *powermeter devices* –either internal DC or external alternating current (AC)– that steadily sample power, sending this output to a *tracing server*. Calls from the application running on the target platform to the `pmlib` application programming interface (API), instruct the tracing server to
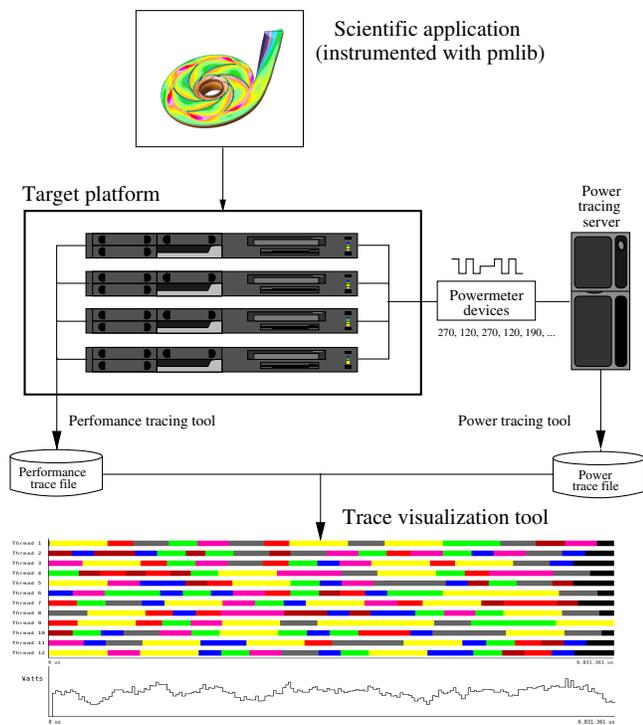
Figure 1. Collecting traces at runtime and visualization of power-performance data.

Figure 2. Single-node application system and sampling points for external and internal powermeters.

start/stop collecting the data captured by the powermeters, dump the samples in a given format into a disk file (*power trace*), query different properties of the powermeters, etc. Upon completion of the application's execution, the power trace can be inspected, optionally hand-in-hand with a performance trace, using some *visualization tool*. Our current setting allows a smooth integration of the framework power-related traces and the performance traces obtained with `Extrae`. The resulting combined traces can be visualized with `Paraver` [6]. Nevertheless, the modular design of the framework can easily accommodate other tracing tools like, e.g., TAU, Vampir, etc.

III. HARDWARE POWER SAMPLING DEVICES

The `pmlib` package interacts with a number of power sampling devices, including i) external commercial products, such as *APC 8653* Power Distribution Unit (PDU) and *WattsUp? Pro .Net*, which are directly attached to the wires that connect the electric socket to the computer Power Supply Unit (PSU), thus measuring external AC for the full platform; and our own internal DC powermeter designs, consisting of an appropriate choice of current transducers that produce data for ii) a commercial data acquisition system (DAS) from *National Instruments* (NI) and, alternatively, iii) our own designs that use a microcontroller to sample transducer data. These devices are described in more detail next, and the connection points are illustrated in Figure 2.

i) External AC powermeters. The APC 8653 PDU has 24 outlets and operates at a sampling rate of 1 Hz, employing the Simple Network Management Protocol (SNMP) to communicate with the tracing server via Ethernet. The WattsUp? Pro .Net also works at 1 Hz and returns samples to the server through an Universal Serial Bus (USB) 2.0 line.

ii) Powermeter using NI DAS. Our own measurement tools have been developed taking into account that they had to measure currents ranging from 1 to 15 A, without introducing significant voltage drops. The selected transducer was the *LEM HXS 20-NP* Hall effect current sensor. The device exhibits high accuracy and linearity, and a very low internal resistance, while being able to measure current in the required ranges.

A set of our designs include several channels with each one comprising a transducer that is connected to one of the power lines leaving from the PSU. Our final system is a modular design, based on stackable 8-channel components that share power and reference voltage, for a total of 32 current channels.

The DAS is composed of the NI9205 module and the NIcDAQ-9178 chassis. The module features 32 16-bit resolution analog-to-digital (AD) channels which can sample data at 7,000 Hz. In principle, the `LabView` software from NI runs in the tracing server, reading the data captured by the DAS from a USB 2.0 port in the chassis. For convenience, we have developed our own daemon/software to interact with the chassis, without the need of LabView, enabling a better integration of the device with `pmlib`.

iii) Microntroller-based powermeters. Our initial designs [5] featured 10 and 25 channels and a Peripheral Interface Controller (PIC) 18 microcontroller from *Microchip*, to perform AD conversion. Each channel consisted of the aforementioned HXS 20-NP transducer and a 10-bit resolution AD channel in the microcontroller. All the channels shared

Figure 3.   Diagram of the communication between client (running a scientific application) and the (pmlib) server.

a reference voltage of 2.5 V generated by the transducers. Data was sent to the host computer over an asynchronous RS232 port, and the sampling rate was therefore limited by the speed of the communications link (115,200 bauds in the selected microcontroller).
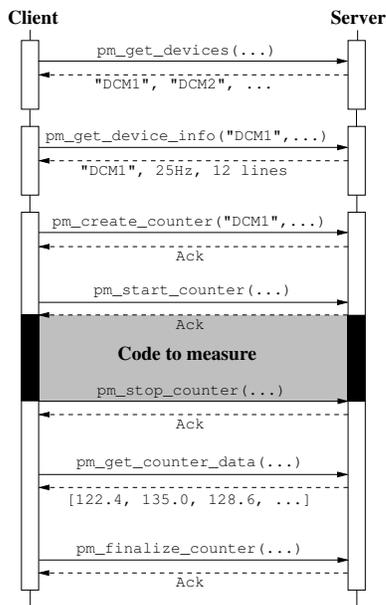
IV. SOFTWARE

The pmlib software package is developed and maintained by our research group to investigate power usage of HPC applications. The current implementation of this package provides an interface to utilize all the above-mentioned powermeters and a number of tracing tools. We next portray the interface of pmlib using a practical example (user's view), offer a few key implementation details (developer's view), and describe the functionality of a new module to gather information about power-related states of the processor cores. We close this section by illustrating the kind of information provided by the framework using a simple parallel application.

A. User's view

Power measurement is controlled from the application using a collection of routines that allows the user to query information on the power measurement units, create counters associated to a device where power data is stored, start/interrupt/continue/terminate power sampling, etc. All this information is managed by the pmlib server, which is in charge of obtaining these data from the devices and returning the appropriate answers, via the interface of the pmlib routines, to the invoking application (client). This client-server interaction is exposed in Figure 3.

```c
int main (int argc, char *argv[])
{
    server_t   server1,  server2;
    counter_t  counter1, counter2;
    line_t     lines1,   lines2;
    device_t   disp;
    char       **list;
    int        i, num_devices,
               freq1=0, freq2=0, aggr1=1, aggr2=1;
    // ... Some other variables...

    // Initializes the servers' structures
    pm_set_server("150.128.82.30", 6526, &server1);
    pm_set_server("127.0.0.1",     6526, &server2);

    // Query on #devices connected to server1,
    // and obtain handles. Then, output information,
    // e.g., for device[0]
    pm_get_devices(server1, &list, &num_devices);
    pm_get_device_info(server1, list[0], &disp);
    printf("Name: %s\n", disp.name);
    printf("Max freq: %d\n", disp.max_frecuency);
    printf("Number of lines: %d\n", disp.n_lines);

    // Selects the lines to measure
    pm_set_lines("0-11", &lines1);
    pm_set_lines("0-31", &lines2);

    // Creates a counter for powermeter DCMeter1
    pm_create_counter("DCMeter1", lines1, !aggr1,
                      freq1, server1, &counter1);

    // Creates a counter for C-states
    pm_create_counter("Cstates", lines2, !aggr2,
                      freq2, server2, &counter2);

    // Starts to collect samples: power, C-states
    pm_start_counter(&counter1);
    pm_start_counter(&counter2);
    // Sampled application code fragment
    dgemm( &transa, &transb, &m, &n, &k,
           &alpha,   &A[k*lda+i], &lda,
                     &B[j*ldb+k], &ldb,
           &beta,    &C[j*ldc+i], &ldc );
    // Stops to collect samples
    pm_stop_counter(&counter1);
    pm_stop_counter(&counter2);

    // ...  Some other nonsampled   ...
    // ... aplication code fragment ...

    // Continue to collect samples: only power
    pm_continue_counter(&counter1);
    // Sampled application code fragment
    dsyrk(&transa, &transb, &m, &n,
          &alpha,  &A[k*lda+i], &lda,
          &beta,   &C[i*ldc+i], &ldc);
    //Stops to collect samples
    pm_stop_counter(&counter1);

    // Dumps collected data onto memory
    pm_get_counter_data(&counter1);
    pm_get_counter_data(&counter2);

    // Prints power data in Paraver format
    pm_print_data_paraver("out.prv", counter1,
                          lines1, 0, "us");
    // Prints c-states data in Paraver format
    pm_print_data_paraver_cstates("cstates.prv",
                                  counter2, lines2,
                                  0, "us");

    //Finalizes the counters
    pm_finalize_counter(&counter1);
    pm_finalize_counter(&counter2);
    return 0;
}
```

Figure 4.   Example of use of pmlib.

Figure 4 displays a detailed example illustrating the use of `pmlib`. The code first declares the most important variables. Next, two server structures are initialized with their respective Internet Protocol (IP) addresses and the port that will be used for the communication with both servers. Here, the first server returns power samples, and it is located in a separated machine to avoid interfering with the parallel application. C-states are recorded using the second server, which is placed in the same machine where the parallel application runs, so that, it can query the files of this machine containing the requested data on C-states. The invocation to the function `pm_get_devices` establishes a communication with the server, to obtain a list with the names of connected powermeters. With this information, the call to `pm_get_device_info`, with one of the detected powermeters, returns more specific information on this device.

With the next two calls to `pm_set_lines`, we select the lines to measure (distinct powermeters may have different numbers of lines). Next, we also call function `pm_create_counter` twice, to create one counter associated with the `DCMeter1` powermeter and a second one that is bound to the C-states. The measurement is initiated and terminated from the application via routines `pm_start_counter` and `pm_stop_counter`, respectively. In this case, we measure the power and record the C-states during the execution of kernel `dgemm`. The sampling process is momentarily interrupted then, by invoking `pm_stop_counter`, and continued later, with `pm_continue_counter`, to record only power samples for kernel `dsyrk`. Finally, routine `pm_get_counter_data` saves the collected data onto the corresponding counter structure; this information is printed in one of the available formats (in the example, `Paraver` format); and the counters are destroyed using routine `pm_finalize_counter`.

B. Developer's view

The `pmlib` software is developed in Python and consists of two modules: the *settings* file and the *server*. Figure 5 depicts how the server works. The daemon starts by initially reading the settings file, which contains configuration information on the powermeters available in the system. Afterwards, a new thread is created per powermeter in order to manage and receive data from these devices. The server then creates as many counters (i.e., new thread instances) as required by the clients.

The main threaded classes implemented by the server are:

- `Device`. This class reads the data collected by a specific powermeter and stores all the active counters that measurement.
- `Counter`. This class manages all the operations performed on a counter. It is stored in the `Device` object
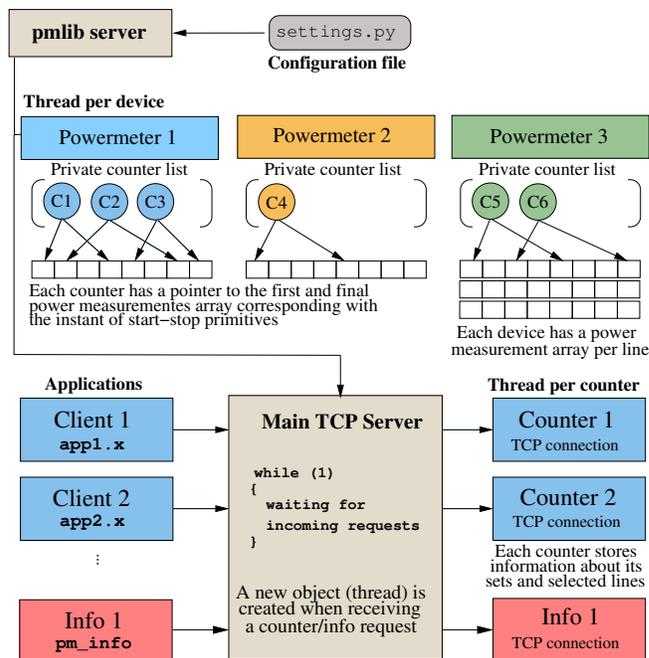
Figure 5. Internal workings of the `pmlib` server.

it is associated with, and contains the data acquired while the counter is running.

- `Info`. This class comprises information about the devices and their configuration.

Figure 5 shows that the server can receive two types of *requests*: a query on information about a device or an operation on a counter. In the first case, the server creates an `Info` object to obtain the required data from the settings file, and sends this back to the client.

If the operation is a request to create a counter, the server allocates a `Counter` object, which will manage all subsequent operations on it, and stores the structure in the appropriate `Device` object. After the creation of a counter, the client should invoke `pm_start_counter` to instruct the server to start recording samples, and `pm_stop_counter` to stop doing it. The client can also use `pm_continue_counter` to restart the recording process, and force the server to record samples from other fragments of the application code in the same counter, generating different `sets` of data. Finally, all collected data can be retrieved by invoking `pm_get_counter_data`.

C. A module to detect power-related states

Most current processors, from those designed for mobile devices to desktop and HPC servers, adhere now to the the Advanced Configuration and Power Interface (ACPI) specification [7], which defines an open standard for device configuration and power management from the operating system.

For our power monitoring purposes, the ACPI specification defines a series of *processor states* (or power modes),

collectively known as C-states, that are valid on a per-core basis. For example, the C-states available in one of our power-monitored platforms (a server with two Intel Xeon E5504 processors) are:

- **C0**. The core is in operating state (i.e., active or executing instructions).
- **C1**. The core is inactive, but can return to an executing state essentially instantaneously.
- **C3**. The core maintains all software-visible state, but may take longer to wake up.
- **C6**. The core does not need to keep its cache coherent, but maintains other state. Some architectures have variations on the C6 state that differ in how long it takes to wake the processor.

A core in state C0 can be in one of several performance states, referred to as P-states. These modes are architecture-dependent, though P0 always corresponds to the highest-performance state, with P1 to Pn being successively lower-performance modes. In practice, the P-states differ in the operation voltage-frequency pair, with P0 being always bind to the highest pair.

Our power framework obtains a trace of the C- and P-states of each core. For example, in order to obtain information on the C-states, a daemon integrated into the power framework accesses the file `/sys/devices/system/cpu/cpuX/cpuidle/stateY/time`, for each cpu `X` and state `Y`, with a user-configured frequency. The daemon reads values from this file corresponding to the total time (in microseconds) spent in a certain state. This value is then subtracted from the previous read, normalized, and stored together with a timestamp in a file with a user-selected format.

Note that the state-recording daemon necessarily has to run on the target application and, thus, it introduces a certain overhead (in terms of execution time as well as power consumption) that, depending on the software that is being monitored, can become nonnegligible. To avoid this effect, the user is advised to experimentally adjust the sampling frequency of this daemon with care.

Figure 6 offers a graphical example of the information that can be collected with our power-tracing framework, when combined with the performance tracer `Extrae` and the visualization tool `Paraver`. The view there corresponds to the execution of a synthetic parallel benchmark that randomly issues three types of computational kernels: `dgemm` (matrix-matrix product), `dtrsm` (triangular system solve), and `sleep`. The test was run using 8 threads on a platform equipped with two Intel Xeon E5504 cores at 2.00 GHz, with 24 GBytes of RAM. The performance trace in the top plot displays task activity per core; the second plot corresponds to the aggregated power dissipated by the mainboard of the machine, captured with the NI powermeter operating at 1 KHz; the C-states trace in the third plot represents the variations that cores experience between processor states

C0, C1, C3 and C6 (with a sampling frequency of 10 Hz). The final part reports the same information contained in the performance and C-states traces, but in numerical format.

V. Related Work

An excellent survey on hardware, software, and hybrid tools for power profiling is given in [8]. We next briefly review some of these efforts, in particular, those proposing solutions related to our framework.

PowerMon2 [9] is a hardware device which, coupled between the computer's PSU and mainboard, samples the power running through the DC lines, offering a basic software interface.

PowerPack [10] employs a commercial DC powermeter from NI, much like ours, connected to the lines leaving from the PSU. This package also performs a number of tests with the purpose of identifying which lines feed different components such as disks, memory, network interface controllers (NICs), processors, etc. This information is then offered to the user who can gain insights on where and how applications consume power. PowerPack exhibits a user-friendly interface, and targets applications running on single-node platforms, though PowerPack's information can be "manually" aggregated for parallel Message Passing Interface (MPI) applications.

HDTrace [11] is a package to trace and simulate the power-performance footprint of MPI programs on a cluster. This software supports MPICH2 and the parallel file system Parallel Virtual File System (PVFS). Recently, it has also been extended to identify power hot spots in these applications, using information from commercial AC powermeters.

Compared with these other efforts, we believe that our project provides a more complete framework with stronger integration and a better modular design.

VI. Conclusions

We have presented a power-tracing framework composed of internal/external powermeters, a power tracing modular package, power-related modules, etc., that is easily integrable with standard performance tracing and visualization tools. The framework offers highly useful information on power usage of scientific workloads running on a variety of parallel platforms, from MPI applications operating on a moderate-scale cluster to multi-threaded codes that execute on a multicore+GPU platform.

We are currently using this framework to develop more energy-efficient HPC linear algebra libraries, which leverage idle periods during the execution using dynamic frequency-voltage scaling and avoiding busy-waits. In the future, we plan to extend this work to other numerical codes and scientific applications in general, by integrating it into a practical runtime task-scheduler.
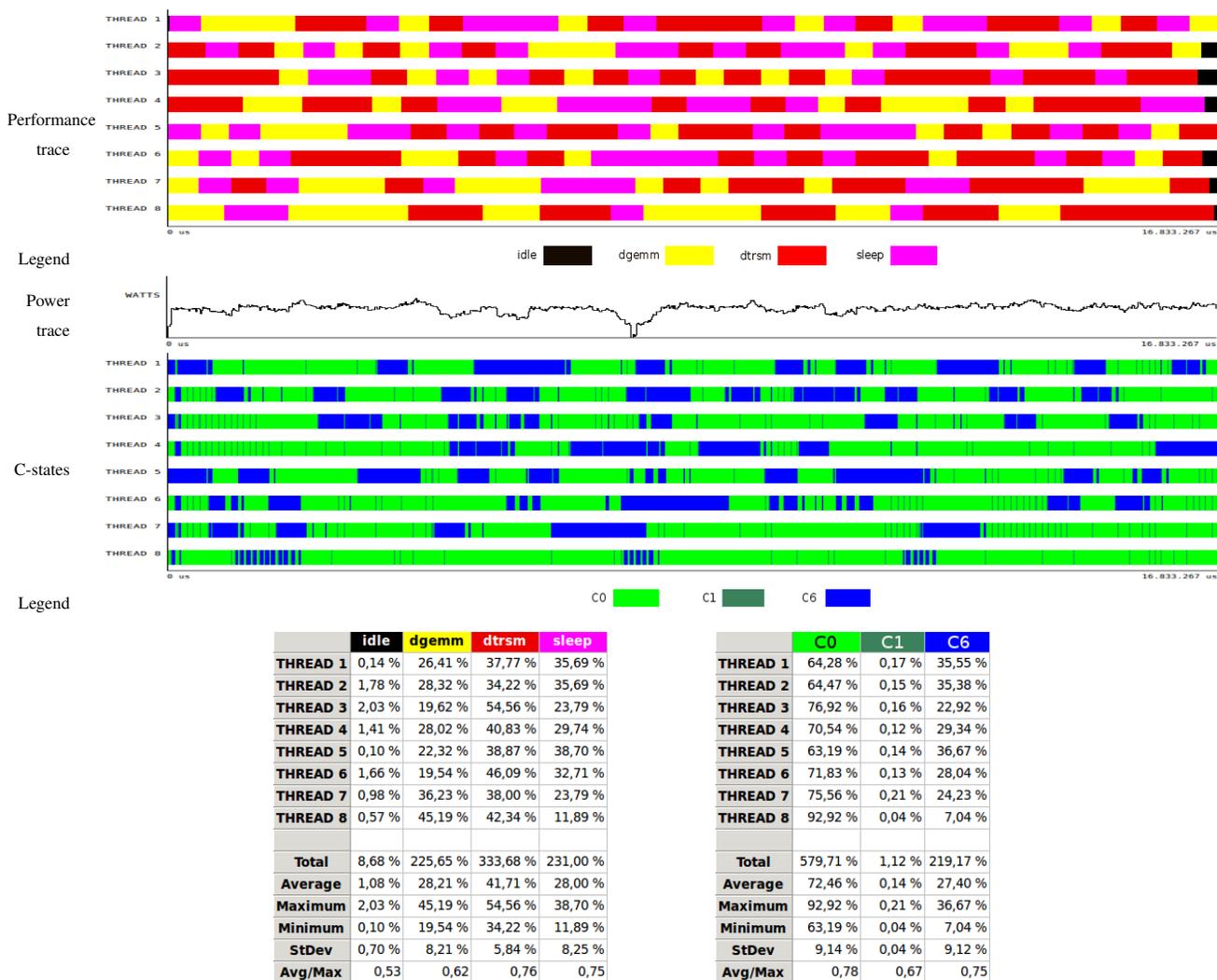
Figure 6. Example of performance and power traces captured by `Extrae` and the proposed power framework, visualized with `Paraver`.

REFERENCES

[1] J. Dongarra et al., "The international exascale software project roadmap," Int. J. of High Performance Computing & Applications, vol. 25, no. 1, 2011, pp. 3–60.

[2] M. D. et al., "The HiPEAC vision," http://www.hipeac.net/roadmap, [retrieved: January, 2013].

[3] W. Feng, X. Feng, and R. Ge, "Green supercomputing comes of age," IT Professional, vol. 10, no. 1, 2008, pp. 17 –23.

[4] "The Green500 list," http://www.green500.org, [retrieved: January, 2013].

[5] P. Alonso et al., "Tools for power and energy analysis of parallel scientific applications," in Proc. of 41st International Conference on Parallel Processing–ICPP, Sep. 2012, pp. 420–429.

[6] "Paraver: the flexible analysis tool," http://www.cepba.upc.es/paraver, [retrieved: January, 2013].

[7] HP Corp., Intel Corp., Microsoft Corp., Phoenix Tech. Ltd., Toshiba Corp., "Advanced Configuration and Power Interface Specification," 2010.

[8] S. Wang, H. Chen, and W. Shi, "SPAN: A software power analyzer for multicore computer systems," Sustainable Computing: Informatics and Systems, vol. 1, no. 1, 2011, pp. 23–34.

[9] D. Bedard, A. Porterfield, R. Fowler, and M. Y. Lim, "PowerMon 2: Fine-grained, integrated power measurement," RENCI, North Carolina, Tech. Rep. TR-09-04, 2009.

[10] R. Ge, X. Feng, S. Song, H. Chang, D. Li, and K. W. Cameron, "Powerpack: Energy profiling and analysis of high-performance systems and applications," IEEE Transactions on Parallel and Distributed Systems, vol. 99, 2009, pp. 658–671.

[11] J. Kunkel, "HDTrace-a tracing and simulation environment of application and system interaction," Dept. of Informatics, Sci. Comp., Universität Hamburg, Tech. Rep. 2, 2011.

Energy Feedback for Smart Grid Consumers: Lessons Learned from the Kukui Cup

Robert S. Brewer, Yongwen Xu, George E. Lee, Michelle Katchuck, Carleton A. Moore, Philip M. Johnson
Department of Information and Computer Sciences
University of Hawaii at Manoa
Honolulu, HI, USA
{rbrewer, yxu, gelee, katchuck, cmoore, johnson}@hawaii.edu

Abstract—**To achieve the full benefits of the Smart Grid, end users must become active participants in the energy ecosystem. This paper presents the Kukui Cup challenge, a serious game designed around the topic of energy conservation which incorporates a variety of energy feedback visualizations, a multifaceted serious game with online educational activities, and real-world activities such as workshops and excursions. We describe our experiences in developing energy feedback visualizations in the Kukui Cup based on in-lab evaluations and field studies in college residence halls. We learned that energy feedback systems should address these factors: they should be actionable, that domain knowledge must go hand in hand with energy feedback systems, and that this feedback must be "sticky" to lead to changes in behaviors and attitudes.**

Keywords-Serious games; energy feedback; energy; energy literacy; smart grid.

I. INTRODUCTION

The development of the Smart Grid and the two-way communication that it provides have enabled a variety of new customer-facing possibilities including real-time feedback on electricity usage, real-time pricing, and demand response. However, to make full use of this potential, end-users of the Smart Grid will need to be engaged about their electricity use, and become more energy literate. We believe that in addition to a Smart Grid, we need Smart Consumers.

In this context, we have developed the Kukui Cup Challenge, a serious game [1] (a game with additional goals beyond just entertainment) designed around the topic of energy. The Kukui Cup includes a variety of energy feedback visualizations [2] designed to inform and engage the players about their energy use. The Kukui Cup also includes a multifaceted online game with educational activities, and real-world activities such as workshops and excursions [3].

The Kukui Cup is designed to provide players with insight into how their behaviors affect energy consumption and production. Such behaviors occur on a spectrum, from the short-term, immediate impact behaviors such as turning off lights, to the longer-term, collective impact of behaviors such as considering the energy policies of political candidates when deciding how to vote. Creating a challenge that helps players understand energy from this wide scope sets the Kukui Cup apart from other similar "energy game" initiatives. It also

impacts on our understanding of effective feedback for smart grid customer-facing applications.

Based on our experiences designing and evaluating energy feedback in the Kukui Cup, we have three recommendations for designing energy feedback systems for smart grid consumers: they should be actionable, that domain knowledge must go hand in hand with feedback systems, and that this feedback must be "sticky" to lead to changes in behaviors and attitudes. This paper explores how we came to these conclusions, and what evidence we have collected that supports these conclusions.

In this paper, we first describe the Kukui Cup system, followed by an explanation of how energy goals and baselines are used in the Kukui Cup. With that foundation, we discuss our results from developing and deploying the Kukui Cup in the field over two years in the areas of designing energy feedback visualizations, the importance of energy literacy in understanding energy feedback, and our use of a serious game to encourage users to engage with the energy feedback information. Finally, we end with sections describing our plans for future work and conclusions.

II. THE KUKUI CUP

College residence hall energy competitions have become a widespread mechanism for engaging students in energy issues, with more than 160 taking place or being planned for the 2010–2011 academic year in North America [4]. Residence hall energy competitions are events where residence halls or floors within a residence hall compete to see which building will use the least energy over a period of time. The competitions tap into both the residents' competitive urges, and their interest in environmental issues. However, unlike home residents, the dormitory residents typically do not financially benefit from any reduction in electricity use resulting from their behavior changes, since residence hall fees are flat-rate and do not change based on energy usage. Since they lack even a monthly bill as feedback, residents are completely unaware of their energy usage,

Residence hall energy competition technologies range in complexity from simple web pages with weekly electricity data to complicated web applications [5, pp. 6–11]. An early adopter of the residence hall energy competition, Oberlin

Figure 1.   The Smart Grid Game widget, displaying level 2 actions.

College, developed a real-time electricity consumption feedback system as described by Petersen et al. [6]

To build on this area of active energy work, we decided to target our serious game to college students living in residence halls. The Kukui Cup extends the typical college energy competition into a broader energy *challenge* where electricity consumption feedback is only one part of a larger game experience for players. The challenge is named after the kukui nut (also known as candlenut), which was burned by Native Hawaiians to provide light, making it an early form of stored energy in Hawai'i.

In a Kukui Cup challenge, residents are grouped into teams based on where they live. Different floors of a building or entire buildings can be formed into teams. The electricity usage of teams is measured either through manual meter readings or through automated meter data collection. In addition to the energy competition, the Kukui Cup has a energy literacy competition where players can earn points by engaging in educational and social activities on the challenge website. The point system provides a way to motivate players to explore and use the system, as the Kukui Cup is currently deployed as an extracurricular activity.

Much of the point competition revolves around a section of the challenge website called the Smart Grid Game (SGG). The Smart Grid Game consists of rows of actions arranged into columns based on a particular topic (similar to the popular game show "Jeopardy"), shown in Fig. 1. Clicking on a square in the SGG shows details about the action and explains how players can complete the action to earn points. There are several types of actions: short YouTube videos on energy and sustainability topics, activities like measuring the flow rate of a shower, excursions such as visiting a farm that produces all its own electricity, and commitments such as carpooling or not eating meat. There are also creative actions such as writing a poem about energy or a letter to the editor on a sustainability topic. The flexibility of the SGG allows us to provide a wide variety of interesting actions for players to take part in.

The completion of each action (with the exception of

commitments) is verified through the challenge website before points are awarded. For activities, players are usually asked a randomly-selected question, and their answer is placed in a queue for challenge administrators to review. The administrator can approve or reject the submission, and can provide feedback on the players' answers. The game also supports activities that are verified by submission of an uploaded image such as a photo or screenshot.

A. Running a Kukui Cup

A Kukui Cup challenge consists of multiple components working together to provide the entire game experience. For challenges using real-time energy data, the open source WattDepot [7] system is used to collect, store, and analyze the data. The challenge website and associated game mechanics are provided by the open source Makahiki system [8]. The current educational content is tailored to the needs of college students living in residence halls in Hawai'i, but can be tailored to suit other audiences or goals.

B. Field Studies

In addition to in-lab evaluations and beta tests, there have been two sets of field studies of Kukui Cup challenges. The first Kukui Cup challenge took place over 3 weeks starting in October 2011 in four residence halls for first-year students on the University of Hawai'i at Mānoa campus containing a total of approximately 1070 residents. Pairs of floors, referred to as *lounges*, were the team unit in the 2011 Kukui Cup.

The second set of challenges started in September 2012. The University of Hawai'i (UH) Kukui Cup is taking place in the same four residence halls with approximately the same number of residents, but over the entire 9 month academic year. The first month of the competition was an intensive period with multiple real-world events taking place each week, while the remaining months will be less intensive. The goal of the much longer time frame is to discourage short-term and unsustainable behaviors (such as forgoing all electronic device use).

In addition to the 2012 UH Kukui Cup, Hawaii Pacific University (approximately 200 residents) and the East-West Center (approximately 130 residents) have run their own challenges using the Kukui Cup system with our support.

III. BASELINES AND GOALS

Goal setting has been shown to be an effective tool in changing energy consumption behavior [9], [10] and are a common component of energy feedback mechanisms. Setting achievable goals is important from a game play perspective, so goals must typically be based on previous energy use. The most common way to generate a goal is to calculate a *baseline* of energy usage based on past energy usage, and then set the goal as some percentage reduction from the baseline.
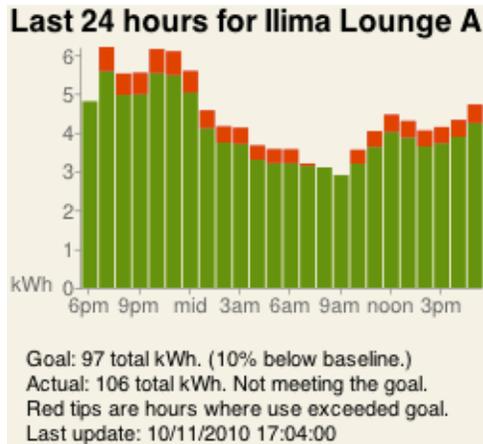
Figure 2.   A bar chart visualization of energy use as compared to a goal

Two of the most common ways to calculate the electricity baseline are to average recent prior usage (such as the last two weeks), or to average usage from previous years. Both of these methods are problematic because they assume that this previous usage is representative of future usage, even though there are many factors that can significantly alter electricity use over time including: occupancy, weather, activities (e.g., studying for a big midterm exam), and changes to the building infrastructure such as efficiency upgrades. Any of these factors can lead to the baseline being an inaccurate predictor of future usage in an energy competition, as described by Johnson et al. [11].

Since baselines can be poor predictors of future electricity use, comparing actual electricity use to the baseline in order to determine how much electricity was "saved" by an intervention is misleading and can tempt designers to make claims about energy saved that cannot be substantiated. However, comparison of actual electricity usage to a goal generated from a baseline can be helpful as a game mechanic to motivate players to conserve energy.

In the 2011 Kukui Cup, we used a baseline that was derived from an average of the two weeks prior to the challenge. In the 2012 Kukui Cups, we have switched to a dynamic baseline [11] that consists of the average electricity usage for the two previous weeks, but the baseline is recomputed every day throughout the challenge. The dynamic baseline means that as the challenge progresses, the baseline will include usage during the challenge. In essence, a goal generated from a dynamic baseline requires a team to reduce their energy usage compared to the recent past. Since the baseline is not a static value picked once before the challenge, anomalous conditions during the period before the challenge will soon be replaced with new, more representative data.

IV. ENERGY FEEDBACK DESIGN EXPERIENCES

Feedback on electricity consumption has been used as a means for facilitating energy conservation by researchers in the human-computer interaction community [2] as well as in the broader energy efficiency [12]–[14] and environmental psychology [9], [10] communities. One reason for this focus on feedback is undoubtedly the hidden nature of electricity, so feedback provides an awareness that is otherwise unavailable.

One of the fundamental principles of energy feedback in the Kukui Cup is that it be *actionable*. While any energy feedback may implicitly encourage energy conservation behaviors simply by making energy use visible, this does not meet our definition of actionable. A feedback display that shows that a home has used 20 kWh so far on a particular day leaves the viewer with natural questions: is that a lot? what should I do if I wanted to reduce my energy use?

A. The Energy Bar Chart Visualization

An early attempt at energy feedback for the Kukui Cup is shown in Fig. 2. This "Energy Bar Chart" shows hourly energy use for a team participating in the Kukui Cup over 24 hours as compared to an energy goal. Note that the data shown in this particular figure are simulated. Bars that are entirely green show the actual energy usage for that hour and indicate that the energy use was below the hourly goal. For mixed red and green bars, the main green portion represents the energy goal for that hour of the day, while the red tips of the bars represent the actual usage in excess of the goal.

This form of energy feedback shows the variation in energy use over the course of a day, which is an important energy literacy concept. It also shows in what parts of the day energy use is exceeding the goal, and by how much. By displaying the times during the day when the hourly goals are not being met, residents could focus on understanding what activities are going on during those periods.

As (naive) designers, we felt that this visualization provided a great deal of useful feedback both clearly and concisely. However, results of an in-lab evaluation were unequivocal: the visualization provided too much information, the meaning of its components was not obvious, and the "actionable" aspects were not obvious. This energy feedback visualization was a failure, and we began a redesign to address its deficiencies.

B. The Daily Energy Goal Game Feedback Visualization

To make our energy feedback easier to understand and also more actionable, we developed the Daily Energy Goal Game (DEGG) visualization shown in Fig. 3. The three most prominent components of the DEGG are: the energy consumption so far during the current day, the energy goal so far for the current day, and a traffic light that shows in the most straightforward way whether the team is meeting their

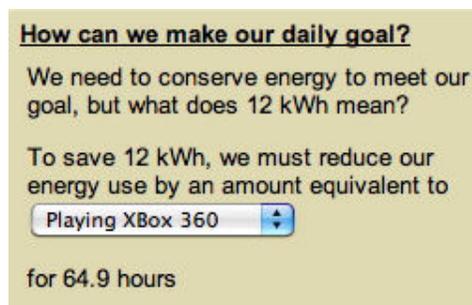Figure 3.   The Daily Energy Goal Game feedback visualization

Figure 4.   The "How can we make our daily goal?" widget

energy goal. The display updates once every 10 minutes with new energy data.

We picked the daily time frame for the game for two reasons. First, having a daily goal makes behavior changes more visible and feedback more immediate than a longer time frame such as weekly or monthly. Second, by concentrating on a daily goal, teams that are performing poorly on a particular day can redouble their efforts to do better the next day. Similarly, a team that does particularly well for one day cannot rest on their laurels, as they must make an effort to conserve every day. This game design reflects our belief that changing energy behaviors is a marathon and not a sprint: radical short-term changes made to win an energy competition are unlikely to be sustainable, and therefore are of very limited utility in achieving long-term energy conservation.

Residential energy use varies in intensity over the course of a day: typically low when people are sleeping and much higher during evening hours. For the students in the residence halls in our studies, the energy usage peak occurs at approximately midnight, and the lowest between 8 and 9 AM, which is considerably different than an average single-family home. There is also daily variation between days of the week, as the activities taking place on a Monday night are different than those on a Saturday night. To account for the hourly and daily variation in energy use, we computed hourly and daily baselines for energy use, and the goal value is a percentage reduction from the baseline. The energy consumption and goal values displayed in the DEGG are computed over the time period from midnight to the current time. This choice of time frame is particularly important for the goal value, because if a daily goal value were simply spread linearly over the course of a day, players would see their energy use as always under the goal during low-usage periods, and going above the goal during the high-usage periods, possibly to a degree that makes it impossible to meet the goal for that day.

The DEGG also links the energy conservation competition with the point competition. When a team meets their daily energy goal, each team member is awarded an administrator-

configured number of points. This linkage provides an additional incentive for players to pay attention to the energy competition, because successfully reducing energy use below the goal can significantly increase team point totals.

Below the traffic light display of the DEGG is a list of actions from the Smart Grid Game that players can take to either learn more about energy, or directly help reduce their energy use. The actions displayed depend on what actions the player has already completed in the rest of the system. The DEGG is highly actionable because it provides direct links to actions that players can take to reduce their energy usage, tailored to the opportunities available in their residence hall.

Evaluation of the DEGG visualization during actual game play indicates that players do not have a problem understanding this visualization. The stoplight image provides a clear, unambiguous signal, and the actual/goal numbers provide further context. In addition, the visualization is explicitly paired with links to descriptions of appropriate actions for that player in the context of the game and the team's current energy use. Log data indicates that players do click on these links in order to understand how to take action based on the energy feedback. This energy feedback was a success and is included in the current version of the Kukui Cup.

C. The "Wii Hours" Energy Feedback Visualization

In another energy feedback design effort, we created a small widget below the DEGG titled "How can we make our daily goal?". This widget, shown in Fig. 4, showed how much the player's team energy usage was above the goal, and provided a drop-down menu of electrical devices commonly present in student rooms: laptops, XBox 360, Wii, etc. When a device was selected from the menu, the system would display the approximate number of hours of device use that would equal the amount of team energy use over the goal value. The time value was intended to show players how much device use they would need to *forego* in order to get back on track to their energy goal, and develop their intuition about the relative power use of different devices (i.e., plasma TVs use much more power than Wii game

consoles). Therefore, a short time value could point out an easy way to make the goal, and a long time value would indicate less significant energy conservation.

However, during in-lab evaluations of the system, we found that multiple subjects misinterpreted the time value, thinking that high time values were bad rather than good. Since the Wii was the device on the list with the smallest power use (20 W) compared to an XBox 360 or Playstation 3, it led to the highest time values. Some subjects drew the conclusion that using a Wii was worse than using an XBox 360 or Playstation 3, which was precisely the opposite goal of this widget. One subject even took the time to use our in-game team discussion forum to post the message "don't play wii" after using the widget! Because of this example, we dubbed this confusion the "Wii problem".

Clearly, energy feedback that can lead at least some players to the opposite conclusion than intended is a failure. The "Wii Hours" visualization never made it into production, and we are still searching for a design variant that can convey this information in an unambiguous fashion to players with minimal energy literacy.

V. Domain Knowledge and Energy Feedback

Energy feedback systems provide data on some aspect of behavior with the goal of reducing negative environmental impact [2]. However, they often assume users possess some level of domain knowledge about the environmental topic they hope to address. The term *energy literacy* has been used to describe the understanding of energy concepts as they relate both on the individual level and on the national/global level.

Some examples of energy literacy are: understanding the difference between power and energy; knowing that a microwave uses much more power than a refrigerator, but that the refrigerator will use much more energy over the course of the day; and knowing how electricity is generated in one's community.

Unfortunately, all indications are that energy literacy is low in the United States. DeWaters and Powers have developed an energy literacy survey instrument for middle and high school students. They found that the student mean attitude scores were 73%, but that knowledge scores lagged far behind (42% correct) [15]. Based on their findings, they make some recommendations, such as energy curricula be "hands on, inquiry based, experiential, engaging, and real-world problem solving...", and using the campus as a "learning laboratory". Similarly a nationwide survey of adults on energy by Southwell et al. found that the average respondent answered fewer than 60% of the energy knowledge questions correctly [16].

One energy literacy topic that we emphasize in the Kukui Cup is the difference between power and energy, power being the rate at which energy is being consumed or produced (measured in watts) and energy is the quantity of work that can be performed by a system (measured somewhat confusingly for electricity in kilowatt-hours). In the Kukui Cup we explain this relationship as being analogous to speedometer and odometer in a car.

Through answers submitted to the online activities in the Kukui Cup, we can see that many players have trouble understanding the concepts of power, energy and their interrelationship. Players often confuse the two concepts and often fail to grasp the time sensitivity of power, and thereby considering devices that consume a lot of power as "bad" irrespective of how long they are actually used. When the users of visualizations do not understand the concepts that are being visualized, understanding of the visualizations becomes much more difficult. It is for this reason that we claim that energy feedback systems should incorporate educational components, or risk being unintelligible to users. However, we reject the notion that power and energy, watts and kilowatt-hours are too complicated and that users should be provided instead with analogies to cars driven or hamburgers eaten. These energy concepts are important for effective customer participation in the smart grid, and should not be reduced to analogies alone.

VI. Energy Feedback, Stickiness, and Serious Games

A meta issue for all energy feedback systems is how to ensure that they continue to be "sticky" for users, as a feedback system that users do not view will be unable to accomplish anything. There are indications that the long-term impact of energy feedback may be diminished due to habituation. Froehlich suggests that the average user will spend less than one minute per day exploring their energy consumption behaviors [17]. A study by Houde et al. of households using Google PowerMeter found an "immediate decrease in electricity consumption, but in the long term these electricity savings decrease and disappear." [18] This finding suggests that a primary concern for any energy feedback system is ensuring that users continue to interact with it over the long term. Put another way, energy feedback alone is not enough to accomplish the goal of long-range customer engagement with their energy consumption.

One solution to the lack of stickiness of energy feedback systems is the incorporation of game play. Serious games like the Kukui Cup provide an alternative route to promote both learning and engagement with energy feedback. It is for this reason that we designed the Kukui Cup as a serious game that incorporates electricity consumption feedback as one aspect of the game experience, rather than an energy feedback system that has been "gamified".

While games are not the only way to promote long-term engagement with energy issues, we submit that any normal energy feedback system will quickly be abandoned by users once the novelty wears off. There must be a continuing

reason for users to revisit the system that even the most novel and interesting energy feedback systems lack.

VII. Future Work

The Kukui Cup is an ongoing project and we continue to build upon our initial work. The first area of future work is the 2012 UH Kukui Cup. The 2012 UH Kukui Cup will last nine months and should shed light on several issues. Can we maintain player interest over longer periods despite fewer prompts from intensive marketing and events? Can the new player-provided educational content fill the gap of the longer challenge duration? What are the results of the DEGG with dynamic baselines over a long time frame?

The 2012 Kukui Cups happening at Hawaii Pacific University and the East-West Center will also offer new insights into how challenge administrators outside our research group design Kukui Cup challenges tailored to their organization, and how different student populations perform.

A longer range goal is to integrate the Kukui Cup with Hawai'i's smart grid efforts. The Kukui Cup is currently a effort-intensive program, so scaling to hundreds of thousands of players will require scaling the management of the challenge, finding a means of funding, and a way for players to incorporate household energy data fairly, in a completely heterogenous environment.

One final area of research is longitudinal studies of players after the game is over and they have moved out of the residence halls. We want to find out whether the Kukui Cup experience actually had lasting impacts on players, and whether they were able to continue any new behaviors after leaving the context of the residence hall.

VIII. Conclusion

We have described the Kukui Cup serious game, and our results from field trials of the system. We have discussed some of the energy feedback visualizations we developed, including both those that succeeded and those that failed. Based on our experiences, we provide three areas that energy feedback systems for the smart grid should address: they should be actionable, they must address users lack of domain knowledge, and they must find ways to be sticky.

Acknowledgments

References

[1] M. Zyda, "From visual simulation to virtual reality to games," IEEE Computer, vol. 38, no. 9, pp. 25 – 32, Sep 2005.

[2] J. Froehlich, L. Findlater, and J. Landay, "The design of eco-feedback technology," in Proceedings of the 28th international conference on Human factors in computing systems - CHI '10. New York, New York, USA: ACM Press, Apr. 2010, pp. 1999–2008.

[3] R. S. Brewer, G. E. Lee, and P. M. Johnson, "The Kukui Cup: a dorm energy competition focused on sustainable behavior change and energy literacy," in Proceedings of the 44th Hawaii International Conference on System Sciences, January 2011, pp. 1–10.

[4] C. Hodge, "Dorm energy competitions: Passing fad or powerful behavior modification tool?" Presentation at the 2010 Behavior Energy and Climate Change conference, November 2010. [Online]. Available: http://www.stanford.edu/group/peec/cgi-bin/docs/events/2010/becc/presentations/2C_ChelseaHodge.pdf [retrieved: January, 2013].

[5] G. E. Lee, "Makahiki: An extensible open-source platform for creating energy competitions," Master's thesis, University of Hawaii, June 2012. [Online]. Available: http://csdl.ics.hawaii.edu/techreports/2011/11-01/11-01.pdf [retrieved: January, 2013].

[6] J. E. Petersen, V. Shunturov, K. Janda, G. Platt, and K. Weinberger, "Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives," International Journal of Sustainability in Higher Education, vol. 8, no. 1, pp. 16–33, 2007.

[7] R. S. Brewer and P. M. Johnson, "WattDepot: An open source software ecosystem for enterprise-scale energy data collection, storage, analysis, and visualization," in Proceedings of the First International Conference on Smart Grid Communications, Gaithersburg, MD, October 2010, pp. 91–95.

[8] G. E. Lee, Y. Xu, R. S. Brewer, and P. M. Johnson, "Makahiki: An open source game engine for energy education and conservation," Department of Information and Computer Sciences, University of Hawaii, Honolulu, Hawaii 96822, Tech. Rep. CSDL-11-07, January 2012. [Online]. Available: http://csdl.ics.hawaii.edu/techreports/2011/11-07/11-07.pdf [retrieved: January, 2013].

[9] L. J. Becker, "Joint effect of feedback and goal setting on performance: A field study of residential energy conservation," Journal of Applied Psychology, vol. 63, no. 4, pp. 428–433, 1978.

[10] J. H. van Houwelingen and W. F. van Raaij, "The effect of goal-setting and daily electronic feedback on in-home energy use," The Journal of Consumer Research, vol. 16, no. 1, pp. 98–105, June 1989.

[11] P. M. Johnson, Y. Xu, R. S. Brewer, G. E. Lee, M. Katchuck, and C. A. Moore, "Beyond kWh: Myths and fixes for energy competition game design," in Proceedings of Meaningful Play 2012, October 2012, pp. 1–10.

[12] S. Darby, "The effectiveness of feedback on energy consumption," Environmental Change Institute, University of Oxford, Tech. Rep., 2006. [Online]. Available: http://www.eci.ox.ac.uk/research/energy/downloads/smart-metering-report.pdf [retrieved: January, 2013].

[13] A. Faruqui, S. Sergici, and A. Sharif, "The impact of informational feedback on energy consumption–a survey of the experimental evidence," Energy, vol. 35, no. 4, pp. 1598–1608, 2010.

[14] B. Foster and S. Mazur-Stommen, "Results from recent real-time feedback studies," American Council for an Energy-Efficient Economy (ACEEE), Tech. Rep. B122, February 2012. [Online]. Available: http://aceee.org/research-report/b122 [retrieved: January, 2013].

[15] J. E. DeWaters and S. E. Powers, "Energy literacy of secondary students in New York State (USA): A measure of knowledge, affect, and behavior," Energy Policy, vol. 39, no. 3, pp. 1699–1710, 2011.

[16] B. G. Southwell, J. J. Murphy, J. E. DeWaters, and P. A. LeBaron, "Americans' perceived and actual understanding of energy," RTI Press, Tech. Rep. RR-0018-1208, 2012. [Online]. Available: http://www.rti.org/pubs/rr-0018-1208-southwell.pdf [retrieved: January, 2013].

[17] J. Froehlich, "Moving beyond line graphs: The history and future of eco-feedback design," Presentation at the 2010 Behavior Energy and Climate Change conference, 2010. [Online]. Available: http://peec.stanford.edu/docs/events/2010/becc/presentations/3D_JonFroehlich.pdf [retrieved: January, 2013].

[18] S. Houde, A. Todd, A. Sudarshan, J. A. Flora, and K. C. Armel, "Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence," The Energy Journal, vol. 34, no. 1, pp. 87–102, 2013.

A Large-scale Power-saving Cloud System Composed of Multiple Data Centers

Toshiaki Suzuki, Tomoyuki Iijima,
Isao Shimokawa, and Toshiaki Tarui

Central Research Laboratory
Hitachi, Ltd.
Yokohama, Kanagawa, Japan
{toshiaki.suzuki.cs, tomoyuki.iijima.fg,
isao.shimokawa.sd, toshiaki.tarui.my}@hitachi.com

Shinichi Kuwahara, Hidenori Takagi,
and Tomohiro Baba

Telecommunications & Network Systems Division
Hitachi, Ltd.
Kawasaki, Kanagawa, Japan
{kuwahara_s, takagi_hide}@itg.hitachi.co.jp,
tomohiro.baba.mn@hitachi.com

Abstract— A large-scale power-saving cloud system-composed of multiple data centers (DCs) and a wide-area network (WAN) connecting them is proposed. In this system, to reduce power consumption of the DCs and the WAN, virtual machines (VMs) are migrated and data routing paths are optimized under the condition that quality of service (QoS) is maintained by simultaneously providing necessary CPU resources and network bandwidth for services by a VM. To address the issue of excess VM migration (causing network congestion) due to separate control of "server resource" and "network resource" by a conventional power-saving scheme, the proposed system controls power consumption by cooperation between an inter-DC management server and a WAN management server. To determine an appropriate resource allocation, conditions for various resources (such as CPU loads and bandwidth consumed by network switches) are monitored in real time. In addition, future loads for the resources are periodically predicted. An appropriate VM reallocation is only executed when necessary resources after the reallocation can be guaranteed. A prototype system comprising 200 VMs, 200 servers, and four DCs was developed and evaluated. The evaluation results indicate that the system can achieve power saving by VM migration between DCs under the condition that the necessary CPU resource and network-access bandwidth for providing services by a VM are maintained.

Keywords- power saving; QoS; cloud system; virtual-machine migration; network congestion; resource allocation

I. INTRODUCTION

Lately, the amount of electric power consumed by information and communication technology (ICT) systems has been dramatically rising [1] in conjunction with the increasing number of data centers (DCs) being constructed. As one of the biggest issues concerning ICT systems, including DCs, power-saving measures have therefore been attracting lots of attention. [2].

Subsequently, to address the above-mentioned power-consumption issue, technical developments and standardizations aiming to make ICT systems more power efficient have been actively promoted. For example, "server-resource virtualization" (that is, saving power consumed by a server on the basis of optimization of necessary resources) has been under research and development [3], [4]. In addition, many standardization activities, such as those undertaken by the Energy Management Working Group (EMAN) in the Internet

Engineering Task Force (IETF) [5], the Institute of Electrical and Electronics Engineers (IEEE) [6], the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) [7], and the Distributed Management Task Force, Inc. (DMTF) [8], are continuing.

Although the above-mentioned activities have aimed at reducing the electric-power consumption of ICT systems, the power consumptions of the "server resource" and "network resource" are controlled separately. Power-saving control has therefore been optimized for each resource, and total optimization of electric-power consumption while maintaining service quality provided by a large-scale cloud system comprising multiple DCs and a wide-area network (WAN) to connect them has not been focused on by conventional activities. Besides, if electric-power saving for one resource is conducted separately, it might cause a serious problem for other resources. For example, an excessive aggregation of servers by virtual-machine (VM) migrations might degrade access quality to a VM since data flows are also aggregated to the same routing path; as a result, network link bandwidth is exceeded, and network congestion occurs.

We are aiming to develop efficient power-saving control scheme for both network and server resources while guaranteeing network and server "quality of service" (QoS), such as bandwidth and CPU power, by integrated power-consumption management of both network and server resources. In a previous work [9], we proposed a power-saving cloud system managed by one control system. In the present work, aiming at total electric-power saving for both WAN and DCs resources, we propose a large-scale power-saving cloud system managed by cooperation between a WAN management server and integrated DC management servers.

The rest of this paper is organized as follows. Section II explains the requirements of a power-saving cloud system. Section III proposes a large-scale power-saving cloud system managed by a WAN management server and integrated DC management server. The proposed system simultaneously saves electric power and guarantees access bandwidth to a VM. Section IV describes a prototype system and presents some results of a performance evaluation. Related works are shown in section V and section VI concludes the paper.

II.   REQUIREMENTS OF POWER-SAVING CLOUD SYSTEM

A power-saving cloud system provides various services and resources, such as application software, CPU processing power, and storage, via a network. To create a power-saving cloud system and to reduce total electric-power consumption during off-peak hours (such as late evening), only the minimum resources required for providing cloud services should be activated.

To control the power of a target system, average loads on physical servers, VMs on those servers, and network nodes should be monitored. In addition, VMs should be appropriately reallocated according to the future loads on servers and VMs predicted under a predefined threshold during off-peak hours. After the appropriate reallocation of VMs, unnecessary physical servers should be turned off. The nodes or ports on the nodes that transmitted data to unnecessary physical servers should also be turned off or switched from active mode to sleep mode. Furthermore, service quality (such as access bandwidth to a VM) should be guaranteed before, as well as after, the power-saving control by VM migration. In addition, a power-saving scheme should be applied to not only small cloud systems comprising a single DC but also large-scale systems comprising multiple DCs.

To summarize the above-mentioned requirements, the power-saving control should be executed according to the following procedures, namely, four power-saving policies.

- Policy 1: Power consumption of the DC can be reduced by turning off unnecessary physical servers that are no longer used after an appropriate reallocation of VMs by VM migration in the DC.
- Policy 2: Power consumption of the DC can be reduced by turning off unnecessary physical servers and network nodes that are no longer used after aggregation of running physical servers and data transmission routes by VM migration in the DC.
- Policy 3: Power consumption of the DC can be reduced by turning off unnecessary physical servers and nodes in the DCs that are no longer used after aggregation of physical servers and data transmission routes by VM migration between DCs based on cooperation between DC management and WAN management.
- Policy 4: Power consumptions of the DC and WAN can be reduced by turning off nodes or their ports in the WAN that are no longer used after VM migration between DCs and aggregation of data-transmission routes.

The above four policies are resource-control procedures from the viewpoint of power saving. In addition, resources should also be controlled from the viewpoint of service quality. More specifically, power consumption of the system should be reduced by aggregation of both server resources and network resources while service quality of a network path between an end user and the VM providing application services is maintained.

III.   PROPOSED POWER-SAVING CLOUD SYSTEM

A. System Architecture

The typical structure of the proposed power-saving cloud system is shown schematically in Fig. 1. The system is composed of multiple DCs and a WAN connecting them. More specifically, the DC consists of multiple switches (SWs) for transmitting data, servers for providing various services, a DC management server for controlling resources in the DC, and an inter-DC management server for controlling multiple DC management servers. The WAN consists of multiple SWs, an "integrated-mining-of-flow" (IMF) apparatus for monitoring network conditions, and a WAN management server for controlling resources in the WAN.

In the power-saving cloud system, the DC management server monitors the loads of servers, VMs, and SWs in the DC in real time. In addition, it predicts future loads of these resources according to statistical analysis (such as autoregressive model analysis [10]) based on the past history of loads. Besides, to reduce electric-power consumption on the DC side, it determines and controls an appropriate reallocation of resources such as VMs and routing paths.

To reduce power consumption of the WAN, the IMF monitors loads of SWs in the WAN. The WAN-management server receives statistical-monitoring data and predicts future loads on each SW. Electric power consumed by the WAN is saved by optimizing data-routing paths and turning off SWs or their ports that are no longer used.

In summary, power consumption of the total system is reduced by reallocating VMs between the DCs appropriately on the basis of cooperation between multiple DC-management servers and the WAN-management server.
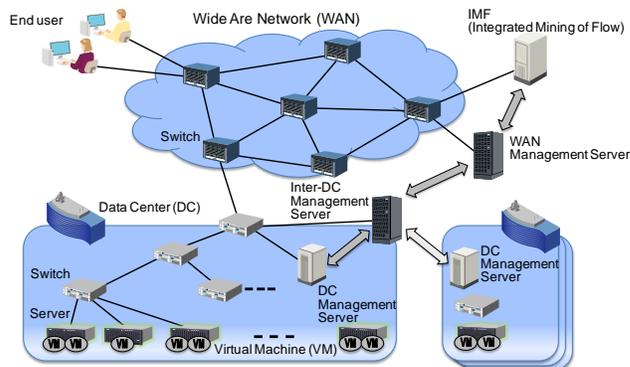

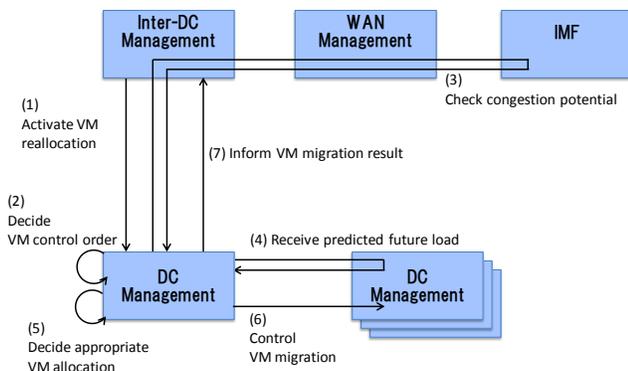Figure 1. Proposed power-saving cloud system

Figure 2. Process steps of proper VM resource reallocation

B. Overview of Power-saving Scheme by VM Migration

The process steps of a typical power-saving scheme by VM migration based on multiple management servers are shown schematically in Fig. 2. In the proposed system, the inter-DC management server activates power-saving control according to the loads on the physical servers and VMs [step (1)]. The DC management server determines the order of VM migration [step (2)]. The DC management server checks a "congestion potential" via the inter-DC management server, the WAN management server, and the IMF [step (3)]. For VM migration between DCs, the DC management server receives predicted loads on alternative physical servers from other DC management servers [step (4)]. To move the VM according to the predicted loads on the outside servers and the effectiveness of the power saving, the DC management server determines one alternative server [step (5)] and triggers an actual VM migration [step (6)]. The VM-migration result is transmitted from the DC management server to the inter-DC management server [step (7)].

C. Detailed VM Resource Reallocation

The seven above-mentioned steps for resource reallocation are explained in detail in the following:

1) VM reallocation trigger by inter-DC management server: The inter-DC management server starts or stops optimizing reallocation of VMs to each DC management server when the loads on servers and VMs are low (such as late evening).

2) Determination of VM reallocation order by DC managemnet server: The DC management server determines the order to reallocate running VMs for each virtual local area network (VLAN). The reallocation order is determined according to (i) decending order of idle power, (ii) ascending order of number of running VMs on a server, (iii) ascending order of assigned CPU resources, and (iv) ascending order of assigned memory resources.

3) Checking of congestion potential in WAN by DC management server: To maintain access quality to the VM after VM migration to another DC, the DC management

server receives the congestion potential concerning the WAN from the IMF via the inter-DC management server and the WAN management server. The congestion potential is evaluated according to the history of the monitored data and predicted future loads in the case of fluctuation of bandwidth for each port of the switch. If there is any posibility of network congestion in the future, data-routing paths including the congestion point are not used for VM migration.

More specifically, the IP address of the VM to reallocate, the identifier of the source DC, and the identifier of the VLAN to which the VM belongs are transmitted from the DC management server to the inter-DC management server. A list of alternative DCs that can accommodate the migrated VM and above-mentioned information from the DC management server is then transmitted from the inter-DC management server to the WAN mananagement server. After that, information about a routing path (from a WAN edge point connecting a user to another WAN edge point connecting an alternative DC) and the above-mentioned information from the inter-DC management server are transmitted from the WAN management server to the IMF. The congestion potential at the routing path between the user and the altenative DC is sent from the IMF to the DC management server via the WAN management server and the inter-DC management server.

4) Determination of target server for VM migration by DC management server: The DC management server detemines the appropriate VM reallocation by considering all alternative DCs. Specifically, all severs that can provide enough resources to run the intended VM in the future and guarantee access qualtiy to the VM at the same time are selected as alternative servers for the reallocation of the VM. The most effective server for power saving is then selected as a final target server for the VM migration.

More specifically, the DC management server predicts future loads on the CPU and consumption of the bandwidth resource by the intended VM. In addtion, it gets information concerning predicted available future resources (such as CPU and bandwidth) for the alternative servers in other DCs from other DC management servers. It finally determines one target server to which the intended VM is reallocated by comparing the received available future resources for all alternative servers in other DCs and the amount of necessary resources for the intended VM in the future.

5) Determination of VM reallocation by DC management server: The DC management server determines whether target servers can provide enough resources (such as CPU processing power and memories) for running the intended VM in the future. The only servers that can provide enough resources are registered as alternative servers for VM migrations. In addition, the DC management server determines whether switches on the routing path between the entrance of the DC and the

alternative server in another DC can provide enough bandwidth for the intended VM after the VM migration. It checks the congestion potential for the routing path between the WAN edge connecting the DC and another WAN edge connecting an end user on the basis of the monitored information from the IMF. To determine the most appropriate alternative server, the DC management server checks all the above-mentioned evaluation points, i.e., CPU load, network congestion, and bandwidth. The most appropriate server that can meet the requirements stated in Section II and has the most effective power-saving advantage is then selected as the target server for the VM migration by the DC management server.

6) VM migration by DC management server: The VM migration is executed according to the trigger of the DC management server. As for VM-migration methods, various technologies have been developed by several organizations [3], [4], and these technologies can be used for an alternative VM-migration scheme by combining them with the proposed power-saving cloud system. After the VM migration, the DC management server updates stored topology information. In addition, to predict future load, when the VM has been migrated to a server in another DC, the history of the VM's resources (such as CPU load) is moved to another DC management server.

7) Information about VM-migration completion sent from DC management server to inter-DC management server: After all VM migrations have been executed, the completion of all VM reallocations is transmitted from the DC management server to the inter-DC management server. In addition, the migration histories from the source servers to destination servers are transmitted from the DC management server to the inter-DC management server. The inter-DC management server receives the migration histories and stores them. These histories are used when the migrated VMs are returned to the original allocated servers when CPU load increases.

D. Power-consumption Model

A power-consumption model for the proposed cloud system is defined as follows. The amount of power (P_{All}) consumed by the cloud system is given by formula (1), where P_{IT} means power consumption of IT equipment, and P_{NET} means power consumption of network nodes. Formula (2) indicates P_{IT} is calculated by a summation of power consumption (P_{SV}) of servers since the proposed system includes multiple servers as IT equipment. Here, i ($i = 1, 2, 3, . . , N$) mean the number of the server. In addition, n means CPU load (%) on the server. P_{SV} is given by formula (3). $P_{idle(i)}$ means the power consumption of the ith server during idle time, and $P_{max(i)}$ means power consumption during maximum load. Formula (4) gives P_{NET} of a network calculated by the summation of the power consumption of each node. Here, k ($k = 1, 2, 3, . . , M$) mean the number of

the node. In addition, m means load (%) on a node in terms of bandwidth. The power consumption of the node (P_{NODE}) is given by formula (5). $P_{idle(k)}$ means power consumption by the kth node during idle time, and $P_{max(k)}$ means power consumption during maximum load. Here, P_{SV} and P_{NODE} are assumed to fit a linear function, as shown in Fig. 3. The relations between the power and CPU loads and between the power and traffic are independently evaluated in advance. According to that evaluation, the relation between power consumption and load (traffic) fits a linear function well (as shown in Fig. 3).

$$P_{All} = P_{IT} + P_{NET} \qquad (1)$$

$$P_{IT} = \Sigma_i P_{SV(i)}[n] \qquad (2)$$

$$P_{SV(i)}[n] = P_{idle(i)} + (P_{max(i)} - P_{idle(i)})(n/100) \qquad (3)$$

$$P_{NET} = \Sigma_k P_{NODE(k)}[m] \qquad (4)$$

$$P_{NODE(k)}[m] = P_{idle(k)} + (P_{max(k)} - P_{idle(k)})(m/100) \qquad (5)$$
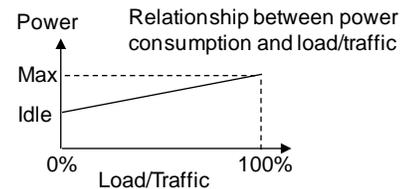
Figure 3. Assumed power consumption based on load/traffic

IV. EVALUATION OF PROPOSED SYSTEM

A. Evaluation System

The evaluation system is shown schematically in Fig. 4, and the number of pieces of ICT equipment is listed in Table I. In the system, switches, servers, and VMs in the DCs are emulated by open-source software, while switches in the WAN and management servers are real apparatuses. The performances of the power-saving control of the DCs and WAN are evaluated in detail in the following sections.
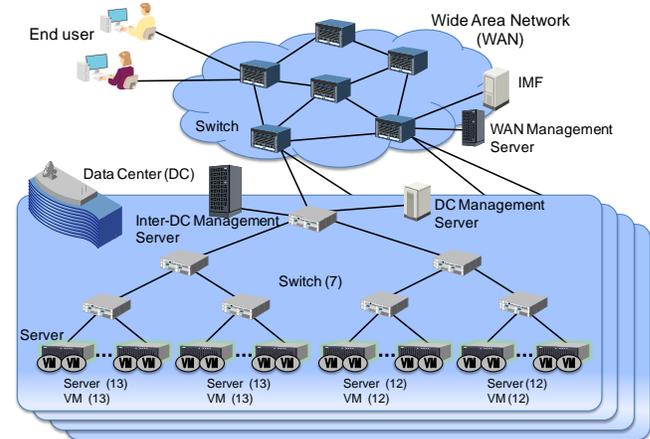
Figure 4. Evaluation system

TABLE I. NUMBER OF PIECES OF ICT EQUIPMENT

	Item	Number of pieces of ICT equipment in 4 DCs
1	WAN management server	1
2	IMF	1
3	SW in the WAN	6
4	DC	4
5	Inter-DC management server	1
6	DC management server	4
7	SW in the DCs	28
8	Server in the DCs	200
9	VM on the servers	200

B. Evaluation of Power-saving Control for DCs

The effectiveness of the power-saving control for DCs per day was evaluated. First, a CPU load model of a VM in the DC is assumed. The bandwidth consumed by the VM for one day is also assumed. The power consumption by DCs for one day is evaluated according to these assumptions.

1) Workload model for a VM per day

The assumed loads on the CPU as well as the incoming flow to and the outgoing flow from a VM are schematically shown in Fig. 5. As depicted in the figure, the peak load is set only one time (around noon), and the loads during business hours are high while the loads during the night for the CPU, incoming flow, and outgoing flow are low. The effectiveness of the power-saving control scheme is evaluated by comparing two cases: executing appropriate VM reallocations and not executing them.

The topology of the DC is shown in the lower part of Fig. 4. The specifications and number of pieces of each apparatus in the DC are listed in Table II.
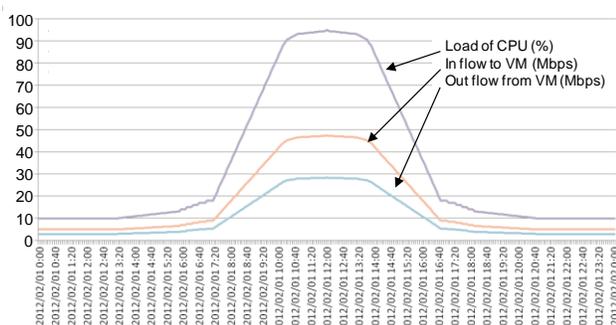
Figure 5. Load model of VM (CPU and in/out flow)

TABLE II. SPECIFICATIONS AND NUMBER OF PIECES OF EACH APPARATUS IN ONE DC

	Apparatus	Idle power	Max. power	Number
1	Server (Model 1)	120 W	170 W	17
2	Server (Model 2)	110 W	150 W	17
3	Server (Model 3)	177 W	251 W	16
4	VM	—	—	50
5	Switch	350 W	450 W	7
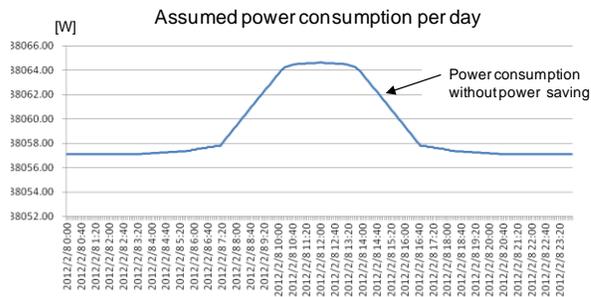
Figure 6. Electric power consumption of a DC per day

*2) Electric-power consumption of a DC per day*

Electric-power consumption of a DC for one day (under the assumed loads for each server shown in Fig. 5) is shown in Fig. 6. The effectiveness of the power-saving control scheme under the following three conditions was evaluated. In the first condition, the VM is reallocated when the CPU loads are less than 75%. In the second and third conditions, reallocations are executed under CPU loads of 50% and 25%, respectively. On the other hand, when the load on the CPU is over these thresholds, reallocated VMs are returned to the original locations to guarantee service quality.

3) Energy consumption of DCs per day

The evaluated fluctuations of power consumption of all DCs for the three above-mentioned power-saving control conditions (CPU loads of 75%, 50%, and 25%) are shown in Fig. 7. The result in the case of no VM reallocation is also shown in the figure for comparison. As shown in the figure, the effectiveness of the power-saving control scheme under the three conditions is verified.

In addition, the results for VM reallocation keeping VM access quality and energy consumption per day are listed in Table III. The number of VMs is shown in the upper row, while the number of servers (SV) is shown in parentheses in the lower row. According to the table, some VMs are migrated between DCs (since the number of VMs in the DC is changed after appropriate VM reallocations). In addition, the number of running servers is dramatically reduced after the VM migration. Here, the CPU resource for a server is assumed to be enough for six VMs with CPU loads of 50%. The reductions in energy consumption at CPU loads of 25%, 50%, and 75% are 45.2%, 45.7%, and 47.6%, respectively.

Figure 7. Electric-power consumption of a DC per day

TABLE III. ELECTRIC-ENERGY CONSUMPTION OF DCS PER DAY

	Optimization timing	DC1	DC2	DC3	DC4	Energy consumption per day (kWh)
		VM (SV)	VM (SV)	VM (SV)	VM (SV)	
1	No optimization	50 (50)	50 (50)	50 (50)	50 (50)	913.421
2	CPU load: 25%	60 (5)	48 (4)	48 (4)	44 (4)	500.388
3	CPU load: 50%	54 (9)	48 (8)	48 (8)	50 (9)	496.395
4	CPU load: 75%	52 (13)	48 (12)	52 (13)	48 (12)	478.323

C. Power-saving Evaluation for WAN

Power-saving control for a wide-area network (WAN) for one day was evaluated. In particular, the effectiveness of the power-saving scheme based on bandwidth control by link aggregation was evaluated. The topology of the evaluated WAN is shown in Fig. 4. In addition, the specifications of the switches in the WAN are the same as those listed in Table II.

Power-saving control by appropriate data routing (including link-aggregation control) was executed after appropriate VM reallocation between DCs. The fluctuation of power consumption of the WAN is shown in Fig. 8. In addition, energy consumptions under the three types of control are compared in Table IV. According to these results, the reductions in energy consumption achieved by the power-saving control scheme under CPU loads of 25%, 50%, and 75% are 10.4%, 12.0%, and 13.7%, respectively.
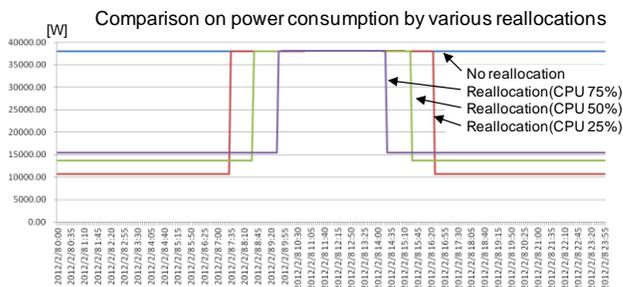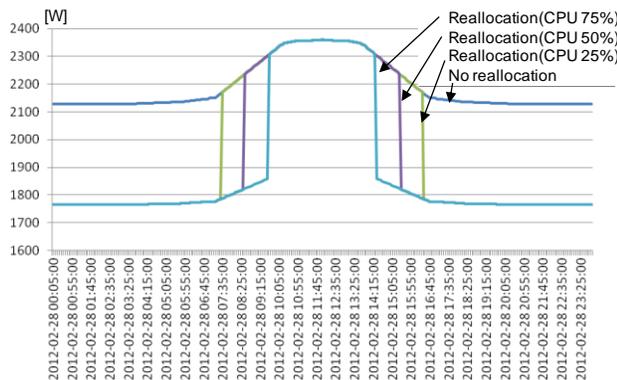

Figure 8. Electric power consumption of WAN per day

TABLE IV. ELECTRIC-ENERGY CONSUMPTION OF WAN PER DAY

	Optimization timing	Electric-energy consumption per day (kWh)	Reduction (%)
1	No optimization	52.470	—
2	25% CPU load	46.989	10.4
3	50% CPU load	46.194	12.0
4	75% CPU load	45.265	13.7

TABLE V. ELECTRIC-ENERGY CONSUMPTION OF ENTIRE CLOUD SYSTEM PER DAY

	Optimization timing	Optimization term	Electric-energy consumption (kWh)	Reduction (%)
1	No optimization	—	965.891	—
2	CPU load: 25%	15h00m	547.377	43.3
3	CPU load: 50%	17h00m	542.589	43.8
4	CPU load: 75%	19h10m	523.588	45.8

D. Power-saving Effect for Entire Cloud System

The effectiveness of the power-saving control scheme for an entire cloud system is shown in Table V. The reductions of energy consumption achieved by the power-saving control scheme under CPU loads of 25%, 50%, and 75% are 43.3%, 43.8%, and 45.8%, respectively. As shown in the table, energy consumption is reduced by approximately 40%. In addition, the highest reduction is accomplished under CPU load of 75%.

E. Discussion of Power-saving Effect

According to the results of this evaluation of a large-scale power-saving cloud system composed of multiple DCs and a WAN, energy consumption of the entire system is reduced by about 40%. With regard to only the power saving for the DCs, energy consumption is reduced by over 45%. On the other hand, energy consumption of the WAN is reduced by only about 10%.

The reason that the reduction of energy consumption of the DCs is high is the effectiveness of turning off unnecessary servers after appropriate VM reallocation. On the other hand, the reason that the reduction of the energy consumption of the WAN is low is that unnecessary switches were not turned off (since turning off unnecessary links (network ports) is only possible for the assumed system).

With regard to power saving for the entire system, the reductions in energy consumption achieved by the power-saving control scheme under CPU loads lower than 25%, 50%, and 75% are 43.3%, 43.8%, and 45.8%, respectively. On the other hand, the periods for the resource optimization under the three above conditions are 15 hours, 17 hours, and 19 hours and 10 minutes, respectively. When the power-saving control is executed under a CPU load of 75%, the period for the optimization is the longest, and reduction in energy consumption is the highest. These results verify the effectiveness of the proposed power-saving control scheme.

V. RELATED WORK

In previous researches, many power-saving schemes for ICT systems have been proposed. For example, power-saving schemes for node and link levels [11], [12] have been proposed. These schemes are useful for our proposed system

to reduce power consumption for the link level. In addition, power-saving schemes [13]-[16] for the network level have been proposed. Power-saving schemes for the DC/server level [17]-[20] have also been proposed.

In conventional power-saving researches like those mentioned above, network and DC/server resources are controlled separately. Therefore, integrated management for maintaining network QoS and reducing energy consumption of servers is addressed in the current study.

VI. CONCLUSION

A large-scale power-saving cloud system comprising multiple DCs and a WAN is proposed. As for this system, VMs are reallocated under condition of guaranteeing necessary CPU resources and network bandwidth for providing cloud services. Energy saving for the entire system is executed by cooperation between a DC management server and an inter-DC management server. The energy saving for the DCs is executed by VM migration between DCs and aggregation of running servers under maintained network-access quality to a VM. On the other hand, the energy saving for the WAN is executed by controlling link aggregation according to the required bandwidth for transmitting user data between a first WAN edge connecting the user and a second WAN edge connecting the DC.

A prototype system, composed of 200 VMs, 200 servers, and four DCs, was developed and evaluated. The evaluation results verify that the functions for reallocation of VMs between DCs and control of link aggregation can reduce power consumption of the DCs and WAN under maintained service quality. In addition, they show the possibility of energy saving by approximately 40% (under the conditions assumed in this evaluation). Moreover, they also show that power-saving control should be executed when CPU load is 75%, i.e., not when CPU load is 50% or 25%.

The power-saving cloud system will be further evaluated in the case that switches in the WAN under various CPU loads and consumed bandwidths are turned off. In addition, while aspects of QoS concerning VM access are partially evaluated in [9], power-consumption control while keeping QoS should be evaluated in detail. Besides, the prototype power-saving cloud system will be enhanced so that it can handle multiple-use cases, i.e., multiple users.

ACKNOWLEDGMENT

REFERENCES

[1] C. L. Belady, Microsoft Corporation, "Projecting Annual New Datacenter Construction Market Size," Mar. 2011 http://cdn.globalfoundationservices.com/documents/Projecting_Annual_New_Data_Center_Construction_PDF.pdf [retrieved: Jan., 2013]

[2] GreenTouch, "Our Mission," http://www.greentouch.org/index.php?page=about-us/ [retrieved: Jan., 2013]

[3] VMware, Inc. Homepage, http://www.vmware.com/ [retrieved: Jan., 2013]

[4] Xen Homepage, http://www.xen.org/ [retrieved: Jan., 2013]

[5] The Internet Engineering Task Force (IETF), "Energy Management Working Group (EMAN WG) Charter," http://datatracker.ietf.org/wg/eman/charter/ [retrieved: Jan., 2013]

[6] Institute of Electrical and Electronics Engineers (IEEE), "IEEE802.1," http://www.ieee802.org/1/ [retrieved: Jan., 2013]

[7] International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Study Group 13 (SG13), "Question 21/13 – Future Networks," http://www.itu.int/ITU-T/studygroups/com13/sg13-q21.html [retrieved: Jan., 2013]

[8] Distributed Management Task Force, Inc. (DMTF), "Cloud Management," http://www.dmtf.org/standards/cloud [retrieved: Jan., 2013]

[9] T. Suzuki et al., "Power-saving ICT platform that guarantees network bandwidth for cloud-service systems," TS-A4: Cloud Computing Technical Session, World Telecommunications Congress, Mar. 2012

[10] International Business Machines Corp. Homepage, http://www-01.ibm.com/software/analytics/spss/products/statistics/forecasting/ [retrieved: Jan., 2013]

[11] M. Yamada, T. Yazaki, N. Matsuyama, and T. Hayashi, "Power efficient approach and performance control for routers," Proc. of IEEE International Conference on Communications (ICC Workshops 2009), pp.1-5, June 2009

[12] Y. Fukuda, T. Ikenaga, H. Tamura, M. Uchida, K. Kawahara, and Y. Oie, "Performance Evaluation of power saving scheme with dynamic transmission capacity control," Proc. of IEEE Globecom Workshops 2009, pp.1-5, Nov./Dec. 2009

[13] J. Baliga, R. Ayre, K. Hinton, W. V. Sorin, and R. S. Tucker, "Energy consumption in optical IP networks," Journal of Lightwave Techology, vol. 27, no. 13, pp.2391-2403, July 2009

[14] C. Lange, D. Kosiankowski, R. Weidmann, and A. Gladisch, "Energy consumption of telecommunication networks and related improvement options," IEEE Journal of selected topics in quantum electronics, vol. 17, no. 2, pp.285-295, March/April 2011

[15] Y. Zhang, P. Chowdhury, M. Tornatore, and B. Mukherjee, "Energy efficiency in telecom optical networks," IEEE Communications surveys & tutorials, vol. 12, no. 4, pp.441-458, Fourth quarter 2010

[16] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the future internet: A servey of existing approaches and trends in energy-aware fixed nework infrastructures," IEEE Communications surveys & tutorials, vol. 13, no. 2, pp.223-244, Second quarter 2011

[17] J. Baliga, R. W. A. Ayre, K. Hinton, and R. S. Tucker, "Green cloud computing: Balancing energy in processing, storage, and transport," Proc. of IEEE, vol. 99, no. 1, Jan. 2011

[18] S. Yang, L. Chen, H. Tseng, H. Chung, and H. Lin, "Designing automatic power saving on virtualization environment," IEEE International Conference on Communication Technology (ICCT 2010), pp.966-970, Nov. 2010

[19] Y. Yao, L. Huang, A. Sharma, L. Golubchik, and M. Neely,"Data centers power reduction: A two time scale approach for delay tolerant workloads," Proc. of IEEE Infocom 2012, pp.1431-1439, Mar. 2012

[20] T. Imada, M. Sato, and H. Kimura, "Power and QoS performance characteristics of virtualized servers," Proc. of IEEE/ACM International Conference on Grid Computing, pp.232-240, Oct. 2009

Electric Vehicle Route Assistance Using Forecast on Charging Station

Marie Nestor Mariyasagayam and Yuichi Kobayashi

Information and Communication Technologies Lab.
Hitachi Europe SAS - ERD - Sophia Antipolis
Valbonne, France
e-mail: marienestor.mariyasagayam@hitachi-eu.com, yuichi.kobayashi@hitachi-eu.com

Abstract—**This paper presents a method to provide route assistance for electric vehicles by forecasting availability of charging spots in a charging station. The method uses location and reachability information of other electric vehicles along the route in order to estimate charging spot occupancy ratio at the charging station. By using such information, a risk factor for charging at a station along a planned route of an electric vehicle is calculated. The risk factor is continuously monitored, and when risk becomes high, a new route based on other charging stations located along the route is proposed to the driver. By dynamically monitoring travel routes based on charging station capacity and population of electric vehicles around the charging station, a prediction can be made about the availability of charging stations. Thus, in order to have faster travel times, electric vehicle drivers can be opportunistically advised to take appropriate routes with less crowded charging stations.**

Keywords-Electric Vehicle; Charging Station; Route Assistance; Telematics Service

I. INTRODUCTION

Electric Vehicle (EV) enables pollution-free and noiseless driving. But, they come at a cost when compared to vehicles with combustion engine: the range (distance) of driving [1]. The actual distance traveled by an electric vehicle without recharging its battery is relatively lesser when compared to driving distances of vehicles based on gasoline. It was one of the prime reasons for electric vehicles to lose the race when they were already deployed in the 19th century, even before vehicles with internal combustion engines came into consumer market [2].

In the 21st century, with growing interest from consumers, and policy makers, the EV market is beginning to see growth again. EV manufacturers are at their initial stages of deploying fully electric vehicles. Many efficient battery usage mechanisms [3], [4] are being developed in order to increase the range of an EV, and on-going collaborative research efforts [5] aim at improving charging station infrastructure in order to have faster charging. Nevertheless, a level of incertitude still exists about destination reachability [6], especially when routes involve longer distances, or even shorter routes which are troubled by traffic incidents such as traffic jam, road blockade, etc.

In the current scenario, one of the solutions prime to the successful deployment of electric vehicles on a large scale is to providing route assistance services taking into account charging stations along the traveling route. It helps to extend the actual range of an EV by planned stopovers for charging at a charging station along a travel route. As charging infrastructure deployments are rapidly growing, routing EVs via charging station has gained considerable importance in the recent years [7], [8], and has become a prime factor for wide-spread the deployment of electric vehicles.

At present, the number of electric vehicles is not so large compared to vehicles based on combustion engines [9]. Soon, this number is seen to increase steadily [10]. When the numbers of EVs increase on a route, the probability for an EV to find a spot immediately, upon its arrival at a charging station decreases. Also, there is high potential for a charging station to remain fully occupied due to increased demand for charging. Thus an EV driver may encounter one or more combination of the following problems:

- The EV driver may be made to wait in order to get charged due to queuing of electric vehicles at a charging station. Depending on the vehicle type, battery requirements, and charging infrastructure capabilities, present fast-charging infrastructure services require anywhere between 10 to 60 minutes [11] for a full battery re-charge. Thus, when compared to gasoline stations, the order of waiting time is relatively higher for an EV driver to get re-charged at a charging station. It can ultimately lead to additional delay in the driver's planned journey, in case there is a necessity for re-charging along the travel route.

- Increased charging demand at one station can lead to gradual overloading of subsequent charging stations along the route because drivers tend stop for charging at the subsequent stations when charging is not possible at the current station. As a result, an EV driver may not be able to find a suitable charging station at all for a considerable length of the planned route.

- It might also happen that the driver can be charged at a higher price per unit of electric power, e.g., a charging station may propose to charge EVs using their stand-by power during times of overloading and power shortage. This may not be appreciated by drivers who are concerned about cost-factor.

Apart from EV drivers, simultaneous charging demand from large number of electric vehicles on a particular section of route can lead to breakdown of conventional power grid. Field tests [12] and results based on analytical framework [13] show current grid architecture is unable to cope with peak charging demands especially when large numbers of EVs try to charge simultaneously via several charging stations controlled by a particular grid. This can disturb smooth functioning of electric power distribution and operation. However, to tackle such problems, large-scale research collaboration [14] are on-going to demonstrate

power grid operational stability and address issues such as power voltage, surplus electricity and frequency fluctuation.

Taking into account the above-mentioned problems, to provide route assistance to an electric vehicle through a charging station that is susceptible to overloading by surrounding electric vehicles becomes risky for destination reachability of the EV driver.

The rest of the paper is organized as follows: Section II presents the prior works that are relevant to the subject of electric vehicle route assistance. Section III describes the overall system architecture. Section IV introduces the parameters necessary for risk factor calculation. Section V provides a set of rules for deciding whether to re-route over other optimal paths. Example calculations are provided in Section VI. Section VII identifies a set of features relevant to electric vehicle route assistance mechanisms, and provides a comprehensive list of availability of such features in the route assistance methods discussed in prior works. Section VIII shows a simple evaluation of the proposed method. Finally, Section IX provides concluding remarks and further extension for future work.

II. RELATED WORK

At present, route assistance services are generally performed independently for each electric vehicle [7], [15], [16], taking into account a navigation route from a start location to a destination location traveled by the electric vehicle, considering location of charging stations along the travel route.

Kobayashi et al. [7] proposed a route search method which calculates routes with stop-overs at charging stations according to the available range of EVs. This is done to have extra battery charge when current remaining charge is not enough to reach the destination. Eisner et al. [15] considered a mathematical model for optimal routing of EVs using battery capacity constraints in order to have energy-efficient routing. This method exploits energy recuperation capacity of EVs during deceleration phases or when going downhill along the travel route. It has been observed through experiments that such mathematical model gives rise to several optimal paths especially considering battery constraints. An optimal routing method [16] based on energy consumption rather than just using shortest route to destination is proposed. The method takes estimated the energy consumption of vehicle when traveling along different routes towards a chosen destination, and selects the most economical route in terms of energy consumption. By selecting most economical routes, the necessity to re-charge at a charging station can decrease or at least the number of stop-overs along the route can be decreased relatively. Nevertheless, in order to have extended range and when distance to destination is longer than actual range possible with current battery level of an EV, re-charging remains inevitable. Thus a viable option is to calculate routes such that there are charging stations near-by in case there will be a need for re-charging.

There are other route search algorithms not specific to electric vehicles, but can be eventually applied to EVs as well. Faez et al. [17] proposed a route search algorithm based

on real-time traffic information provided by a sensor network consisting of road side terminals. Having real-time traffic information gives an idea of how long a vehicle can stay in a particular route segment under current traffic conditions. Services for EV route assistance can use such information to estimate how long an EV will approximately take, in order to travel a section of the road under actual traffic conditions. Kono et al. [18] proposed a route search system which uses information such as real-time traffic information and geometric information such as inclination of the route, then calculates optimum route in terms of not route distance but gas consumption by avoiding congested points or ascending slope. Such idea can be applied to electric vehicle, by making EVs to take routes that have relatively lesser ascending slopes or can be coupled with the idea presented by Eisner et al. [15] to increase the range of an electric vehicle.

In the prior works, it can be observed that route planning for electric vehicles is typically based on the shortest distance to the charging stations or by selecting route that consumes least amount of electric power. There is no account of other EVs planning to use the same charging station, nor is the actual availability of charging spot in a charging station, upon the arrival of an EV, considered. When many EVs are planning to use the same charging station, and if route plans are made without considering such information, it is highly likely that one or more EV users may suffer due to problems mentioned in Section I.

This paper presents a method to enable route assistance to an EV user, by taking into account the population of other electric vehicles in the vicinity of a charging station, and also by predicting the availability of charging spots in the station.

III. OVERALL SYSTEM ARCHITECTURE

The overall system is illustrated in Fig. 1. It consists of the following components:

1. A service center.
2. Charging stations.
3. Traffic centers.
4. Electric vehicles.

Service center: This is the main component that provides route assistance service to the electric vehicles. It is connected to the traffic centers, charging stations and electric vehicles through a communication network, and can exchange data with them through the network. The service center contains one or more servers that are used to provide the service. For the purpose of description, the "server of the service center" will be known as "server" in the rest of this document. The service center has access to the location information of the charging stations. It can either store such information in a database, or can request for update from the traffic centers or directly from the charging station. The service center has also location information about electric vehicles, which is received from the traffic centers and/or directly from the electric vehicles through the communication network.

It is important to note that privacy and data protection remain an open issue, and there are claims [19] that location

sensing methods are a threat to privacy of the driver participating in the service. Several debates and related developments are taking place at different levels through industry-academia partnership projects [20], [21] and government reforms [22] in order to address issues related to privacy and data protection. Until proper consensus is reached at all levels, this work assumes that such issues can be handled using known privacy-preserving [23], [24] or service subscription by having consent of the participating entities, etc.

Charging station: It contains one or more charging spots. A charging spot is used by an electric vehicle to charge its battery. The charging station provides information to the server through the communication network about current charging conditions such as maximum number of charging spots, currently occupied spots, rate at which charging is being carried out at a particular charging spot, the current battery level of the electric vehicle, the maximum battery charge capacity or the intended level of charging at the charging spot. Such information is used by the server to estimate availability of charging spots for other electric vehicles reaching the charging station.

Traffic center: This component manages road traffic related information collected from roadside sensors and/or vehicles through communication network. It is capable of providing road traffic related information to the server through the network. The traffic related information consists of collection of localization data, speed, direction of travel, estimated traveling time between geographic locations, and location of charging stations. Such information is used by the server to estimate the time required by an electric vehicle to reach a particular charging station, under actual traffic conditions. In addition, location information of electric vehicles along the route, and status of their battery charges are also provided by the traffic center to the server. This additional information is used by the server to predict likelihood of occupying a charging spot at a charging station by other electric vehicles.

Electric vehicles: These are the beneficiaries of the route assistance service provided by the service center. They are connected to the network by wired or wireless means. From the electric vehicle, the service center can receive location information, status of battery charge, and intended travel route or destination. With help of such information, the server identifies charging stations that are located along the travel route of the electric vehicle.

IV. FORECASTING OF CHARGING STATION AVAILABILITY

The principle idea behind the route assistance method is to associate a risk factor (RF) for charging at a charging station (CS) along the travel route. The RF is calculated based on charging station availability (i.e., absence or presence of charging spots), a time-to-reach factor calculated based on travel duration of other electric vehicles along the route to reach the charging station, and a ratio between the EV population traveling towards the charging station and the total EVs within a perimeter of the charging station. The server uses the current position of the ego-EV (i.e., the electric vehicle to which the route assistance service is

offered) to calculate a risk factor for charging at the next intended charging station. For its calculation it uses data from the charging station, the electric vehicle and the traffic center.
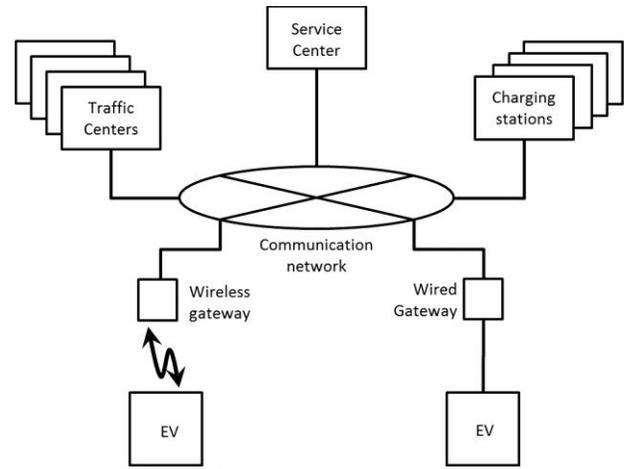
Figure 1. Overall System Architecture.

Figure 2 shows the flowchart of the route assistance process. In order to calculate the risk factor for charging at a particular station, the server needs to compute the following parameters:

- Charging station capacity (P_{CAP})
- EV population (EV_{POP} and $EV_{POP-EGO}$)
- Time-To-Reach Factor (TTRF)

In addition to the above factors, the server calculates the following factors that are used during the re-route decision making process:

- Free Charging Spots (FCS)
- Minimum Queueing Time (MQT)

The use of above parameters and the method by which they are calculated are shown from sections IV-A to IV-E.

A. Charging Station Capacity(P_{CAP})

With the help of location information received from the vehicle, the server can identify the nearest charging station along the traveling direction (S1) from its database. The server then queries the following information from the charging station (shown in steps S2 and S3 of Fig. 2):

- Maximum number of charging spots (FC)
- Currently occupied spots (OC)

Subsequently, the charging station capacity P_{CAP} (S4) can be calculated using (1):

$$P_{CAP} = OC / FC. \qquad (1)$$

P_{CAP} is inverse to the risk factor for charging. For values approaching towards 1, the availability of a charging spot in a charging station decreases. Hence, the risk factor for charging at a charging station for an EV increases.

B. Estimation of EV Population

With the help of the location information of the charging station, a perimeter is constructed. The perimeter is a midway-distance between the next charging station and the previous charging station (example available in Section VI). The server gets information of the number of EVs that are present within the perimeter (step S5), along with their respective location and moving direction. It then calculates the EVs that are located within the perimeter EV_{POP} and the number of EVs traveling towards the charging station of the ego-vehicle ($EV_{POP-EGO}$) in step S6. The population of electric vehicles has a direct effect on the charging spot availability at a charging station. If more vehicles travel towards the charging station, the risk that the charging station can get occupied is also higher. $EV_{POP-EGO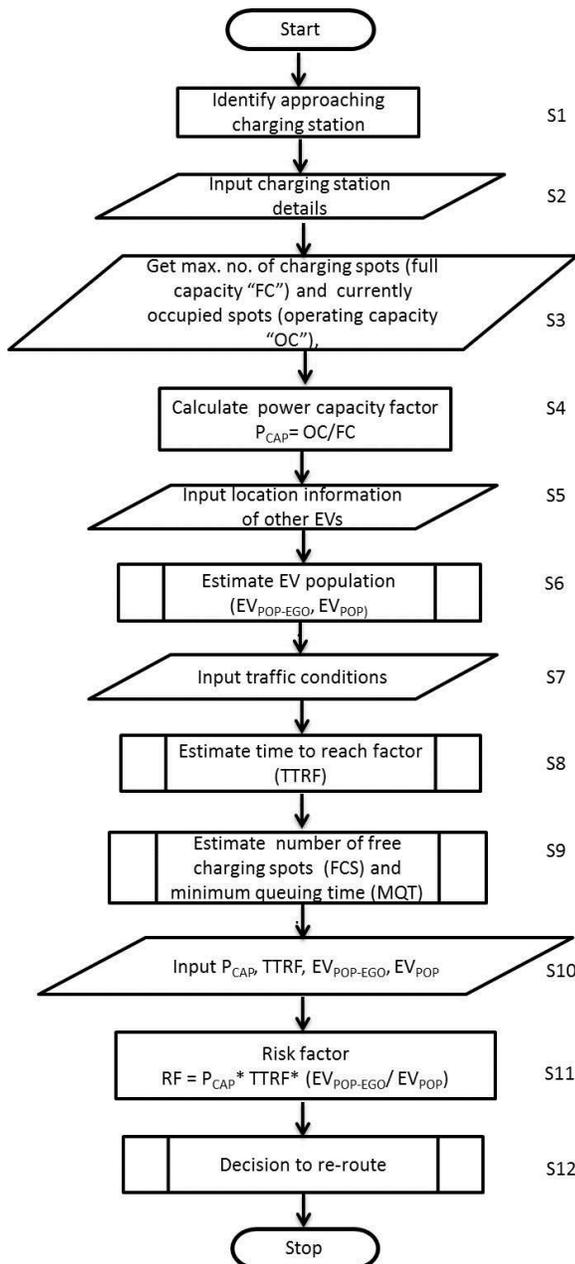}$ is the actual number of electric vehicles that travel in the direction of the charging station. The value EV_{POP} is used as a weighting factor when calculating the risk factor.

C. Time-To-Reach Factor (TTRF)

The server receives information about traffic conditions from the traffic center (step S7). It contains current traffic information of vehicles, speed of the traffic flow and the status of the route (accident, road works, etc.) along the traveling direction of the electric vehicle. The data is used to calculate the instantaneous travel time along a certain section of the planned route. This is equivalent to the time duration that an EV can take to reach a charging station along its route depending on the existing traffic conditions. Also, with the input received in step S5 on the location information of other EVs, the server estimates the time taken by other vehicles located within the perimeter and driving towards the location of charging station. This time taken to reach is called as "Estimated Time-to-Reach" (ETR). An average and the standard deviation of the ETR are calculated. Then a Time-To-Reach Factor (TTRF) is given using Table I.

TABLE I. A MAPPING OF TTRF VALUES FOR DIFFERENT ETR$_{EGO}$ VALUES BASED ON MEAN AND STANDARD DEVIATION OF ETR VALUES OF EVS.

ETR$_{EGO}$ value	TTRF
$ETR_{EGO} \leqslant \mu - 3\sigma$	1/7
$\mu - 3\sigma < ETR_{EGO} \leqslant \mu - 2\sigma$	2/7
$\mu - 2\sigma < ETR_{EGO} \leqslant \mu - \sigma$	3/7
$\mu - \sigma < ETR_{EGO} \leqslant \mu + \sigma$	4/7
$\mu + \sigma < ETR_{EGO} \leqslant \mu + 2\sigma$	5/7
$\mu + 2\sigma < ETR_{EGO} \leqslant \mu + 3\sigma$	6/7
$ETR_{EGO} > \mu + 3\sigma$	1

The logic of the table is such that for higher ETR values, a higher TTRF value is chosen, i.e., vehicles which are far away from the charging station have lower probability (higher risk) to find a charging spot because of other vehicles which may be closer to the charging station. For lower values of ETR, a lower value of TTRF is selected, which in turn provides a lower risk factor: vehicles closer to the charging station. This time factor is used as a simple measure of likelihood of occupancy of a charging spot in a charging station. Within the perimeter of a charging station, the more time it takes for an ego-EV to reach the charging station, the higher is the probability for a charging spot to be occupied by another electric vehicle, traveling in the direction of the charging station, and ahead of the ego-EV.

D. Free Charging Spots (FCS)

The server receives information about the number of charging spots occupied (S61 of Fig. 3). The information contains current charging station information and conditions such as:

- Number of charging spots (N).
- Rate of charging at Nth charging spot (R_N).
- Maximum charge required by EV (Ch_{maxN}) at Nth charging spot.

Figure 2. Route Assistance Process

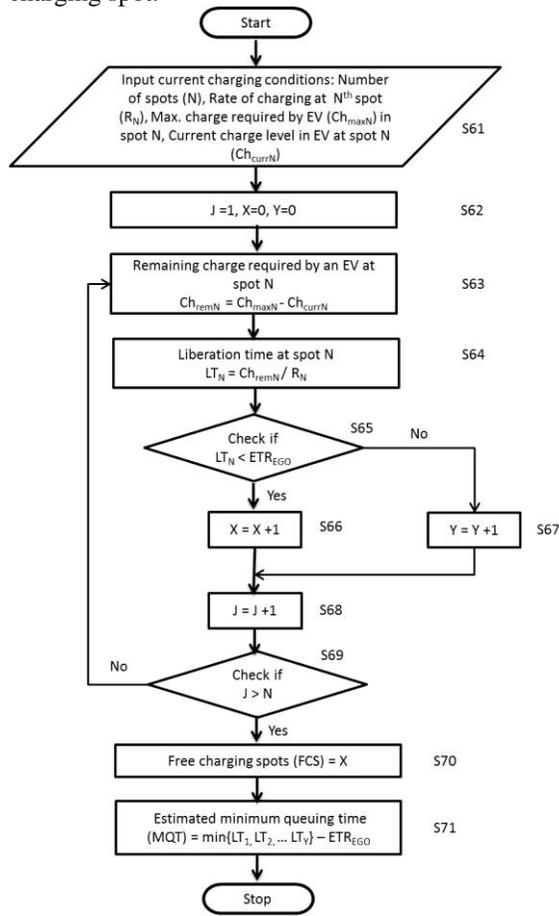- Current charge level in EV (Ch_{currN}) at Nth charging spot.

Figure 3. Estimation of Free Charging Spots (FCS) and Minimum Queuing Time (MQT).

With the above information, the server estimates a liberation time for each charging spot at the station. Liberation time (LT) is an estimate of the time duration until which a charging spot will be occupied by an electric vehicle. It is the ratio of the charge required by an electric vehicle to the rate of charging provided by the charging spot. The calculation procedure is shown from steps S62 through S69 of the flowchart in Fig. 3.

E. Minimum Queueing Time (MQT)

Assume that a charging station consists of N charging spots, and their respective liberation times are LT_1, LT_2, LT_3... LT_N. Let X be the total number of spots that have liberation time (i.e., they become free) less than the arrival time of the ego-vehicle (i.e., ETR_{EGO}), and Y be the number of spots that have liberation time greater than ETR_{EGO}.

$$MQT = \min \{LT_1, LT_2, LT_3, \ldots LT_Y\} - ETR_{EGO} \qquad (2)$$

For the ego-vehicle to queue at the time of its arrival, X=0. In order to estimate how long an electric vehicle has to

wait just after its arrival at the charging station, a set of liberation times that have values greater than ETR_{EGO} have to be formed. From this set, a minimum value is selected. The difference between the minimum value and the ETR_{EGO} gives an estimate of the minimum queuing time.

F. Risk Factor (RF)

The risk factor is calculated using parameters obtained in Sections IV-A, IV-B and IV-C. Equation (3) is used to calculate the risk factor.

$$RF = P_{CAP-EGO} * (EV_{POP-EGO}/EV_{POP}) * TTRF \qquad (3)$$

RF value lies between 0 and 1, where 0 means no risk and 1 means maximum risk. This can also be expressed in terms of percentage. The risk factor associated with the charging station is continuously monitored, and when the risk factor exceeds a given threshold, decision making process is initiated (details in Section V).

V. DECISION TO RE-ROUTE

This section presents a set of conditions that need to be checked after the risk factor goes beyond a determined threshold and before advising a new route to the driver. This is the last step (S12) of the route assistance process shown in Fig. 2. The flowchart of the decision making process is shown in Fig. 4.

The server assumes a threshold factor for risk (RF_{TH}). If the calculated risk factor is greater than threshold, a decision to re-route using new charging station is based on the following conditions (step S12 of Fig. 2):

1. The server estimates the number of electric vehicles that will reach the current charging station before ego-EV. If this number is greater than the estimated free charging spots, then it can be known whether there will be any immediate spot for charging at the arrival of the ego-EV.
2. The location of subsequent charging station CS_{NEW} in the traveling direction of the ego-EV is identified (S84). The current battery level of the EV is used to check reachability to the newly identified charging station in S85. A new time to destination is calculated via CS_{NEW}. From this it can be known whether the electric vehicle is able to reach the new charging station if proposed.
3. Then it is checked whether the sum of MQT and time to destination via current charging station, is greater than the new time to destination via charging at CS_{NEW} (and its MQT_{NEW} if required). From such comparison, it can be known whether passing by new charging station is faster than by just using the current charging station (even though it involves a minimum waiting time). This condition check is also used to void routing vehicles through CS_{NEW}, when waiting times at CS_{NEW} are larger than the current station (i.e., already many vehicles are being queued at CS_{NEW}).
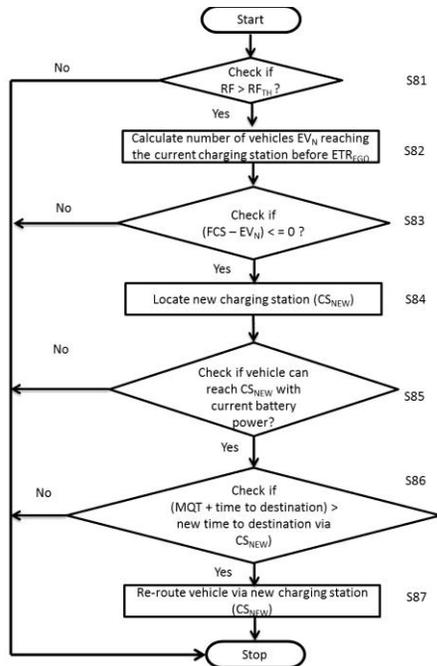
Figure 4. Deciding to re-route.

When conditions 1, 2 and 3 are true, the user is informed of new route by the server through the communication network.

VI. ROUTE ASSISTANCE SERVICE EXAMPLE

In this section, an example of route assistance service is shown along with a sample calculation for risk factor using four electric vehicles: ego-EV, EV_1, EV_2 and EV_3 as shown in Fig. 5. Only vehicles ego-EV, EV_1, EV_2 are moving in the direction towards the charging station CS_2, which is the charging station being targeted by ego-EV.
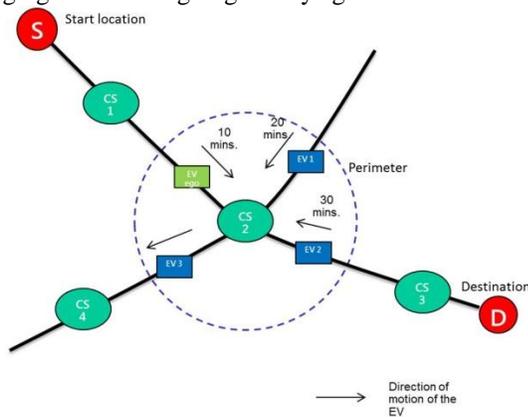

Figure 5. Example calculation of Time-To-Reach Factor (TTRF)

Assume ETR of ego-EV, EV_1 and EV_2 are 10, 20 and 30 minutes respectively, in order to reach CS_2. Then the TTRF value for ego-EV can be calculated according to explanation in Section IV-C and Table I as follows:

$$TTRF = 3/7 \qquad (4)$$

where, $\mu = 20$, $\sigma = 8.16$, $\mu - 2\sigma < ETR_{EGO} \leqslant \mu - \sigma$

Assume that CS_2 has 3 charging spots out of which 2 are occupied by EVs for charging, then charging station capacity calculated according to (1) as follows:

$$P_{CAP-EGO} = 2/3 \qquad (5)$$
$$EV_{POP} = 4 \qquad (6)$$
$$EV_{POP-EGO} = 3 \qquad (7)$$

Then, the risk factor calculated according to (3) will evaluate to

$$RF = 0.214 \qquad (8)$$

In other words, the risk percentage for charging at CS_2 for the ego-vehicle is about 21% when it reaches the charging station in 10 minutes.

In the example used, the time to arrival of EV_{EGO} (10 minutes) < EV_1 (20 minutes) < EV_2 (30 minutes). Table II gives risk percentages calculated for all vehicles charging at CS_2 for various power capacity levels (P_{CAP}) at the charging station.

TABLE II.   RISK PERCENTAGE OF EV_{EGO} EV_1 AND EV_2 FOR VARIABLE P_{CAP} VALUES AT CHARGING STATION CS_2

Vehicle	$P_{CAP} = 0$	$P_{CAP} = 1/3$	$P_{CAP} = 2/3$	$P_{CAP} = 1$
EV_{EGO}	0%	11%	21%	32%
EV_1	0%	15%	28%	43%
EV_2	0%	18%	35%	54%

Assuming a threshold of 50% for risk, the server does not initiate any route re-calculation neither for EV_{EGO} nor for EV_1. But, when EV_2 is considered, for the case when $P_{CAP-EGO}$ is 1 (i.e., when all charging spots are occupied at CS_2), the risk percentage is more than 50%. So, for EV_2, the server identifies a subsequent charging station along the travel route, and re-calculates route towards the destination using the newly identified charging station. Then a decision is made whether to propose this new route to the EV_2 driver based on the following conditions (as explained in Section V):

1. Can EV_2 reach the next charging station with the current battery capacity (i.e., whether vehicle able to reach newly chosen charging station)?
2. If the estimated number of available charging spots at CS_2 are more than the number of electric vehicles reaching the charging station before EV_2 (i.e., whether vehicle is unable to have a free charging spot upon arrival)?
3. Whether passing via another charging station is faster than passing via current station, taking into account the minimum queuing time required?

When the above conditions evaluate to true, the new route is advised to the driver.

VII. SUMMARY OF FEATURES RELATED TO EV ROUTE ASSISTANCE

This section summarizes a list of features that are relevant to electric vehicle route assistance, and the

advantage of using such features in optimizing route search for electric vehicles. Table IV shows a listing of the features, their corresponding advantage in optimization of route search for electric vehicles, and their availability in the method M1 described in this paper, and other methods M2 [7], M3 [15] and M4 [16] presented in the prior works under Section II. From Table IV, it can be observed that some features are also used for route assistance of non-electric vehicle as well, and hence are not new. But, given a limited range of electric vehicles, these features have much more impact, and a stronger role to be play while providing route assistance. In general, the features used for optimization target mainly two metrics for efficient route assistance: (i) travel time and (ii) energy consumption. Methods 1 and 2 focus on the first metric in order to minimize travel time when passing by charging stations. Methods 3 and 4 focus on the second metric in order to minimize energy expenditure while traveling along a route. For a given a set of routes, methods 3 and 4 prioritize route selection based on the estimate of the total amount of energy spent while traveling along the route. When battery level in an EV is such that there is enough range to reach destination, methods 1, 2, 3 and 4 can be conveniently used for route assistance of the electric vehicle. When several routes are available with enough battery range for each route, methods 1 and 2 can provide the shortest route towards destination, whereas methods 3 and 4 can select the route with lowest energy consumption.

On the other hand, when routes are similar (i.e., in terms of energy consumption), and when the distance to destination is longer than range that could be achieved with current battery level of an EV, there is no option left, but to re-charge. In such situation, it is advantageous to consider route navigation passing by charging stations. As previously mentioned in the introductory section, the order of waiting times in a charging station are relatively larger for an EV when compared to vehicles based on combustion engines. When charging is imminent to continue an onward journey, route planning via charging stations becomes a viable option. By forecasting and continuous monitoring of the charging stations along the route, drivers can be advised to take different optimal routes in order to pass by charging stations with lower or zero waiting time.

VIII. EVALUATION OF THE PROPOSED METHOD

This section evaluates the proposed route assistance method using travel time and energy consumption metrics. For the purpose of a simple hypothetical evaluation, consider Fig. 6. There are three possible routes to a destination D from start position S of an electric vehicle. They are:

- **Route 1**: S→C1→D
- **Route 2**: S→C2→D
- **Route 3**: S→C3→D

C1, C2 and C3 are the charging stations that are located along the routes. Routes 1 and 2 partly include passage via hilly areas where the energy expenditure for a vehicle can become higher when compared to traveling over a flat plain. Each of the routes is divided into segments based on the

location of charging stations. Traveling via each segment on a particular route will incur a certain amount of energy expenditure (E), segment travel time (ST), charging time (CT) and queuing time (QT) in a charging station.
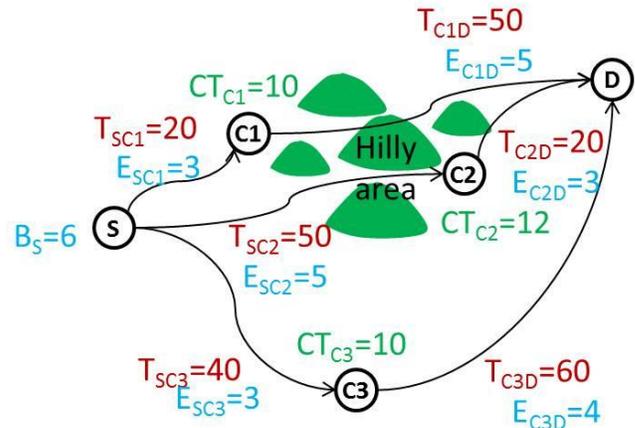
Figure 6. Evaluation of proposed method using energy expenditure and travel time metrics

The total travel time for each route is equivalent to the sum of the segment travel times, charging times and the queuing times at the charging station.

$$\text{Travel Time TT} = \sum ST_n + \sum CT_x + \sum QT_x \qquad (9)$$

where 'n' is the number of segments and 'x' is the number of charging stations.

The energy expenditure for each route can be calculated by summing up the individual energy expenditures of each segment for that particular route.

$$\text{Energy Expenditure EE} = \sum E_n \qquad (10)$$

Using (9) and (10), the total travel time and the energy expenditure for each of the routes are calculated respectively for the case presented in Fig. 6, and are shown in Table III. It is assumed that the initial available charge (B_s) is at least enough to reach a charging station along the given route. In this evaluation, time is expressed in minutes and energy expenditure in kWh. Two cases are considered: Case A: Without knowledge of QT; Case B: With knowledge of QT.

TABLE III. EVALUATION RESULT BY PROPOSED METHOD.

	Case	Route #	QT (min)	TT (min)	EE (kWh)
A	When QT = 0 or no queuing time is considered	1	0	80	8
		2	0	82	8
		3	0	110	7
B	When QT is considered	1	5	85	8
		2	2	84	8
		3	5	115	7

TABLE IV SUMMARY OF FEATURES RELATED TO EV ROUTE ASSISTANCE AND THEIR AVAILABILITY IN DIFFERENT METHODS
(O: AVAILABLE; X: NOT AVAILABLE)

Feature	Advantage	M1	M2 [7]	M3 [15]	M4 [16]
Charging spot availability	Enables to predict if an EV requires queuing when reaching a charging station, and eventually choose stations without any queues in order to minimize travel delay.	O	X	X	X
Queuing time in a charging station	Minimize queuing delay in a charging station (when queuing cannot be avoided). When several charging stations are present along the route of an EV, this feature allows selecting the charging station with least amount of queuing time.	O	X	X	X
Shortest travel time to charging station (including time to charge)	Allows having faster route towards destination.	O	O	X	X
Real-time traffic information	Enables to estimate extra energy consumption due to idle times and traffic incidents (e.g., traffic jam) along the route, and thereby allows predicting charging necessity or route diversion according to situation.	O	X	X	X
Include routes based on lesser EV population	Allows route traveling via possibly less-crowded charging stations, and as a result an EV may find a charging spot easily upon its arrival at a station along the travel route which is sparsely populated with EV traffic.	O	X	X	X
Shortest path to charging station	Allows having lesser total traveling distance at the end of the journey, in case a vehicle has to deviate from its route in order to find a charging station, when no suitable charging station is available along the route.	X	O	X	X
Battery range check	Enables appropriate selection of charging stations (for methods M1 and M2) or selection of appropriate route segment (for method M3) when multiple routes with several charging stations are available towards a destination.	O	O	O	X
Use descending routes	EVs can recuperate energy by using downhill slopes to augment battery charge in order to extend their range.	X	X	O	O
Avoid ascending routes	Driving uphill on a steep road requires higher power from the engine, which increases vehicle energy consumption. By avoiding such routes, energy consumption can be reduced.	X	X	O	O
Select lowest energy consumption route	When several routes are possible towards a destination, selecting a route with lowest energy consumption can reduce demand for re-charging.	X	X	O	O

Considering case A, methods M1 and M2 will select route 1 (i.e., the route with the least travel time) and methods M3 and M4 will select route 3 (i.e., the route with least energy consumption).

Considering case B, since method M1 is aware of the queuing time at the charging stations, it will select route 2, method M2 will still select route 1 because it does not have knowledge about the queuing time at the charging stations, and methods M3 and M4 will select route 3.

Thus by knowing queuing time in addition, the proposed method is able to provide a more optimal route in terms of travel time.

IX. CONCLUSION AND FUTURE WORK

This paper presented a method to provide route assistance for electric vehicles taking into account the availability of a charging station and other EVs that are susceptible to use the charging station. An approaching charging station, along the route of an electric vehicle, was associated with a risk factor for charging based on: EV population within a perimeter of the charging station, estimated arrival times of the EVs to the charging station, and the availability of charging spots in the station. By continuous monitoring of the estimated risk factor, alternative routes or charging stations were advised to the EV driver when risk percentage to charge at a station became greater. This was done to reduce delay in the EV driver's journey when a stopover is required for getting charged. Thus, by predicting risk percentages for charging at a station,

EV drivers can be advised to take alternative routes that have lesser risk percentages, and which are comparatively faster for destination reachability. In Section I, the pros and cons of electric vehicle were discussed and route assistance via charging stations were shown as a promising approach to speed-up the deployment of EVs on a large scale. Some of the problems that an EV driver can encounter while driving via charging station were also discussed. Existing methods for route assistance were discussed as prior works in Section II. Section III showed the overall architecture needed to realize the route assistance service. The parameters necessary for calculating risk factor, and the decision making process for new route selection were explained in Section IV, and Section V, respectively. In Section VI, exemplary risk factor calculation, and the conditional checks that were required to re-route an electric vehicle were shown. For the reader's digest, a tabular summary of features that are used in optimization and their advantages were presented in Section VII. Observing the list of features, two principal metrics were identified for route assistance of EVs: travel time and energy consumption. EV route assistance methods can use these two metrics to trade-off and optimize route search for electric vehicles. Based on the two metrics, a simple evaluation of the proposed method was shown in Section VIII.

A detailed comparison of the proposed method with other route assistance methods using real time data is reserved for future work.

REFERENCES

[1] J. Larminie and J. Lowry, "Electric vehicle technology explained," Wiley, 2012.

[2] E.H. Wakefield, "History of the Electric Automobile," The Society of Automobile Engineers, SAE International Publishing, 1994.

[3] N. Kutkut, H. Wiegman, D. Divan, and D. Novotny, "Design considerations for charge equalization of an electric vehicle battery system," IEEE Transactions on Industry Applications, 1999, vol. 35, no. 1, pp. 28–35.

[4] L. Lam and R. Louey, "Development of ultra-battery for hybrid-electric vehicle applications," Journal of power sources, vol. 158, no. 2, August 2006, pp. 1140–1148.

[5] NissanNews.com. (September) GE, Nissan sign R&D agreement to fast track broader adoption of electric cars. http://www.nissan-global.com/EN/. Retrieved: January 2013

[6] T. Franke, I. Neumann, F. Bühler, P. Cocron, and J. Krems, "Experiencing range in an electric vehicle: Understanding psychological barriers," Applied Psychology: An International Review, 2011, doi: 10.1111/j.1464-0597.2011.00474.x

[7] Y. Kobayashi, N. Kiyama, H. Aoshima, and M. Kashiyama, "A route search method for electric vehicles in consideration of range and locations of charging stations," in IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, 2011, pp. 920–925.

[8] J. Ferreira, A. Silva, and J. Afonso, "EV-Cockpit–Mobile Personal Travel Assistance for Electric Vehicles," Advanced Microsystems for Automotive Applications, Springer Berlin Heidelberg, 2011, pp. 247–257, doi: 10.1007/978-3-642-21381-6_24

[9] U. Eberle and R. von Helmolt, "Sustainable transportation based on electric vehicle concepts: a brief overview," Energy & Environmental Science, 2010, vol. 3, no. 6, pp. 689–699.

[10] M. Tran, D. Banister, J. Bishop, and M. McCulloch, "Realizing the electric-vehicle revolution," Nature Climate Change, 2012, vol. 2, no. 5, pp.328–333.

[11] C. Botsford and A. Szczepanek, "Fast charging vs. slow charging: Pros and cons for the new age of electric vehicles," The International Electric Vehicle Symposium and Exposition, EVS24, Stavanger, Norway, May 2009, pp. 1-9.

[12] W. Kempton, et. al, "A test of vehicle-to-grid (V2G) for energy storage and frequency regulation in the PJM system," Results from an Industry-University Research Partnership, University of Delaware, Pepco Holdings, Inc PJM Interconnect, and Green Mountain College, November 2008.

[13] J. Taylor, A. Maitra, M. Alexander, D. Brooks, and M. Duvall, "Evaluation of the impact of plug-in electric vehicle loading on distribution system operations," In Power & Energy Society General Meeting, July 2009, PES'09, IEEE, pp.1-6, doi: 10.1109/PES.2009.5275317

[14] Hitachi Press Release, "U.S. and Japan Companies Collaborate on Smart Grid Project in Hawaii - Hitachi to Serve as Project Leader", May 17, 2011. http://www.hitachi.com/New/cnews/110517b.html. Retrieved: January 2013.

[15] J. Eisner, S. Funke, and S. Storandt, "Optimal route planning for electric vehicles in large networks," in Proc. 25th Assoc. Advancement Artificial Intell. Conf., San Francisco, CA, August, 2011. https://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/view/3637/4005. Retrieved: January 2013.

[16] A. Artmeier, J. Haselmayr, M. Leucker, and M. Sachenbacher, "The optimal routing problem in the context of battery-powered electric vehicles," in Workshop: CROCS at CPAIOR-10, Second International Workshop on Constraint Reasoning and Optimization for Computational Sustainability, Bologna, Italy, June 2010.

[17] K. Faez and M. Khanjary, "Utospf: A distributed dynamic route guidance system based on wireless sensor networks and open shortest path first protocol,". ISWCS'08. IEEE International Symposium on Wireless Communication Systems, 2008, pp. 558–562.

[18] T. Kono, T. Fushiki, K. Asada, and K. Nakano, "Fuel consumption analysis and prediction model for eco route search," in 15th World Congress on Intelligent Transport Systems and ITS America's Annual Meeting, New York, November 2008, pp. 1-9.

[19] B. Parno and A. Perrig, "Challenges in securing vehicular networks." In Workshop on Hot Topics in Networks (HotNets-IV), College Park, MD USA, November 2005, pp. 1-6.

[20] SeVeCom (Secure Vehicular Communication) http://www.sevecom.org/. Retrieved: January 2013

[21] PRESERVE: Preparing Secure V2X Communication Systems http://www.preserve-project.eu/. Retrieved: January 2013

[22] European Commission, "How will the EU's reform adapt data protection rules to new technological developments?", January 2012. http://ec.europa.eu/justice/data-protection/document/review2012/factsheets/8_en.pdf Retrieved: January 2013

[23] P. Papadimitratos, L. Buttyan, T. Holczer, E. Schoch, J. Freudiger, M. Raya, Z. Ma, F. Kargl, A. Kung, and J-P. Hubaux. "Secure vehicular communication systems: design and architecture," Communications Magazine, IEEE, 2008, vol. 46, no. 11, pp. 100-109.

[24] R. Lu, X. Lin, H. Zhu, P. H. Ho, and X. Shen, "ECPP: Efficient conditional privacy preservation protocol for secure vehicular communications," In INFOCOM 2008, The 27th Conference on Computer Communications, IEEE, 2008, pp. 1229-1237.

Green Service Discovery Protocol in Mobile Ad Hoc Networks

Janine Kniess

Santa Catarina State University (UDESC)

Joinville - SC, Brazil

janine@joinville.udesc.br

Orlando Loques, Célio V. N. Albuquerque

Fluminense Federal University (UFF)

Niterói - RJ, Brazil

loques@ic.uff.br, celio@ic.uff.br

Abstract—**Energy efficiency is an important aspect of green computing. In order to achieve green computing in service discovery protocols for MANETs, it is essential for researchers to investigate energy efficiency mechanisms able to prevent collisions and interferences, which both represent a waste of channel bandwidth and energy. In this paper, we propose a service discovery protocol that applies a data aggregation scheme in intermediate nodes to reduce messages replies in these networks. The results show that this mechanism makes it possible to maintain the trade off between the discovery success rate and the message reduction, hence, minimizing the energy consumption.**

Keywords-energy; discovery; selection services.

I. INTRODUCTION

A mobile ad hoc network (MANET) is a self-organizing and dynamically reconfigurable wireless network which, operates in the absence of a fixed infrastructure [1]. In MANETs, most mobile devices have limited processing capabilities and operate using limited peak voltage batteries, which imposes a stringent limitation on the instantaneous energy consumption they can suport [2]. In addition, wireless communications are prone to collisions and interferences due to the broadcasting nature of the radio medium. In this context, careful transmission mechanisms prevent collisions and interferences, which both represent a waste of channel bandwidth and energy [3]. Some of the issues identified in MANETs are service discovery, mobility management and energy consumption. Service discovery is an essential component for the usability of such networks on the grounds that service discovery enables devices to use their functions to automatically locate network services.

The extremely dynamic nature of MANETs has motivated the development of their application in different scenarios like MANETs in assisting emergency missions and taxi service. In such applications, the service provider's location is not previously known. As a consequence, a common approach to disseminate requests is to use broadcast mechanisms [4]. This is often the case where multiple providers can offer the specified service. In this case, a data aggregation approach can be used to minimize the reply transmission cost in network.

This paper proposes a location-aware service discovery protocol for MANETs composed of an autonomic, dis-

tributed and location-aware service selection mechanism that uses a data aggregation scheme in intermediate nodes to combine the replies coming from different providers, minimizing the number of transmissions, thus improving the network performance and saving energy. Also the protocol provides a service discovery mechanism that adjusts a search area for each individual request in order to save energy.

In applications such as MANETs in assisting emergency missions, the maximum response time to attend one request is essential to guarantees that service discovery is successful. Thus, we identified a number of aspects as being critical to a successful outcome: the geographic location where the service provider is being requested; the maximum response time (we suppose that the provider needs to arrive at the place where the service is required within a maximum time); the speed at which the service provider moves; and the number of providers that must be requested. The proposed protocol takes into account the above mentioned aspects in the service discovery process.

Briefly, our contributions are the proposals of: (1) a service selection mechanism that applies a data aggregation scheme in intermediate nodes to filter the responses aiming to reduce the exceeding replies and save energy; (2) a service discovery mechanism that adjusts a search area for each individual request; and (3) a service invocation mechanism that specifies how the service providers will be accessed and used by the requester node.

This paper is organized as follows: Section 2 presents the motivation scenario and the main related proposals regarding service discovery for MANETs; Section 3 introduces the Location Aware Discovery Protocol (LADP); Section 4 describes the evaluation and the results and, finally, Section 5 presents the final conclusion and suggests further research directions.

II. APPLICATION CONTEXT AND RELATED WORK

Service discovery in MANETs has emerged as a useful application in scenarios where the service providers are mobile and must be efficient and capable to attend a request within a maximum time, for instance, MANETs in assisting emergency missions.

A. Application Scenario

In this context, suppose an area that has been struck by natural disaster. In this case, we believe that the distribution of a specialized rescue team (formed by vehicles, robots and humans beings, for example), interconnected by a wireless mobile ad hoc network could contribute greatly to the rescue of survivors.

In the proposed scenario, some of the mobile elements of the rescue team provide a resource, i.e., an ambulance, a robot with the ability to access places that are not reachable by man, such as areas with a contamination risk, or even a human being transporting medication. The nodes that provide services are the providers and those that request the service are the requesters. We assume that each node in the network is aware of its geographic position by means of a localization system, such as a Global Positioning System (GPS), and the time in the devices is synchronized. For instance, is possible that a node identifies the existence of a leak of a given type of gas and, from specialized algorithms, the node can derive the need for a provider, such as a fire engine, which could be at the location in a matter of minutes. After identifying which resource is needed, the node will send a request message to the network searching for the appropriate resource.

B. Related Work

Service discovery is the action of finding and locating a service in the network [5] and service discovery protocols can be generally classified into three categories [4]. In chronological order we mention: fixed network protocols such as Jini [6] and Service Location Protocol (SLP) [7]; wireless single−hop networks such a Distributed Embedded Application Platform space (DEAPspace) [8] and Bluetooth [9]; and wireless multi−hop networks. Neither the first category (fixed network protocols), nor the second one (wireless single−hop networks) are feasible for decentralized environments such as the one presented in the scenario above. In these environments, it is possible to create multi-hop networks. The following protocols stand out in this last category: Konark Gossip [10], FTA (Field Theoretic Approach) protocol [11], the P2P Discovery Protocol (P2PDP) [12] and the cross-layer approach of Varshavsky et al. [13].

Although considerable previous research has been done into service discovery for distributed systems, few protocols have considered the development of distributed aggregation responses in the selection phase in order to save energy. Service selection is the phase that comes after service replies are gathered by the service requester.

The approaches that have been proposed by Lenders et al. [11], Gomes et al. [12] and Varshavsky et al. [13] provide a mechanism for autonomic service selection. However, energy consumption issues are not considered in these works.

The FTA approach proposed by Lenders et al. [11] is based on the theory of electrostatic fields. Requests to an instance of a given service type are routed selectively in the direction of the provider that generated the highest field gradient. However, this approach does not scale well when different types of services are available. In Varshavsky et al. [13] the selection service function is integrated into the routing mechanism in a cross-layer approach. Moreover, the selection occurs only when the replies arrive at the requester node. In Gomes et al. [12] it shows a suppression vicinity service selection mechanism that can discard exceeding replies during the reply transmission. However, this mechanism filters response messages through the reverse path, i.e., messages are routed by a reverse path traversed by the request. This concept can easily fail in MANETs due to highly dynamic topologies.

Unlike the mentioned approaches, our work also considers that the providers must move as far as the place where the resource is needed. The other approaches consider that the services can be remotely accessed and used.

III. THE PROPOSED SERVICE DISCOVERY PROTOCOL

In this section, we will firstly describe how the Location Aware Discovery Service (LADS) [14] can save energy in MANETs. Secondly, we will present the Location Aware Service Selection (LASS) scheme to aggregate replies in intermediate nodes, and to select the best providers during the reply transmission. Table 1 summarizes the notation used in this paper.

A. Location Aware Discovery Service (LADS)

This section describes the strategy adopted by the service discovery mechanism, LADS to save energy in service discovery for MANETs.

LADS works as follows. Suppose a node in the network needs information about service providers and sends discovery messages. The LADS mechanism limits the search diameter R_i, on the basis of the maximum speed that a node can reach, v_{max} (each type of resource knows this value), and the maximum response time for one request, Δ_{tmax}. Using R_i, this mechanism prevents unnecessary request and reply transmissions in the network. The diameter R_i is given by the equation:

$$R_i \leftarrow v_{max} \times \Delta_{tmax} \tag{1}$$

The mechanism defines the diameter using v_{max} so that the search area includes the greatest number of apt providers. Given the pair (i,j), being i the requester and j the provider, it is assumed that the speed (v_j) of the latter is known.

After sending a message, the requester node starts the timer, $\Delta_{trequester}$. This timer defines the maximum amount of time during which the requester will wait for responses, and this time is proportional to the diameter of the request. If a requester does not receive responses within this time, it will resend the discovery message. $\Delta_{trequester}$ is given by,

Table I. SUMMARY OF NOTATION

Symbol	Definition
$vmax$	maximum node speed
Δ_{tmax}	maximum response time for one request
R_i	search diameter for requester node
$n_{provider}$	reply counter
$max_{provider}$	maximum number of responses specified by the requester
$\Delta_{trequester}$	client node timer
$\Delta_{tintermediate}$	intermediate node timer
β	maximum time for the closest intermediate node to store responses before forwarding them

$$\Delta_{trequester} \leftarrow K \times (R_i/R_t) \qquad (2)$$

The value of K is defined as $\alpha + \beta$. The value of α specifies the forward and backward delay for one hop on the network. The value of β specifies the maximum amount of time that the intermediate node closest to the requester can store one response before forwarding it. The range of the antenna is given by R_t, and R_i/R_t represents an estimate of the hops number. The closest intermediate nodes maintain the replies stored for longer time in compared with more distant nodes.

In the discovery message, the requester node sends the following information: a node identification, its coordinates, $coord_X$, $coord_Y$, the maximum response, Δ_{tmax}, the service, s, and the number of desired providers. If node j receives a request from node i, the algorithm verifies the distance (d_{ij}) between both nodes. If $d_{ij} > R_i$, the request is discarded by j because this node is out of the search area. Conversely, if $d_{ij} \leq R_i$, the algorithm verifies the speed of the service provider j. Moreover, the algorithm verifies if this provider offers the searched resource (s) and if the service provider j is available at the moment. If the restriction given by Eq. (3) is satisfied, node j sends a response to node i.

$$d_{ij}/v_j \leq \Delta_{tmax} \qquad (3)$$

Assuming that the provider node j has the resource, but v_j is insufficient, j does not send a reply to i, and it only resends the request message. It is assumed that the nodes have maximum speeds defined. The nodes attend to the requests on demand. Each node maintains information about the resource that is offered by it.

B. Location Aware Service Selection (LASS)

As a result of the discovery process, multiple providers can respond to a service request. The LASS mechanism takes into account such aspects as the requester node's geographic location, the maximum response time to attend one request, the speed that the service provider moves, and the number of service providers desired to select and discard answers. The aim is to reduce the number of reply transmissions from the network.

LASS works as follows. Suppose that node k (intermediate node) receives the reply message from one of its neighbors, for instance, node m. Then, node k starts a timer, named $\Delta_{tintermediate}$, with the function of storing replies. $\Delta_{tintermediate}$ is given by Eq. (4).

$$\Delta_{tintermediate} \leftarrow \gamma \times (1/d_{ik}) \leq \beta \qquad (4)$$

where γ specifies a proportionality factor with the function to maintain the timer $\Delta_{tintermediate}$ less than β. $1/d_{ik}$ denotes the inverse distance between the requester and the provider node. Since the storage period is inversely proportional to the distance between the intermediate node and the requester, the closest nodes store replies for a longer time. For instance, in Fig. 1, $\Delta_{tintermediate}$ of node $I2$ is greater than $\Delta_{tintermediate}$ of node $I1$. This occurs due to the fact that $I2$ is closest to the requester node. The LASS data fusion scheme is shown below.

C. LASS Data Fusion Scheme

When an intermediate node receives one response, the intermediate node starts $\Delta_{tintermediate}$. If the maximum number of responses ($max_{provider}$) that meet the requests profile is reached before $\Delta_{tintermediate}$ expires, the intermediate node aggregates these responses and sends only one response to the requester node. After this step, $\Delta_{tintermediate}$ is canceled by the intermediate node. Others responses received are discarded.
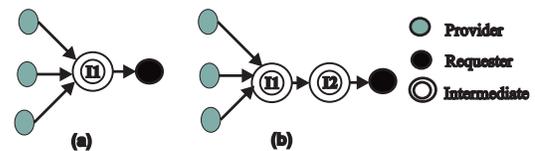

Figure 1: Lass data fusion scheme without aggregation (Figure. 1a) and with aggregation (Figure. 1.b)

In the data fusion scheme, an intermediate node collects responses from different providers, and concatenates two individual parameters in one string, which are $Id_{provider}$ and $ServiceTime_{provider}$. The other parameters(id_{Client}, $id_{Requester}$, $id_{SeqRequester}$, $max_{provider}$) are common for all replies. In the aggregation phase, all parameters are aggregated in one response and transmitted to the requester node. In the response message ($msgReply()$), a provider sends its geographic coordinates, the client identification, the resource identification, the request identification, the sequence number identification, the maximum number of replies, $Id_{provider}$ and the $ServiceTime_{provider}$.

An intermediate node can receive a response directly from a provider or another intermediate node. If the intermediate node receives a response directly from a provider, the response was not aggregated. In this case, while the maximum number of responses is not reached, the intermediate node

stores responses. If the intermediate node receives a response from another intermediate node, the LASS algorithm works as summarized in Fig. 1.

In Fig. 1, let us suppose the intermediate node *I2* receives one response from the intermediate node *I1*. Firstly, *I2* will verify the number of aggregated responses in this message. After this step, *I2* verifies the number of responses stored by it. If $max_{provider}$ was not reached, *I2* stores the responses.

The number of stored responses plus the number of aggregated responses may not exceed $max_{provider}$. In order to solve this problem, the LASS algorithm updates the stored responses on the condition that the aggregated responses have better quality than the stored responses.

The LASS mechanism is presented in Algorithm 1. It is noteworthy that the repeated answers (already treated by a particular provider) are discarded.

Algorithm 1. Service Selection Algorithm

PROCWAITMAXPROVIDER()
1 node k verifies the number of aggregated replies (*number$_{agregate}$*) in *msgReply()*;
2 node k verifies the number of replies already received by it (*number$_{provider}$*);
3 $number_{totalresponses} \leftarrow number_{agregate} + number_{provider}$;
4 **if** $number_{totalresponses} < max_{provider}$
5     $number_{provider} \leftarrow number_{totalresponses}$;
6   node k stores the news responses;
7 **else if** $number_{totalresponses} = max_{provider}$
8   node k aggregates received replies;
9   node k sends *msgReply()*;
10   node k cancels $\Delta_{intermediate}$;
11 **else**
12   node k discards *msgReply()*;

PROCENDTIMERWAITMAXPROVIDER()
1 **if** $\Delta_{intermediate}$ expires
2   node k aggregates received replies;
3   node k sends *msgReply()*;

D. Service Invocation Mechanism

The service invocation mechanism operates after the selection service phase and establishes rules for access and use of the providers already selected. In the service invocation phase, requesters and providers verify the viability of the attendance and the providers physically move to the place where the service is required.

As we can see in Fig. 2a, after the selection service phase represented by (2), the requester node sends a service invocation message (3) to the selected provider (Fig. 2b). Then, the selected provider sends a service confirmation message (4) in which the provider informs that it is available go to the place where the service is required.

In the service confirmation message, the provider sends the following information: request identification ($id_{Request}$),
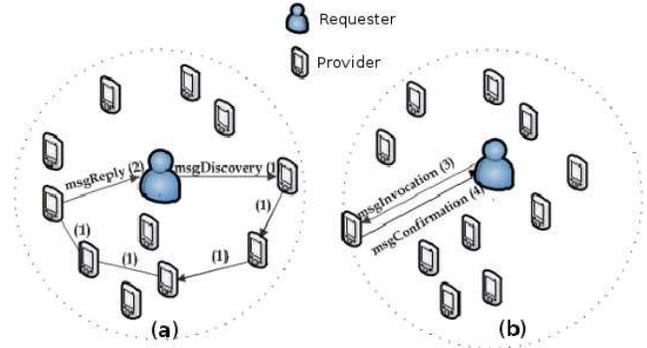

Figure 2: Transition diagram of the messages in the discovery process

provider identification ($id_{Provider}$), and requester node identification (id_{Client}). When a provider informs that it is available go to the place where the service is required it begins to move toward the requester. However, the confirmation message may be lost. In order to attend this fault, we considered that if the provider receives a discovery message of the same requester, the provider restarts the process and sends a reply message (*msgReply*) to the requester. Also, in this case, the provider can respond to other requesters. A provider is only allocated to a requester after sending the confirmation message.

IV. SIMULATION RESULTS

In order to evaluate the proposed approach, we used the Network Simulator (NS-2) [15] and the scenario presented in Section II was mapped for different area sizes. Specifically, 4km^2 (2000m x 2000m) and 25km^2 (5000m x 5000m). The number of nodes in the network varies between 100, 200, 300 and 400, and the node speed varies between 0.5 m/s and 15.0 m/s.

Experiments were carried out where resources were attributed to 10%, 15%, 20%, 25% and 30% of the nodes in the network. Each provider offers one type of resource. A random distribution of resources was conducted. In the simulation we used the Optimized Link State Routing (OLSR) protocol, [16] to send the service discovery messages to the neighbor nodes.

All scenarios started with a 3600 second initialization phase. The minimum number of providers that must be delivered could be 1 and 2, and the maximum response time was set at 10.0 min. All nodes in the network are mobile. However, the requester node remains static while waiting for the reply.

We conducted experiments with four mechanisms: LADS (discovery without selection), LASS+Fusion (discovery with selection and aggregation), Flooding-1 (traditional flooding), and Flooding-2. In Flooding-2, we extracted from the responses obtained with Flooding-1, the responses of apt nodes, that is, the providers that can reach in time where the

resource is needed. The confidence interval presented in the results is 95%.

Each node is equipped with the default wireless network module that NS-2 provides, which has a 250m transmission range. We adjusted the transmission power at 42mW and the receiving power at 55mW. The initial energy of all the nodes is set at 1000J. Each node makes requests during the 28800s duration of the simulation.

A. Results

Fig. 3 shows the service invocation success rate (SI) according to the node's mobility. The service invocation success determines whether or not a client discovers a service after the service invocation phase. SI is defined as number of responses received after the service invocation phase divided by number of requests sent. We measured the service invocation success rate for mechanisms: LADS, LASS+Fusion, Flooding-1, and Flooding-2. In this experiment, 10% of the participating nodes have the resource and the number of providers that must be delivered is 1.
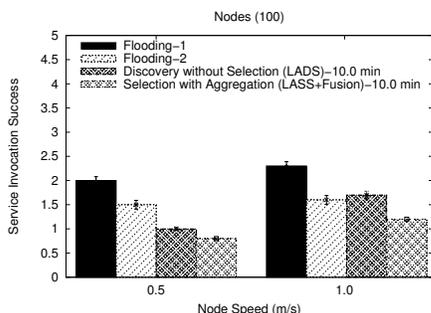

Figure 3: Service invocation success × node speed

The results show that LASS+Fusion performs well for all scenarios. However, we observed that the SI rate obtained with LASS+Fusion when the node speed is 0.5 m/s was less than 1. This result is due to loss of invocation or confirmation messages.

Energy consumption (EC) experiments were performed. In terms of energy consumption, we observed that the LASS+Fusion mechanism outperforms the other evaluated mechanisms without compromising the discovery process. EC is defined as total energy consumption divided by initial energy of the nodes.

In Fig. 4, we analyzed the node energy consumption (EC). For this experiment, we compared the LASS+Fusion mechanism with the LADS mechanism and Flooding-1. The number of nodes in the network is 100. The resource percentage is 10%. It is noteworthy that the Flooding-2 mechanism was not mentioned in this simulation because in Flooding-2, we extracted from the responses obtained with Flooding-1, the responses of apt nodes, that is, the providers that can reach in time where the resource is needed.
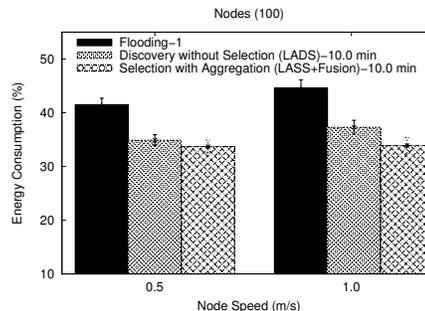

Figure 4: Energy consumption × node speed

From these results, we observed that the LASS+Fusion mechanism saves considerable energy by discarding responses with a longer time limit. For instance, in a scenario where the node speed is 0.5 m/s, the energy consumption was 41.5% with Flooding-1, 34.9% with LADS and 33.7% with LASS+Fusion.

Through this experiment, we concluded that the use of data aggregation for response suppression in service discovery protocols can reduce the energy consumption of network nodes and increase the network lifetime.
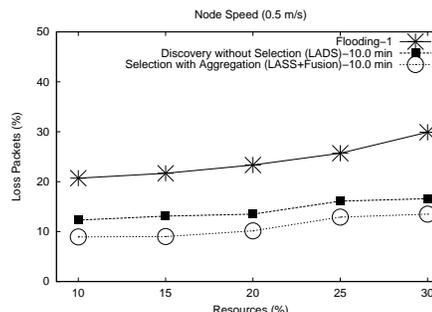
Figure 5: Lost packets × resource percentage

Fig. 5 depicts the loss packets rate (LK) after the service invocation phase. This metric is given by number of lost packets divided by total number of messages generated by LADP. As we can see, LASS+Fusion presented low packet loss in relation to Floding-1. For instance, when 30% of the participating nodes have the resource, the loss rate was 30.90% for Flooding-1, and 13.5% for LASS+Fusion. This result shows that to aggregate replies can also contribute to decrease collisions and avoid packet loss.

In Table II, we present the percentage of success (PE) after the service invocation phase. PE is based on the replies obtained from the Exact Approach.

We implemented the Exact Approach to verify the mobility scenarios, and to show the best provider(s) in each time instant for the requester node. For this experiment, we calculated the PE among the responses obtained with LADS, LASS+Fusion and Flooding-2, with the responses presented

by the Exact Approach.

In this experiment, 10% and 30% of the participating nodes have the resource. The minimum number of providers that must be delivered is 1 and the node speed is 0.5 m/s.

TABLE II. PERCENTAGE OF SUCCESS(%) × NODE SPEED

Resource (%)	Flooding-2	LADS	LASS+Fusion
10%	75.0	70.0	70.0
30%	69.0	75.5	80.52

The results show that the PE with Flooding-2 was 69%, and 80.5% for LASS+Fusion when 30% of the participating nodes have the resource. The PE of the Flooding-2 was lower for scenarios where nodes have more resources. This occurred due to collisions that increased dropped response messages. This exemplifies that reply aggregation can greatly benefit the percentage of success by reducing the number of reply messages and saving energy.

Table III shows the energy consumption related to the number of nodes. Checking Table III, the economy of energy was greater with LASS+Fusion. In a scenario with 400 nodes, the energy consumption of Flooding-1 was 86.92% and with LASS+Fusion was 53.98%.

We conclude that the LASS+Fusion mechanism outperforms the other mechanisms with respect to node energy consumption through the discards of additional responses without compromising the discovery process. In this experiment, 30% of the participating nodes have the resource and the node speed is 0.5 m/s.

TABLE III. ENERGY CONSUMPTION(%) × NODE

Node	Flooding-1	LADS	LASS+Fusion
300	86.87	69.97	53.84
400	86.92	70.10	53.98

Also, were performed experiments with larger areas in order to evaluate the scalability of the protocol. In the experiments presented below, the scenario presented in Section II was mapped for an area of 25 km^2. The number of nodes in the network varies between 50, 100, 150, 200 and 250. The node speed varies from 0.5 m/s to 15.0 m/s and the maximum time limit to service the request is 1.5 min and 15 min.

Fig. 6 shows the service invocation success rate (SI) according to the node's mobility. We measured the service invocation success rate for mechanisms: LADS, LASS+Fusion, Flooding-1, and Flooding-2. In this experiment, 10% of the participating nodes have the resource, the number of providers that must be delivered is 1, the network has 200 nodes and the maximum time limit to service the request is 1.5 min.

The LASS+Fusion mechanism has enabled a greater response suppression without causing a negative impact on the discovery process, i.e., it has reached the goal of 1 (one)
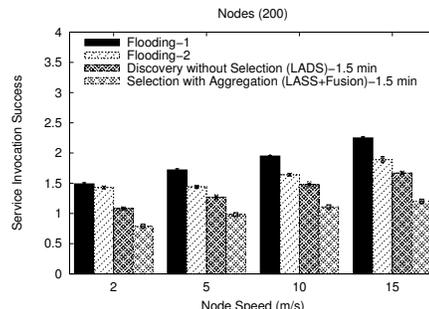

Figure 6: Service invocation success (1.5 min) × node speed

response at all speeds, except for the scenario where nodes move at 2.0 m/s. In this case, the service invocation success rate was 0.79. This behavior is due to the low number of providers within the radius. Besides that, there was not a substantial change in the topology in order for additional providers to be found.

In Fig. 7, we measured the service invocation success rate (SI) according to the node's mobility for the same scenario shown in Fig. 6. However, unlike previously presented scenario, the maximum time limit to attend the request is increased to 15 min and the node speed varies between 0.5 m/s, 1.0 m/s and 1.5 m/s. Therefore, in this scenario, the radius R$_i$ remains the same as in Fig. 6.
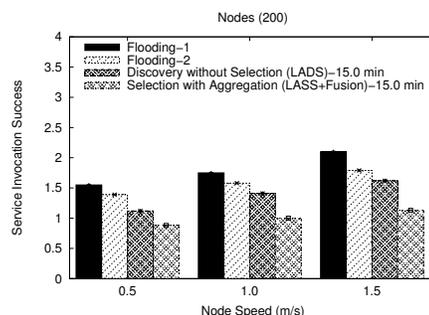
Figure 7: Service invocation success (15 min) × node speed

The results obtained were similar to those presented in Fig. 6. These results indicate that the performance is mainly a function of the radius R$_i$ and that a longer attending time limit can be reached if nodes move at proportionally lower speeds.

V. CONCLUSION AND FUTURE WORK

This paper presented the Location Aware Service Discovery Protocol for MANETs and its mechanisms for service discovery, LADS, and a mechanism for service selection, LASS. The service selection mechanism presented a data aggregation scheme in intermediate nodes.

The results showed that the LASS+Fusion mechanism saves considerable energy through the additional response

discards without compromising the discovery process. The proposed mechanisms help to maintain the battery of the mobile devices longer.

Another contribution is a service invocation mechanism. This phase is generally not considered in service discovery protocols, however, it is crucial to ensure the attendance success.

We observed that the use of the fault-tolerant service discovery mechanism by means of the $\Delta_{trequester}$ timer contributed to keep the service invocation rate within the expected average despite of the existence of faults.

The results of simulations on the Location Aware Service Discovery Protocol showed that the LASS+Fusion mechanism can provide a good scalability and searching performance in different scenarios.

As future works we would like to extend our store and forward strategy (e.g., understand the trade-off between network scalability and the period of time to keep the message stored in a node).

REFERENCES

[1] R. Balaji, H. Jarmo, and S. Bilhanan, "Service discovery framework for manets using cross-layered design," in Proc. 6th international conference on wired/wireless internet communications (WWIC'08), Tampere, Finland, 2008, pp. 152–163.

[2] Q. Li and M. Zhou, "The survey and future evolution of green computing," in Proceedings of the 2011 IEEE/ACM International Conference on Green Computing and Communications (GREENCOM '11). Washington, DC, USA: IEEE Computer Society, 2011, pp. 230–233.

[3] A. Bianzino, C. Chaudet, D. Rossi, and J. Rougier, "A survey of green networking research," Communications Surveys Tutorials, IEEE, vol. 14, no. 1, pp. 3–20, 2012.

[4] A. N. Mian, R. Baldoni, and R. Beraldi, "A survey of service discovery protocols in multihop mobile ad hoc networks," IEEE Pervasive Computing, vol. 8, no. 1, pp. 66–74, 2009.

[5] R. S. Marin-Perianu, "Wireless sensor networks in motion: clustering algorithms for service discovery and provisioning," Ph.D. dissertation, University of Twente, Enschede, 2008. [Online]. Available: http://doc.utwente.nl/60028, [retrieved: 08, 2012].

[6] K. Arnold, The Jini specification, ser. Jini technology series. Addison Wesley, 1999.

[7] E. Guttman, "Service location protocol: Automatic discovery of ip network services," IEEE Internet Computing, vol. 3, no. 4, pp. 71–80, 1999.

[8] M. Nidd, "Service discovery in deapspace," Personal Communications, IEEE, vol. 8, no. 4, pp. 39–45, 2001.

[9] B. Miller and C. Bisdikian, Bluetooth revealed. Prentice Hall PTR, 2001.

[10] S. Helal, N. Desai, V. Verma, and C. Lee, "Konark - a service discovery and delivery protocol for ad-hoc networks," in IEEE Wireless Communications and Networking Conference (WCNC'03), 2003, pp. 2107–2113.

[11] V. Lenders, M. May, and B. Plattner, "Service discovery in mobile ad hoc networks: A field theoretic approach," Pervasive and Mobile Computing, vol. 1, no. 3, pp. 343–370, 2005.

[12] A. T. A. Gomes, A. Ziviani, L. dos S. Lima, M. Endler, and G. Chelius, "Mitigating reply implosions in query-based service discovery protocols for mobile wireless ad hoc networks," in The 11th International Conference on Ad Hoc Networks and Wireless (ADHOC-NOW), vol. 5198. Springer, 2008, pp. 29–42.

[13] A. Varshavsky, B. Reid, and E. de Lara, "A cross-layer approach to service discovery and selection in manets," in The 2nd IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, 2005, pp. 459–466.

[14] J. Kniess, O. Loques, and C. V. N. Albuquerque, "Location aware discovery service and selection protocol in cooperative mobile wireless ad hoc networks," in Proceedings of the 28th IEEE international conference on Computer Communications Workshops (INFOCOM'09). Piscataway, NJ, USA: IEEE Press, 2009, pp. 367–368.

[15] T. Issariyakul and E. Hossain, An Introduction to Network Simulator NS2. Springer, 2009. [Online]. Available: http://books.google.com.br/books?id=cD69He_oU60C, [retrieved: 12, 2012].

[16] T. Clausen and P. Jacquet, "Rfc3626: Optimized link state routing protocol (olsr)," RFC Editor United States, 2003.

Power Line Communication using STC/SFC/STFC over Statistical Indoor Power Line Channels

Jeonghwa Yoo and Sangho Choe

Wireless Communications Laboratory, The Catholic University of Korea

Bucheon, Republic of Korea

mundade@hanmail.net and schoe@catholic.ac.kr

Abstract— **This paper presents a MIMO-OFDM based power line communication (PLC) system using space-frequency coding (SFC), space-time coding (STC), and space-time-frequency coding (STFC) over statistical indoor power line channels. We present a maximum ratio combining (MRC) scheme that provides both multiple antenna diversity gain and multipath fading diversity gain. By simulation, we verify that the proposed MRC is effective to both 2×2 MIMO and SISO over indoor power line channels. On the simulation, we consider both perfect and imperfect channel estimation. We also simulate SFC, STC, and STFC over power line channels, with or without crosstalk between antenna paths; STFC and STC are more robust to crosstalk than SFC; STFC with some added complexity overpowers the other coding schemes in terms of bit-error rate performance. Further, STFC is less sensitive to impulse noise index *A* than STC and SFC.**

Keywords— *PLC; MIMO; OFDM; MRC; multipath fading; impulse noise; crosstalk.*

I. INTRODUCTION

Smart grid that is the renewable energy based future power line network will provide various kinds of power line access services (PAS). For that smart grid PAS, power line communication (PLC) that does not request a separate backbone network unlike other medium has become one of the main alternatives for high-speed bidirectional information exchange among electric power providers, electricity industries, and consumers. Moreover, since an international PLC standard, IEEE 1901 [1], was adopted in 2010, there has been a growing interest in various other PLC applications including home networks and emergency backup networks. PLC is available at a low cost because it does not require any additional infrastructure; further, it is ubiquitous because it is available anywhere where there is electricity and is easy to access with a plug-in power cable.

PLC that uses existing power lines, which are originally installed for power supply, has very poor channel environment such as limited carrier frequency ($f_c < 100$MHz) by power cable attenuation, multipath fading, and impulse noise.

The signal processing model of PLC multipath fading channel is classified into two types: deterministic-type [2] and statistical-type [3], [4]. In the case of deterministic-type, the well-known Zimmermann model has been widely used [2] but it does not consider practical channel statistics. Hence, there are several trials of statistical-type channel models, which are mostly modified versions of Zimmermann model, such as Galli model [3], and Tonello model [4]. Tonello model first takes the uniformly-distributed reflection factors in fading paths into account but it does not show any empirical justification yet (see [3]). Therefore, throughout this paper, we use the Galli's statistical channel model to get the most realistic simulation results for our 2×2 multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) PLC system. We also employ Middleton class A model [5] for the statistical impulse noise sample generation considering the impulse characteristics, which is raised by switching the power supply, the power converters, and so on.

OFDM is widely used in power line channels since it is robust to inter-symbol interference (ISI) caused by the multipath fading delay spread [6]. In recent studies, MIMO using the spatial diversity, turbo coding, and low-density parity-check (LDPC) coding are actively investigated to reduce the transmit bit error rate (BER) and improve the performance of OFDM PLC systems [7], [8]. In this paper, via indoor power line, we implement the three different MIMO coding schemes: space-frequency coding (SFC) [8] using the frequency space diversity gain, space-time coding (STC) [8], [10] using the time space diversity gain, and space-time-frequency coding (STFC) [8] using the time and frequency space diversity gain.

By simulation, we evaluate and compare SFC/STC/STFC based 2×2 MIMO-OFDM PLC systems with the proposed maximum ratio combing (MRC) scheme over statistical PLC channels. Whereas conventional MIMO-OFDM PLC [6], [8] just considers antenna MRC (AMRC) to

obtain spatial diversity gain, the presented MIMO-OFDM PLC employs antenna & fading MRC (AFMRC, also called rake receiver) that effectively combines both multiple antenna (in PLC, a pair of power line conductors forms one antenna port) and multipath fading diversity gain. Conventional MIMO PLC in [8] assumes the ideal antenna paths with no crosstalk, but the proposed MIMO PLC takes crosstalk between antenna paths into consideration.

In this paper, for the theory and analysis of the proposed MIMO-OFDM system model, we consider both perfect channel estimation and imperfect channel estimation. For the simulation of imperfect channel estimation, we employ a least square estimator (see [9]), which is a simple but well-established channel estimator.

Simulation results verify that the presented MRC scheme is superior to the conventional scheme, whether or not crosstalk between antenna channels exists. This scheme improves the BER performance, not only in the 2×2 MIMO, but also in the single-input single-output (SISO); note that SISO just uses fading MRC (FMRC) rather than AMRC. In simulation, we also evaluate the proposed MIMO PLC when the impulse noise index A and the ratio of impulse to Gaussian noise variance τ vary. The contributions of this paper are summarized as follows:

- Propose a SISO-/MIMO-OFDM PLC using a rake receiver;
- Fully evaluate the proposed system over statistical power line channels with fading, impulse noise, crosstalk between antenna paths;
- Analyze the proposed system with both perfect and imperfect channel estimator.

Section II describes the proposed system model including PLC channel characteristics, crosstalk in MIMO channels, SFC/STC/STFC based MIMO-OFDM, and least square channel estimation for MIMO-OFDM. Section III details simulation results of the proposed system and finally, Section IV concludes the paper.

II. System Model

A. Impulse Noise and Fading Channel in PLC

A PLC channel can be characterized with both impulse noise and multipath fading, due to multiple signal reflections caused by power line impedance mismatch. First, for impulse noise, we use Middleton class A model [5], which *pdf* (probability density function) is defined as:

$$p_X(x) = \sum_{m=0}^{\infty} e^{-A} \frac{A^m}{m!} \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{x^2}{2\sigma_m^2}}, \qquad (1)$$

$$\sigma_m^2 = \sigma^2 \frac{m/A+\tau}{1+\tau}, \qquad (2)$$

where $\sigma^2 = \sigma_G^2 + \sigma_I^2$ (σ_G^2 is the Gaussian noise variance and σ_I^2 is the impulse noise variance), $\tau = \sigma_G^2 / \sigma_I^2$, and A is the impulse index.

Second, in a PLC channel, there are two types of channel models; deterministic-one like Zimmermann model [2], and statistical-one like Galli model [3]. First, the transfer function of Zimmermann channel model [2] at the jth antenna path is expressed as:

$$H_j(f) = \sum_{l=1}^{L} H_{j,l}(f), \qquad (3)$$

$$H_{j,l}(f) = g_{j,l} \cdot e^{-(\alpha_0 + \alpha_1 \cdot f^u)d_{j,l}} \cdot e^{j2\pi f(d_{j,l}/v_p)}, \qquad (4)$$

where L is the number of fading paths per antenna path. α_0, α_1, and u are the power line cable parameters, and $|g_{j,l}| \leq 1$ is the weighting factor of the jth antenna and lth fading path [2]. $d_{j,l} / v_p$ is equivalent to the corresponding path delay $T_{j,l}$ (where $d_{j,l}$ represents its length) as follows:

$$T_{j,l} = \frac{d_{j,l} \cdot \sqrt{\varepsilon_r}}{c_0} = \frac{d_{j,l}}{v_p}, \qquad (5)$$

where ε_r is the non-insulation dielectric constant of the cable and c_0 is the speed of light.

Typically, each OFDM subcarrier has flat (constant) frequency channel characteristics due to its narrow bandwidth such that the frequency selective fading channel transfer function of (3) can be translated (digitized) and approximated as follows:

$$H_j(f)\big|_{f=f_c+k\Delta f} \cong H_j(k) = \sum_{l=1}^{L} H_{j,l}(k), \qquad (6)$$

where f_c is the carrier frequency (which is herein assumed to indicate the lower limit of the OFDM bandwidth (BW)), Δf is the subcarrier spacing, and $k (= 0, 1, \ldots, N-1$, where N is the number of subcarriers) is the frequency index. For the time being, operating under the assumption that a good channel estimator, such as a least square method [9] or a pilot assisted method [10], is employed, we assume the ideal channel estimate for all fading paths, i.e. $H_{j,l}(k) = \hat{H}_{j,l}(k)$.

In this paper, for practical MIMO channel simulation, we consider Galli model [2], a modified version of Zimmermann model, where the average of channel gain and delay (spread) in (4), is log-normal distributed. Since the delay is easily dealt with using a cyclic prefix (CP) of OFDM system, the proposed system model just considers the average channel gain at the jth antenna path and lth fading path, defined as

$$\bar{G}_{j,l} = 10^{\bar{G}_{j,l}(dB)/10} = \frac{1}{N} \sum_{k=0}^{N-1} |H_{j,l}(k)|^2. \qquad (7)$$

In simulation, $\bar{G}_{j,l}(dB)$ is assumed to be real, independent, and log-normally distributed in [-1, 1]. The other parameter values such as α_0, α_1, u, and so on, are assumed to be fixed (as seen in Table I, Section III).

B. Crosstalk in MIMO channels

In MIMO channels, there exists crosstalk between antenna paths; note that each antenna path in a MIMO PLC system is formed with a pair of power line conductors. This crosstalk may cause the capacity loss of the MIMO PLC system such that it is not negligible, especially for $f_c \geq 25$ MHz [6]. The 2×2 MIMO channel matrix $\mathbf{H}(k)$ with non-zero crosstalk terms, indicating the ith transmit jth receive antenna path gain $H_j^i(k) \neq 0$ (where $I \neq j$), can be expressed as follows:

$$\mathbf{H}(k) = \begin{bmatrix} H_1^1(k) & H_2^1(k) \\ H_1^2(k) & H_2^2(k) \end{bmatrix} \qquad (8)$$

Let the channel capacity with or without crosstalk be denoted as C_{ct} and C_{nct}, respectively. The capacity-loss ratio (CR) by crosstalk can be defined as [6]

$$\mathrm{CR} = \frac{C_{nct} - C_{ct}}{C_{nct}} \times 100\%, \qquad (9)$$

where the channel capacity is [14]

$$C = \sum_{k=0}^{N-1} \int_{f_c+k \cdot \Delta f}^{f_c+(k+1)\Delta f} \log_2 \det\left(\mathbf{I}_I + \frac{S_{psd}(f)\mathbf{H}(k)\mathbf{H}^H(k)}{N_{psd}(f) \cdot I} \right) df. \qquad (10)$$

$S_{psd}(f)$ and $N_{psd}(f)$ represent the transmitted signal power spectral density and colored noise (impulse plus Gaussian noise) power spectral density, respectively. I is the number of transmit antennas, $\mathbf{I}_I$ is the identity matrix of size I, and $(\cdot)^H$ refers to the Hermitian of $(\cdot)$.

C. SFC/STC/STFC based MIMO-OFDM PLC System

We implement a SFC/STC/STFC based 2×2 MIMO-OFDM PLC system. In a MIMO PLC system, since a pair of electrical wires is converted into a single antenna channel, the number of transmitting and receiving antennas is typically limited to one for 1-phase 2-wire and two for 1-phase 3-wire (including one wire for common ground (or protective earth)). Therefore, MIMO-OFDM is used with a 1-phase 3-wire power line, whereas SISO-OFDM is mostly used with a 1-phase 2-wire power line. This 2×2 MIMO system has two antenna paths that consist of a single antenna path formed with two wires and another antenna path made of other two wires.

In the OFDM transmitter, the kth (= 0, 1, …, N-1) subcarrier modulation signal, $S(k)$, experiences the following inverse fast Fourier transform (IFFT):

$$s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) e^{j2\pi nk/N}, \qquad (11)$$

where $s(n)$ is the nth (= 0, 1, …, N-1) time sample and N is the number of subcarriers.

First, in the case of a SFC based 2×2 MIMO, whose encoder diagram is shown in Fig. 1(a), we can obtain the spatial and frequency diversity which is used to reduce the error probability caused by the interference in the MIMO channel.
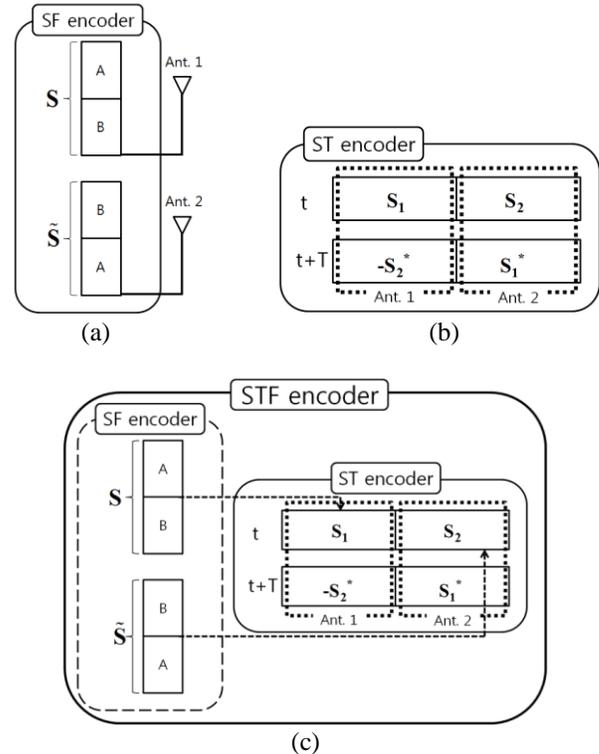

Figure 1. (a) 2×2 SFC encoder (b) 2×2 STC encoder (c) 2×2 STFC encoder.

The following two SF encoder vectors $\mathbf{S}$ and $\tilde{\mathbf{S}}$ are formed by arranging the same subcarrier signal samples in an appropriate order (i.e., vector $\tilde{\mathbf{S}}$ is the circular-shifted version of $\mathbf{S}$ [8]) for this SF encoder.

$$\mathbf{S} = [S(0),...,S(\frac{N}{2}-1),S(\frac{N}{2}),...,S(N-1)]^T, \qquad (12)$$

$$\tilde{\mathbf{S}} = [\tilde{S}(\frac{N}{2}),...,\tilde{S}(N-1),\tilde{S}(0),...,\tilde{S}(\frac{N}{2}-1)]^T, \qquad (13)$$

where $S(k) = \tilde{S}(k), (k=0,1, …, N\text{-}1)$ and $(\cdot)^T$ refers to the transpose of $(\cdot)$. $\mathbf{S}$ and $\tilde{\mathbf{S}}$ are respectively converted to the corresponding time sample vectors, $\mathbf{s}=IFFT\{\mathbf{S}\}$ and $\tilde{\mathbf{s}}=IFFT\{\tilde{\mathbf{S}}\}$, through the IFFT process (see (11)) and then transmitted to the receiver via each antenna path. The cyclic prefix (CP) is added to the OFDM modulated sample vectors ($\mathbf{s}$, $\tilde{\mathbf{s}}$) before their transmission, in order to prevent inter-symbol interference (ISI) due to the maximum delay spread $\mathrm{T}_{max} = \max_l \mathrm{T}_l$. This transmission process occurs at the modulator and SFC encoder, and its corresponding reception process takes place at the SFC decoder and demodulator [11]. The received signal via the power line channel goes through the SFC decoding process at the SFC receiver, including the fast Fourier transformation (FFT) (OFDM demodulation)

and MRC process, to recover its data stream after removing the added CP.

Whereas conventional MIMO schemes in [8] use AMRC to achieve diversity gain by multiplying their optimum weights to different antenna paths, the proposed MIMO system employs AFMRC (also called rake receiver), a combined technique of AMRC and FMRC. The used scheme has one finger (one receiver) for each fading path to achieve the FMRC gain. The receiver complexity of the AFMRC based MIMO increases L-fold as a result of adding FMRC. At the receiver, the received signal via the jth (=1, 2) receive antenna lth (=1, 2, …, L) fading path after the FFT process (which is the reverse process of the IFFT in (11)) is given as

$$Y_{j,l}(k) = \sum_{i=1}^{I}\sqrt{\frac{E_s}{I}}H_{j,l}^{i}(k)S_i(k) + N_{j,l}(k), \qquad (14)$$

where E_s represents the average energy of the transmit signal and $S_i(k)$ is the transmit signal from the ith transmission antenna. $N_{j,l}(k)$ indicates the noise component that is the result of the FFT operation of the time axis impulse plus Gaussian noise signal $n_{j,l}(n)$ with variance σ^2 (see (1)). At the rake receiver, the received signal in (14) is first translated as

$$\overline{Y}_j(k) = \sum_{l=1}^{L}Y_{j,l}(k)\hat{H}_{j,l}^{j*}(k), (j=1,2), \qquad (15)$$

where we can see that all fading path signals multiplied with corresponding channel estimate conjugates are combined. Before the jth antenna received signals in (15) are added together, note that they pass through the reverse process of the circular-shifted process in (12) and (13). Hence, by the SFC decoding process, we derive

$$\hat{Y}(k) = \sum_{j=1}^{2}\tilde{Y}_j(k), \qquad (16)$$

where $\tilde{Y}_j(k)$ is the reverse circular shifted version of $\overline{Y}_j(k)$. Finally, by the maximum likelihood (ML) based maximum ratio combining (called AFMRC) scheme, the recovered symbol is obtained as $\hat{S}(k) = \underset{S\in\Omega}{\arg\min}|\hat{Y}(k)-S|^2$, where Ω indicates the total signal symbol space.

Second, in the case of a STC based 2×2 MIMO, whose encoder diagram is shown in Fig. 1(b), we can obtain the spatial and temporal diversity to improve the system performance. The two transmission vectors in this STC transmitter, $\mathbf{S_1}$ = { $S_1(k) | k = 0,1, …, N$ - 1} and $\mathbf{S_2}$ = { $S_2(k) | k = 0,1, …, N$ - 1} are independent each other and (·)* refers to the conjugate of (·). In Fig. 1(b), we observe that at time t, $\mathbf{S_1}$ and $\mathbf{S_2}$ are assigned to antenna 1 and antenna 2, respectively, and at time $t+T$ (where T is the OFDM symbol time), $-\mathbf{S_2^*}$ and $\mathbf{S_1^*}$ are assigned to antenna 1 and antenna 2, respectively.

At the receiver, the received signal via the jth (=1, 2) receive antenna lth (=1, 2, …, L) fading path after the FFT process is given as

$$Y_{j,l}(k,t) = \sum_{i=1}^{2}\sqrt{\frac{E_S}{2}}H_{j,l}^{i}(k,t)S_i(k,t) + N_{j,l}(k,t), \qquad (17)$$

$$Y_{j,l}(k,t+T) = \sum_{i=1}^{2}\sqrt{\frac{E_S}{2}}H_{j,l}^{i}(k,t+T)S_i(k,t+T) + N_{j,l}(k,t+T), \qquad (18)$$

where $Y_{j,l}(k,t)$ and $Y_{j,l}(k,t+T)$ are the received signals at the time t and $t+T$, respectively. At the rake receiver, the received signal in (17) and (18) is first translated as

$$\overline{Y}_j(k,t) = \sum_{l=1}^{L}Y_{j,l}(k,t)\hat{H}_{j,l}^{j*}(k,t), \qquad (19)$$

$$\overline{Y}_j(k,t+T) = \sum_{l=1}^{L}Y_{j,l}(k,t+T)\hat{H}_{j,l}^{j*}(k,t+T), \qquad (20)$$

where we can note that all fading path signals multiplied with corresponding channel estimate conjugates are combined. Then, by the STC decoding process, we derive

$$Y_1'(k) = \overline{Y}_1(k,t) + \overline{Y}_2^*(k,t+T), \qquad (21)$$

$$Y_2'(k) = -\overline{Y}_1^*(k,t+T) + \overline{Y}_2(k,t), \qquad (22)$$

Finally, at the ML based AFMRC process, we can recover the following two symbols from the two derived signals in (21) and (22),

$$\hat{S}_1(k) = \underset{S\in\Omega}{\arg\min}|Y_1'(k)-S|, \qquad (23)$$

$$\hat{S}_2(k) = \underset{S\in\Omega}{\arg\min}|Y_2'(k)-S|. \qquad (24)$$

Third, in the case of STFC based 2×2 MIMO, it extends the signal space of STC to the frequency axis such that we can achieve the frequency diversity besides the spatial and temporal diversity; its encoder has a serially-combined structure of SFC and STC [8], as shown in Fig. 1(c). In the STFC transmitter, the input of STC is the two output vectors of SFC encoder, i.e., $\mathbf{S}$ and $\tilde{\mathbf{S}}$ (see (12) and (13)).

At the STFC receiver, first, by the STC decoding process of (19), (20), (21), and (22), we obtain $Y_1'(k)$ and $Y_2'(k)$. Then, by the SFC decoding process, including the reverse operation of circular shift and combining, we derive

$$Y''(k) = \sum_{j=1}^{2}\tilde{Y}_j(k), \qquad (25)$$

where $\tilde{Y}_j(k)$ is the reverse circular shifted version of $Y_j'(k)$ ($j=1,2$). Finally, at the ML based AFMRC process, the recovered symbol is obtained as $\hat{S}(k) = \underset{S\in\Omega}{\arg\min}|Y''(k)-S|^2$.

D. Least square Channel estimation for MIMO-OFDM system

In this paper, we use a least square estimator for the channel estimation. Least square channel estimation [9] is

performed after the FFT operation at the receiver. For the channel estimation, the pilot signals which are already known at the receiver are used. In the OFDM system, adjacent G subcarriers are grouped without overlapping. In this paper, we assume that each group has almost the same channel. In each group, the first subcarrier is the pilot subcarrier. Therefore, the total $M\ (=N/G)$ pilot subcarriers are used for channel estimation. In [9], they assume that all the pilot signals have an equal complex value c. The transmitted signals on the kth subcarrier are expressed as

$$S_i'(k) = S_i'(mG+g) = \begin{cases} c, & \text{for } g=0 \\ \text{Information Data}, & \text{otherwise} \end{cases} \quad (26)$$

where $m = 0, 1, \ldots, M$-1 and $g = 0, \ldots, N/G$-1 (for each m).

The received signal can be expressed as

$$Y_{j,l}(k) = \sqrt{\frac{E_s}{2}} H_{j,l}^j(k) S_i'(k) + N_{j,l}'(k), \quad (27)$$

where N' is the crosstalk plus noise term (assume that the crosstalk term is Gaussian distributed). Since all the pilot signals have c value, M estimated channels at the each pilot subcarriers can be written as

$$\hat{H}_{j,l}^j(mG) = \frac{Y_{j,l}(mG)}{c} = H_{j,l}^j(mG) + \frac{N_{j,l}(mG)}{c}. \quad (28)$$

Finally, the channel estimate will be expressed as

$$\hat{H}_{j,l}^j = \frac{\sum_{m=0}^{M-1} \hat{H}_{j,l}^j,(mG)}{M}. \quad (29)$$

In (28), $N'(mG)/c$ is a noise plus crosstalk component of the estimated channel. The estimated channel will have a large value of this component when the signal to noise ratio is small. 2×2 MIMO using SFC or STC has twice as many noise component than SISO when the signals from the two antenna paths are combined (STFC has twice more noise component compared to SFT or STC). The affection of this noise component is shown in simulation results.

III. SIMULATION RESULTS

We simulate the proposed system model with the QPSK constellation under power line channel conditions. The simulation assumes a PLC multipath fading channel with L=6, e.g., whose parameters in Zimmermann model are shown in Table I of [8]. For simplicity, we also assume the same fading channel parameters for the two antenna paths — in the case of PLC channels using indoor power lines; the changing of channel parameters between the antenna paths is almost negligible in practice [7]. We set $N = 1024$ (in the

case of STFC, N equals 2048 for the same data rate), the CP size = 120 (unit: samples), f_c = 30 MHz, Δf (frequency spacing) = 10 KHz, and BW = 10.24 MHz for the simulation; hence the maximum data rate is approximately 18.3 Mbps.

Fig. 2 compares the BER performance between conventional MRC method and the proposed MRC method on SISO-/MIMO-OFDM system. First, in the case of MIMO-OFDM system using 2×2 SFC or 2×2 STC, the proposed MRC (AFMRC) obtains the performance gain of about $2dB$ at 10^{-4} BER compared to conventional MRC (AMRC). Second, in the case of MIMO using 2×2 STFC, AFMRC provides the $2dB$ gain at BER=10^{-4} compared to AMRC. Fig. 2 also compares conventional method (with no MRC) and the suggested method (with FMRC) on SISO-OFDM system. For SISO with FMRC, we can get approximately $2dB$ gain at BER=10^{-4} compared to SISO with no MRC. Therefore, FMRC for SISO system and AFMRC for MIMO system are more effective according to these simulation results. It also verifies that AFMRC is similarly effective at MIMO systems using SFC, STC, or STFC.
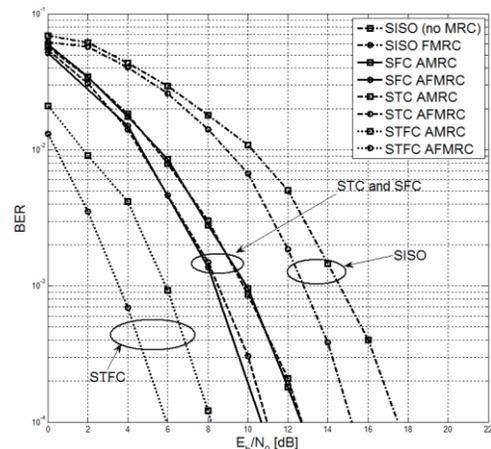

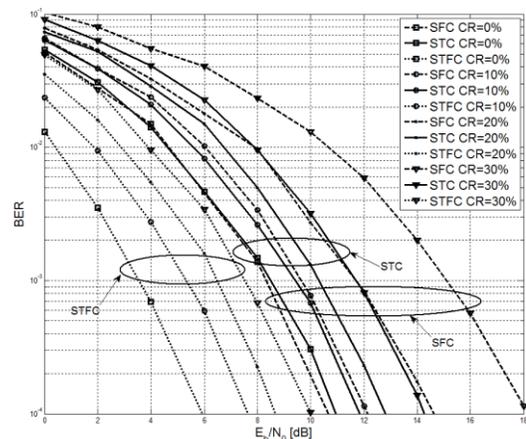Figure 2. Performance comparison of SISO/MIMO-OFDM with different MRC schemes ($A = 0.3$, $\tau = 0.1$).

Figure 3. BER comparison of SFC, STC, and STFC based MIMO-OFDM PLC when crosstalk exists ($A = 0.3$, $\tau = 0.1$).
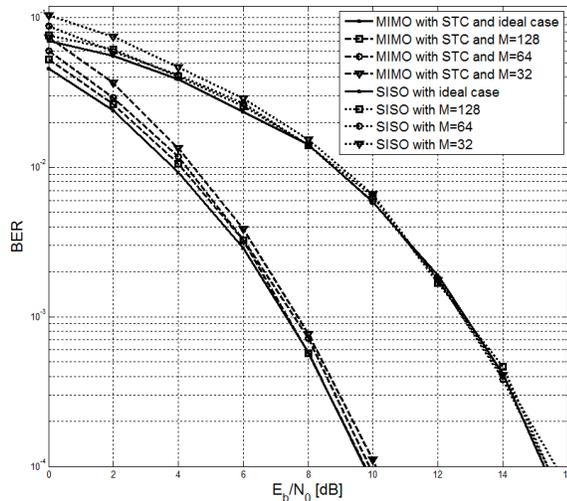
Figure 4. BER comparison of SISO/MIMO PLC with least-square channel estimator when M varies ($A = 0.3$, $\tau = 0.1$).

Fig. 3 compares the BER performance of 2×2 STC, 2×2 SFC, and 2×2 STFC based MIMO-OFDM PLC when the crosstalk between antenna channels exists. Assume that AFMRC is applied to all these MIMO coding schemes. We can observe that as the crosstalk ratio (CR) increases, STC has a relatively reduced BER compared to SFC. Especially, for CR > 10%, the relative performance gain difference between STC and SFC becomes greater. It indicates that STC is less sensitive to crosstalk than SFC. In Fig. 3, we can find out that STFC has a similar tendency as STC. It looks that as CR increases the STFC (STC as well) performance is degraded linearly while the SFC performance is degraded exponentially; specifically, for the 10% increase of CR, the coding gain of STFC degrades $1.5dB$ (similar in STC) in the average sense while that of SFC degrades $1.2dB$ for the increment of CR 0% to 10% and $2.2dB$ for CR=10% to 20%, and $3.5dB$ for CR=20% to 30%. Therefore, STC and STFC have robustness to crosstalk and STFC with additional complexity improves the BER performance of STC further.

In Fig. 4, we evaluate the BER performance of SISO-OFDM PLC and STC-based 2×2 MIMO-OFDM PLC using least square channel estimation methods with different number of pilot subcarriers. Fig. 4 shows that BER decreases as the number of pilot subcarriers increases. It confirms that if we have a larger number of pilot subcarriers, we can estimate the channel more exactly (since the noise component in (28) becomes smaller). Fig. 4 also shows that the BER curve is going to approach to the ideal case as E_b/N_0 [dB] increases. In the case of MIMO PLC, the ideal channel estimator has $0.5dB$ gain at the BER=10^{-2} compared to the least square channel estimator with M=32 pilot subcarriers. However, at the BER=10^{-4}, the gain difference between those two schemes

reduces (about 0.1 to 0.2dB); for the larger M, that difference gets smaller. In the case of SISO, we can observe the similar tendency of BER curves as in MIMO.

However, in Fig. 4, we can observe that SISO improves the BER performance a little bit compared to STC-based MIMO, since the latter requests two times more noise component than the former. SFC-based MIMO shows the similar performance results (even their results are not included here due to the space limit).

IV. CONCLUSIONS

We proposed a MIMO-OFDM based PLC using AFMRC. We compared SFC, STF, and STFC based MIMO-OFDM over statistical indoor power line channels, with or without crosstalk between antenna paths. By computer simulation, we verified that the proposed MRC scheme is effective to both SISO PLC and MIMO PLC applications. It is shown that STC and STFC are more robust to crosstalk than SFC; STFC with additional complexity overpowers STC and SFC in terms of BER over power line channels. Further, we also simulated the proposed MIMO model with imperfect channel estimation.

REFERENCES

[1] IEEE standard for broadband over power line networks: Medium access control and physical layer specifications, IEEE Standard 1901, 2010.

[2] M. Zimmermann and K. Dostert, "A multipath model for the powerline channel," IEEE Trans. Commun., vol. 50, no. 4, Apr. 2002, pp. 553-559.

[3] S. Galli, "A simplified model for the indoor power line channel," in Proc. IEEE ISPLC 2009, Apr. 2009, pp. 13-19.

[4] M. Tonello, "Wideband impulse modulation and receiver algorithms for multiuser power line communications," EURASIP Journal on Advances in Siganl Processing, vol. 2007, article ID 96747, 2007, pp. 1-14.

[5] J. Haring and A.J. H. Vinck, "Coding for impulsive noise channels," in Proc. IEEE ISPLC 2001, Feb. 2001, pp. 103-108.

[6] L. Hao and J. Guo, "A MIMO-OFDM scheme over coupled multi-conductor power-line communication channel," in Proc. IEEE ISPLC 2007, March 2007, pp. 198-203.

[7] R. Hashmat, P. Pagani, and T. Chonavel, "MIMO communications for inhome PLC networks: measurements and results up to 100 MHz," in Proc. IEEE ISPLC 2010, March 2010, pp. 120-124.

[8] B. Adebisi, S. Ali, and B. Honary, "Space-Frequency and Space-Time-Frequency M3FSK for indoor multiwire communications," IEEE Trans. on Power Delivery, vol. 24, no. 4, Oct. 2009, pp. 2361-2367.

[9] S. Wei and P. Zhiwen, "Iterative LS channel estimation for OFDM systems based on transform-domain processing," in Proc. IEEE WiCom 2007, Sept. 2007, pp. 416-419.

[10] N. Kawabata, H. Koga, O. Muta, and Y. Akaiwa, "A high-speed power-line communication system with band-limited OQAM based multi-carrier transmission," IEICE Trans. on Commun., vol. E92-B, no. 2, Feb. 2009, pp. 473-483.

[11] H. Jafarkhani, Space-Time Coding Theory and Practice, Cambridge Univ. Press, 2005, ch. 11.

Opportunities and Challenges of DSM
in Smart Grid Environment

Maria das Neves Queiroz de Macedo, Joaquim Jorge Martins Galo, Luiz Alberto Luz Almeida, Antonio Cezar Castro Lima

Department of industrial engineering postgraduate

Federal University of Bahia

Salvador, Brazil

mnevesmacedo@gmail.com, joaquimgalo2@gmail.com, lalberto@ufba.br, acdcl@ufba.br

Abstract—**The demand side management (DSM) is a set of techniques and policies that can help to provide greater efficiency in the planning and operation of the electrical network. The problem is that the full deployment of the DSM for the totality of consumers requires data availability of the loading curve and also requires devices that can operate in real-time in the system of each consumer, individually. In Brazil, the current power grid does not have these features. This article introduces the basic features of Brazil's electrical sector, covering the issues associated with the challenges of deployment of DSM in current scenario and discusses their deployment in Smart Grid environment, showing how new technologies can help to facilitate the implementation of DSM in Brazil's electrical network.**

Keywords -DSM; Smart Grid; electrical network.

I. INTRODUCTION

The increasing complexity of electrical systems from new markets is a worldwide phenomenon and is resulting from the significant growth of the electric network, which has changed the number of consumer units and also the use of renewable energy sources that can supplement the energy supply avoiding problems of energy shortages [1].

This fact implies the increase of initiatives and actions that contribute to the search for greater efficiency in energy management, involving a greater control in delivering quality, cost reduction and attention to environmental issues [2].

The idea is not only to increase the supply of energy, but also to control the shape of consumption through the use of energy management techniques by the demand side. In addition to conservation policies, the fight against waste involves the use of devices and more modern and efficient technologies [3].

The demand side management (DSM) is a set of techniques and policies that act to equalize the levels of demand of energy consumption throughout the day. The DSM allows the control of the consumer loads in order to operate the system more efficiently.

The DSM can be defined as a program or activities organized by the concessionaire that will affect the amount of time and energy used by consumer to postpone investment in new electrical installations [3].

The major challenge in the implementation of the program of DSM is the quest for knowledge of the daily behavior of the loads of the electrical system, which normally is not available on systems based on electromechanical meters.

In this scenario, the emergence of new technologies like the Smart Grid creates an environment that introduces a convergence between the infrastructure of generation, transmission, distribution, energy, information technology and digital communications infrastructure that enables the exchange of information and control actions among the various segments of the power grid.

This paper discusses issues associated to the DSM before and after deployment of the Smart Grid in Brazil and how new technologies can contribute with new applications for the management of the electrical sector. Section 2 of the article presents the basic characteristics of the electrical system and the challenges of implementing DSM programmes in Brazil in the current scenario. Section 3 discusses the implementation of DSM programmes in Smart Grid environment including opportunities and challenges and Section 4 presents the concluding remarks and future prospects.

II. BRAZIL'S ELECTRICAL SYSTEM AND THE DEPLOYMENT OF DSM

The installed capacity of electric power generation plants of Brazil reached 117,135 MW in 2011, according to Agency National Electric Energy(ANEEL) [4]. Table 1 indicates that the weight of the residential customer is quite significant, about 36%, commercial 24%, industrial 23%

and pubic power 12%, included the public lighting [1] for a total of more than 68 million consumers. The residential sector is what contributes the most to the tip of the electric system.

TABLE I.   ENERGY CONSUMPTION IN BRAZIL IN APRIL 2011

Consumer class	Electrical Energy consumption (MWh)	Number of consumer units
Residential	9.276.139	58.220.152
Industrial	6.446.334	557.120
Trade Services	6.054.848	4.949.915
Rural	1.387.517	3.963.534
Public authorities and other	3.179.562	673.546
Total	**26.344.400**	**68.364.267**

Source ANEEL.

Currently, in Brazil's electricity sector, the generation and transmission systems of energy companies already have systems of automation, supervision and control that use digital technology to monitor processes in virtually all major centres. These systems offer several features such as supervision, remote control and remote measurement through the SCADA (System of control and data acquisition) implemented in the Centres of Operation and indicate the operating conditions of all system automated in real time.

In the case of Distribution System in Brazil (voltage less than 34.5 kV) the reality is very different. Due to its complexity and a high number of consumers, about 68 million according to Table 1, the automation deployment of these systems is only at the beginning and its operation is still performed conventionally. Power measurements for billing are made manually in about 95% of consumer units, from the electromechanical meter reading, which makes for a poor monitoring of the workloads.

To deploy a DSM program efficiently, it is necessary to perform the following steps [5]:
- the study of this market present and growth prospects in the short and long term;
- study of various forms of energy supply, with the costs involved;
- study of characteristics of system loads;
- definition and implementation of an appropriate modeling system loads in the study;
- consumer awareness and encouraging their participation;
- analysis of the overall costs involved and of the evolution of the program,.

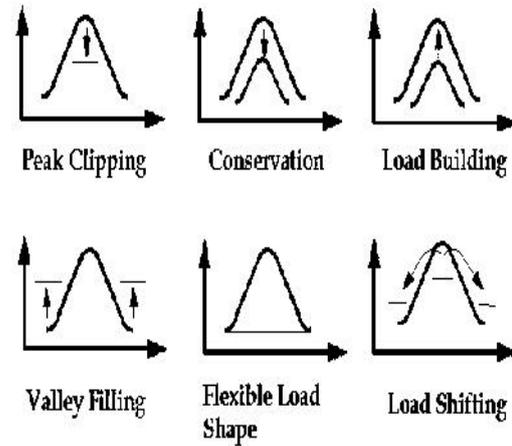The most commonly used DSM techniques are shown in Figure 1.

Figure 1.   Techniques DSM

A.   Peak Clipping

Peak Clipping is a program of load cutting, demand reduction in time for a heavy load. Cutting or reducing the duration of the peak can be reached by direct load control, by shutdown of consumer equipment or by distributed generation.

B.   Filling the valleys

It is a program which encourages the off peak consumption. It builds non-peak consumption periods which is particularly desirable as the cost of production is lower, which causes a decrease in the average price and improves the efficiency of the system. Incentives in several ways. such as discount into account should stimulate special consumers to change their habits.

C.   Conservation strategic

It is a program for seasonal energy consumption reduction mainly through efficient consumption and combating energy waste. This program is quite comprehensive and includes incentives for technological change.

D.   Load building

It is a program to control the seasonal energy consumption increase in building. The dealership employs intelligent systems, processes, more efficient equipment and more competitive energy sources to achieve energetic efficiency.

E. load shifting

It is a program, with the workload transfer period of greatest consumption (peak period to period of lower consumption), move tip out loads, without changing the total consumption. This is also possible with distributed generation.

F. flexible modeling

It is a set of actions and integrated planning between the concessionary and the consumer, subject to the needs of the moment. It is a partnership in order to model consumer loads, without affecting the actual conditions of security, limiting the power and energy that the individual consumer can use at certain times, through the installation of load limiting devices.

The implementation of these techniques requires a daily monitoring of load type of consumer units so that the most appropriate technique can be chosen for each case.

The collection of consumer load curve is the first step in the load characterization study. The reliability of the data obtained is a key point, because this will be the entire basis of the studies for the implementation of policies required to achieve efficiency in energy management [5].

Obtaining of data of daily load curve all consumers of an electric system, without the aid of new technologies is technically impossible or nearly impossible, because the conventional energy meters do not registers these daily curves for each consumer and in this case the curve is typically made from sampling of a small portion of the population and are later using statistical tools to characterize the entire population [6].

In Brazil, the data collection is made through a campaign of measures held every 4 years (or in special campaign), that consists in the installation of demand meters which will record the consumption and the time during a given period of a small sample of the population. This method is used to calculate tariffs.

This sample is chosen randomly and normally should represent the profile of a population that will be analyzed the variables of interest.

The biggest problem in this case is the inaccuracy and lack of representativeness caused by the production of daily load curve of a small portion of the population, every 4 years. Since there is a high probability that the consumption profile change significantly during this period and may not serve as a exact model for the choice of appropriate policies for each type of consumer, once hardly this process reflects the dynamics of consumer habits. In addition, some policies of DSM as the direct load control presents the difficulty of implementation in conventional systems [8].

For these reasons, in Brazil, the implementation of DSM policies has been carried out in isolated cases with few projects pilots deployed and which are still in operation.

III. DSM AND SMART GRID

The Smart Grid is based on the integrated use of information technology, automation, telecommunications and electric network control, that involves smart meters, sensors and digital network management devices, bi-directional, allowing deployment of control strategies and optimization of electric network with real-time data processing [7].

This convergence of technologies provides real-time data volumes with high reliability, enables the power grid to be controlled with more autonomy to the consumer units and enables energy management to be implemented on a decentralised basis, requiring the development of new methods of control, automation, and optimization of the operation of the electric system.

Figure 2 shows a basic scheme of Smart Grid that involves all sectors of the system: generation, transmission, distribution and consumers with or without generation.
The Smart Grid environment introduces features that allow [8,9]:

- The integration of a variety of energy sources of various dimensions and technology, allowing any consumer to produce energy and store or sell the surplus;
- Consumer empowerment from the more decentralised energy management and demand management by remote actuation in consumer devices, registration of consumer behaviour in the processes of planning and operation of the network;
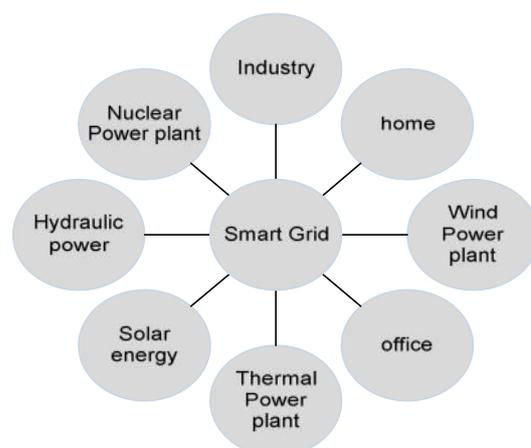

Figure 2.   Basic scheme Smart Grid

- Monitoring and reduction of technical and commercial losses of the electric power sector;

- Trend of sustainability: reduce the environmental impact of electricity-producing system, since it can diversify sources, reducing losses and using sources of low environmental impact;
- The integration of a variety of energy sources of various dimensions and technology, allowing any consumer to produce energy and store or sell the surplus;
- Monitoring and reduction of technical and commercial losses of the electric power sector;
- Trend of sustainability: reduce the environmental impact of electricity-producing system, since it can diversify sources, reducing losses and using sources of low environmental impact;
- Differentiated tariff and dynamics tariff ;
- Power quality monitoring: from data and indicators in real time;
- Creation of a database of load curves: record of consumer units in real time and off line to be used subsequently in energy management programs;
- Self-healing: ability to automatically detect, analyze, respond and restore network failures;
- Make and take advantage of competitive energy markets, favoring the retail market and distributed generation or microgeneration.

The implementation of a program of DSM in a system with a large number of consumer units, as in the case of Brazil, requires high reliability and ability to collect and process data from the network. In addition to devices, computational tools can contribute to deciding which policy is best suited for each consumer profile

The success of the management program is directly associated with the monitoring of loads 24 hour per day, so if you get the behavior of each type of load, observing the days not typical for each load and from this monitoring to provide precise actions for each case.

Despite the Smart Grid is a global trend, the deployment process is normally done slowly and in well defined stages, because it involves many challenges, such as [10]:

- Hardware's standardizing, software and protocol communication of the devices;

- High investment for replacement of conventional devices by microprocessor and change data capture infrastructure in each unit of the electric network;

- Study for regulation and tariff policy that may result from the new system;

- The large volume of data to be processed and its security against attacks

- Choice of functionalities of network devices namely, may enable the remote load or promote cut access to the Web via the mains, or measure the energy per unit equipment.

Brazil has set goals for Smart Grid deployment that requires high investments until 2020. Several proposals are already beginning to be discussed by ANEEL, energy companies and industry sectors and research centers.

The first Smart Grid project of largest-scale in Brazil is called Inovcity and is being deployed in the city of Aparecida and it consists in the installation of 15300 electronic meters in the city, that will transmit in real time all information of each consumer. The new equipment will eliminate the traditional manual reading made on each unit and will avoid measurement errors. Another benefit is to reduce fraud and improve the quality of energy delivered to consumers.

A basic scheme of a system with these characteristics must submit at least three levels of network with two-way communication, as is represented in Figure 3.
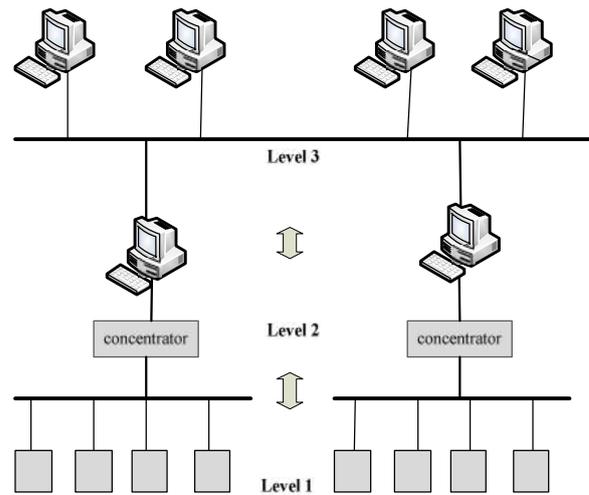
Figure 3.   Control system scheme

Level 1 is the level of consumer units (in case of large consumers or special) or set of consumer units connected by hubs. This network is formed by smart meters, with fixed address on the network, where should be installed the tools that enable each unit to apply all of the management features, like sending real-time data, measures, indicators, system information, load curve, tariff schedule price, consumption, etc.

Level 2 is the concentration of a group of consumer units in a particular region. The information for a full region is centered at this stage with the possibility of use of the

system control tools and access to region information in real time, in addition to data storage for later study.

Level 3 is the network that concentrates and totals information of the system as a whole and carries out communication and information exchange with the SCADA systems of transmission and Generation of the concessionaire.

It is important to note that this architecture, from smart devices have the ability to provide energy management bi-directionally, because currently many consumer units can also store energy or have alternative energy sources that will meet the utility's system at certain times. This can be quite handy, since it can supply the electrical system in times of need or in a contingency, with reliability and security.

Another important factor is the management policy of differentiated tariffs.. The effect of differentiated tariff is of great impact [11]. It is already used in large industrial consumers and it allows a better distribution of load throughout the day. This also is only possible with the installation of smart meters.

In this scenario, it becomes evident the need to have systems with the features of the Smart Grid, able to process a large volume of data in real time and choose the best configuration of the network, as well as to indicate the electrical system trends for the short, medium or long term using tools like the demand projection."

In Brazil, the issues associated with efficient power management and deployment of DSM are widely discussed, but still there are many challenges such as lack of adequate infrastructure and investment to the power sector to modernize.

IV. CONCLUSION

The concept of energy management has been expanded in recent years because of the new demands of the global market, which is alert to indicators of quality, efficiency and sustainability of the electric system.

The challenge is to devise economically viable technical alternatives capable of optimising energy consumption and enable the postponement of investments in the expansion of the energy supply.

Smart Grid is a world trend that presents exactly these features and functionality, as does the electrical system work as a large interconnected system, where there is strong interaction between company and consumers, bi-directionally which enables energy management with the effective participation of consumers.

As described in this article, in the case of Brazil, a long way should be done and there are many issues and technical

and economic obstacles to be overcome, although part of the electrical system (high and voltage mean) presents considerable advances in the automation of its processes.

It is important to note that in this initial stage of change of the electric sector there is the emergence of a new environment full of opportunities in developing applications for the Smart Grid, from new algorithms and artificial intelligence techniques for decision support.

REFERENCES

[1] F. Saffre and R. Gedge, "Demand-side management for the Smart Grid," in Proc. IEEE/IFIP Network Oper. Manage. Symp. Workshops (NOMS Wksps), Apr. 2010, pp. 300–303.

[2] M. Zhou, Y. Gao, and G. Li, "Study on improvement of available transfer capability by demand side management," in Proc. 3rd Int. Conf. Electric Utility Deregulation and Restructuring and Power Technologies, DRPT'08, 2008, pp. 545–550.

[3] A. Mohsenian-Rad, V. Wong, J. Jatskevich, R. Schober, and A. LeonGarcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future Smart Grid," IEEE Trans. Smart Grid, vol. 1, no. 3, pp. 320–331, Dec. 2010.

[4] Balance report National Energy, 2011, Ministry of Mines and Energy – MME, www.aneel.gov.br, November, 2012.

[5] C. W. Gellings and J. H. Chamberling, "Demand-Side Management: Concepts and Methods". Lilburn, GA: Fairmont Press, 1987.

[6] P. Palensky and D. Dietrich, "Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads", IEEE Trans. On Industrial Informatics, Vol. 7, No. 3, Aug. 2011, pp. 381-388.

[7] M. Amin and B. F. Wollenberg, "Toward a Smart Grid: Power delivery for the 21st century," IEEE Power Energy Mag., vol. 3, no. 5, pp. 34–41, 2005.

[8] G. T. Heydt, "The Next Generation of Power Distribution Systems", IEEE Transactions On Smart Grid, Vol. 1, No. 3, December 2010, pp. 225-235.

[9] J. Vasquez, J. Guerrero, J. Miret, M. Castilla, and L. de Vicuñanda, "Hierarchical control of intelligent microgrids," IEEE Ind. Electron. Mag., vol. 4, no. 4, pp. 23–29, Dec. 2010.

[10] E. Koch, P. Palensky, M. A. Piette, S. Kiliccote, and G. Ghatikar, "Architecture for supporting the automation of demand response," presented at the 1st IEEE Ind. Electron. Soc. Industry Forum, Santa Clara, CA, 2008. pp. 177-183.

[11] R. Podmore and M. R. Robinson, "The Role of Simulators for Smart Grid Development", IEEE Transactions On Smart Grid, Vol. 1, No. 2, September 2010, pp. 205-212.

[12] G.T. Heidt et al., "Professional Resources to Implement the Smart Grid," Proceedings of the 2009 North American Power Symposium (NAPS), Starkville, Mississippi, USA, October 4-6, 2009, pp. 123-129.

Real-Time Scheduling for Peak Load Reduction in a Large Set of HVAC Loads

Marco L. Della Vedova and Tullio Facchinetti
Dpt. of Industrial and Information Engineering
University of Pavia, Italy
E-mail: marco.dellavedova@unipv.it, tullio.facchinetti@unipv.it

Abstract—This paper presents a technique to predictably coordinate the activation of Heating, Ventilation and Air Conditioning systems (HVACs) in order to limit the overall peak load of power consumption (peak shaving). The proposed solution represents a viable approach to the Demand-Side Management in the context of a smart grid for this type of loads. The coordination method performs a load shifting based on the discipline of real-time scheduling traditionally studied in the field of computing systems. With this approach, individual constraints on the temperature associated with the activation of each HVAC can be satisfied. The main advantage of the proposed technique is its low computational complexity, which allows to manage large sets of loads. A specific approach is proposed and evaluated to deal with large sets of loads by properly partitioning the load set into subsets (scheduling groups) that are scheduled independently from each other. Simulation results based on realistic parameters show that the peak load can be reduced by 35% in normal working conditions, and up to 60% with respect to worst case situations, without affecting the comfort achieved by each HVAC.

Keywords- Real-time systems; Power system control; Scheduling; Demand-Side Management; Load shifting.

I. INTRODUCTION

The Demand-Side Management (DSM) is one of the key features enabled by the implementation of a smart power grid. The DSM is commonly implemented to improve energy system performance and reliability by limiting the peak load [1].

This paper proposes a technique to achieve predictable Demand-Side Management actions targeted to the reduction of the peak load of power consumption in a smart energy environment. Our approach is based on load shifting to avoid unnecessary simultaneous activations of a set of HVACs to reduce the peak load that is caused by the absence of coordination. A typical goal of a HVAC system is to keep the room temperature within the desired range. Therefore, heating or cooling is provided depending on the actual room temperature, which is affected by the temperature of the external environment. For a given external temperature the activation pattern of a HVAC can be suitably approximated by a periodic activity [2].

The modeling and control approach is derived from the domain of Real-Time scheduling studied in computing systems [3]. The timing behavior of load activations is modeled using parameters traditionally adopted by real-time processing systems, such as periods and activation times. The distinguishing features of our approach are: i) to provide guarantees in worst case conditions; ii) to achieve the user constraints on the temperature controlled by each HVAC; iii) to address the scalability issue in order to deal with large sets of loads.

This paper addresses the load management in worst case conditions. While it can be pessimistic in the average scenario,
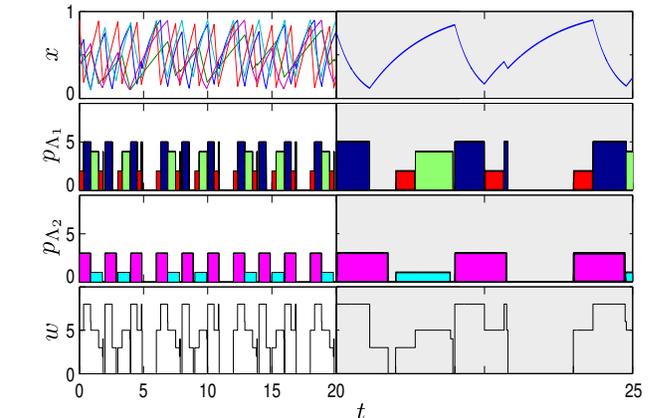

Fig. 1. Example of schedule of several HVACs. The time-scale is enlarged from $t = 20$ to $t = 25$ to better depict the system evolution. In the top chart shows the evolutions of temperatures normalized in their working ranges. The second and third top charts show the consumed power by two independently scheduled groups of HVACs. Active loads are indicated by their color. The bottom chart shows the overall power demand.

it allows a predictable load balancing to limit the impact of low-probability (but still possible) conditions which are behind potentially catastrophic consequences such as power disruption due to overloads [4]. The performance of real-time scheduling applied to electric loads in average conditions is studied in [5], where real-time load scheduling was compared with the case of absence of load control.

The control action is based on the specification of the requirements that need to be enforced on the temperature, i.e., the desired working range for each HVAC. The control policy will meet such requirements, while limiting the peak load. To model and meet the requirements on the desired working conditions, an approach based on the existing real-time scheduling theory has been used, namely Real-Time Physical Systems (RTPS) [6]. RTPSs are a special case of switching hybrid system where the switching signal corresponds to a real-time schedule. A physical variable (the controlled temperature, in our case) changes its value according to the activation the HVAC, and such activations are driven by a real-time scheduler. In practice, the dynamics of the physical process changes when its state is switched due to scheduling decisions. Moreover, the activity of a load may be interrupted and resumed later to account for higher priority loads. The priority of a load is automatically determined by the scheduling algorithm on the basis of the timing constraints.

The focus of the proposed approach is on the scalability to large systems. For this purpose, the discussed method has low

computational complexity, making it applicable to large sets of loads. The idea is to organize the loads into distinct groups that are scheduled independently from each others. This commonly adopted approach is known in the field multiprocessor real-time computing systems as *partitioned scheduling*. Despite it may produce a sub-optimal solution, the partitioned scheduling reduces the complexity to a linear or polynomial problem, and it allows the use of simpler scheduling algorithms having a number of known useful properties.

Fig. 1 shows an example of system behavior where different on/off HVACs are partitioned into two independent groups and scheduled to limit the overall peak load. The figure also shows the behavior of the temperature associated with each load, which depends on the activation state of the load.

The reminder of the paper is organized as follow. In Section II is presented a brief analysis of the state of the art in DSM literature. Section III describes the system model in both dynamics and timing terms. Section IV states the properties of the system and shows how is possible to get adequate timing parameters for the switching signal. Section V tackles the problem of the large set of loads: a partitioning method is proposed to achieve both good results in peak load reduction and small complexity in algorithms that allows the method to scale well with the number of loads. Section VI quantifies the performances of the proposed approach with simulated examples. Final remarks are in Section VII.

II. RELATED WORKS

The literature on power systems addressing DSM approaches is wide. A recent overview and categorization of DSM approaches is available in [1]. Some works focuses on the modeling aspects, without proposing a control method based on those models [7]. Optimization methods are often leveraged to minimize the peak load. However, the adopted solutions are based on off-line algorithms that can not cope with the dynamic and heterogeneous nature of large systems [8], [9]. Approaches based on artificial intelligence are also available, such as [10], where fuzzy logic is used to control a set of thermal loads. However, the properties of these methods (and predictability in particular) are not formally proved. In [11] the authors present a priority-based approach to the load management. It clearly shows different types of loads and their working constraints. The weak point of the approach is the manual assignment of priorities. Our approach, instead, while being inherently based on the assignment of priorities, provides an automatic assignment of priorities on the basis of timing parameters. In fact, the scheduler dynamically activates the load having the highest priority. This is a distinguishing feature of real-time scheduling methods.

RTPSs have been firstly presented in [6], dealing with affine dynamic systems. In [6] a partitioning-based method to manage large load sets is mentioned as a possibility and not explicitly integrated in the model. Before the formalization of RTPSs, the use of real-time scheduling for the management of electric loads was proposed in [12] and [13]. In [12] the focus is on the optimal partitioning of loads to manage large sets of loads. However, no physical variables are associated with

the loads. In [13] such association is addressed, but the issue of large systems is not considered. Moreover, in [13] state variables are characterized by integrator dynamics, while affine dynamics are considered in this paper. The application of real-time techniques to the load scheduling is investigated in [14] considering constraints on state variable variations and modeling errors. Errors are characterized by a statistical distribution, and they are compensated using a feedback technique based on the measurement of the state variable value in correspondence to request times.

III. SYSTEM MODEL

This section describes the system model, including the adopted model for HVACs and its modeling terms of real-time parameters.

A. Physical system

The power system considered in this paper is modeled as a set of n electric on/off HVACs. HVACs are independent from each others. A simple but accurate model for a HVAC system is proposed in [9], and it is recalled in the following. The adopted model describes a first order dynamic system, which has been proven to capture the behavior of HVAC loads accurately.

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = \frac{X_o(t) - X(t) - X_g s(t)}{\tau} \tag{1}$$

In (1), $X(t)$ is the internal air temperature of the room, $X_o(t)$ is the outside air temperature, X_g is the temperature gain of the air-conditioner, τ is the actual time constant of the room expressed in minutes, and $s(t) : \mathbb{R} \to \mathbb{B}$ is the current state of activation of the load: $s(t) = 0$ if the load is not active at time t and $s(t) = 1$ otherwise. The above model characterizes the behavior of a single HVAC. The controller must guarantee that the inside room temperature lays within a comfort range:

$$X(t) \in \left[X^{\min}, X^{\max} \right], \quad \forall t \tag{2}$$

In [15] this type of loads is called *controllable load*, since it can be shed to achieve the peak load reduction. Preferably, in this specific case a feedback control scheme should be integrated to regulate the temperature upon external temperature variations. A suitable method for this control is proposed in [16].

B. Real-time modeling of the switching signal

Load activations are driven by a switching signal. Such a signal is generated by a centralized controller, called *scheduler*. Considering the entire system, the *schedule* is defined as the function $s : \mathbb{R}^+ \to \mathbb{B}^n = [s_1 \ldots s_n]$. The distinguishing point of the proposed approach is that the switching signal is generated by a real-time scheduling algorithm, such as the Earliest Deadline First algorithm (EDF) [3].

The total power demand in any given time instant t is the sum of the power consumed by all active loads at time t. On the other hand, the scheduling action performed by

the scheduler produces the effect to reduce the unnecessary simultaneous activations of loads. As a result, the scheduling action is able to limit the peak load by avoiding unnecessary simultaneous activations.

Considering the above observations, the modeling and control problem translates to the assignment of proper values to timing parameters and constraints associated with the HVACs. For this purpose, a set of real-time parameters are associated to each HVAC. The adopted model derives from the periodic task model studied in real-time computing systems [17]. The generic i-th load is associated with the tuple $\lambda_i \doteq (T_i, C_i, P_i)$, where:

- $T_i \in \mathbb{R}^+$: it is the minimum time interval between two consecutive request times, or *period*; a *request time* $r_{i,k}$ is defined as the k-th request for activating the load; it holds $r_{i,k+1} - r_{i,k} = T_i$, $k \in \mathbb{N}$;
- $C_i \in \mathbb{R}^+ : C_i \leq T_i$ represents the required duration of the load activation time within each period T_i;
- $P_i \in \mathbb{R}^+$ is the nominal power of the i-th load.

Real-time parameters are used by the scheduling algorithm to generate the switching signal, i.e., the schedule. The values of timing parameters will be bounded to enable the schedulability analysis in the worst case. For this purpose, the following definition is introduced:

Definition 1 (Valid schedule). *A schedule s is said to be* valid *if it assigns to each load an amount of activity time equal to C_i within each time interval $[r_{i,k}, r_{i,k+1}]$. Formally, it holds:*

$$\forall i, \forall k \quad \int_{r_{i,k}}^{r_{i,k+1}} s_i(t)\,\mathrm{d}t = C_i \tag{3}$$

Note that the definition of valid schedule is slightly different from the one applicable to traditional real-time systems. In particular, in traditional real-time systems a less-then-equal relation is allowed, since C_i refers to the Worst-Case Execution Time (WCET). The WCET is the longest possible execution time of a real-time computing task. The WCET is used to perform the schedulability analysis in the worst case, while during the system behavior the actual duration of a task can be less than WCET. An equality is instead formulated in (3). This is required to achieve the requirements on the state variable variation.

To derive the results in following sections, two common figures used in real-time systems are introduced: the load utilization $U_i \doteq C_i/T_i$ and the total utilization $U^{\mathrm{tot}} \doteq \sum_{i=1}^{n} U_i$. While $U_i \leq 1$ is the fraction of time in which the i-th load is active, U^{tot} represents the total fraction of activity time of the whole load set. The total utilization U^{tot} is particularly useful. In fact, it is used to perform a test (the so-called *schedulability test*) to determine whether a given scheduling algorithm can successfully schedule the load set [3]. Under proper assumptions, a scheduling algorithm $\mathcal{A}$ is able to schedule a load set if $U^{\mathrm{tot}} \leq U^{\mathrm{lub}}(\mathcal{A})$, where $U^{\mathrm{lub}}(\mathcal{A})$ is the least upper bound on the total utilization that guarantees the schedulability when using the $\mathcal{A}$ scheduling algorithm.

C. The feasibility problem

As from previous definitions, the considered system model is composed by a dynamic system, the desired working range, the real-time parameters and a scheduling algorithm. While the dynamic system and the working range are related with the underlying physical process, i.e., the controlled HVAC, real-time parameters and the scheduling algorithm can be selected by the system designer. The selection should be made in order to obtain a feasible RTPS, according with the following definition:

Definition 2 (Feasibility). *Given the timing parameters describing the load set, a RTPS is said to be* feasible *if and only if user requirements are satisfied by every valid schedule.*

Equation (3) characterizes a class of switching signals within the set of all possible scheduling patterns. The RTPS *feasibility problem* concerns the identification of the class of valid switching signals such that the requirements on the controlled temperature are guaranteed. This problem translates to the identification of suitable values for C_i and T_i to drive the evolution of physical variables in compliance with user requirements.

The analysis is based on the observation that the scheduler generates a valid switching signal among all the possible valid signals. Therefore, the analysis is performed considering the worst case signal, i.e., the signal that brings to the worst possible situation in terms of user requirements violation. This allows to assess the behavior of all other "less critical" valid switching signals.

D. Peak load minimization

The application of RTPSs proposed in this paper is to limit the peak load of power demanded from a set of HVACs, while meeting requirements on the controlled temperature. The activity of an HVAC is controlled by the scheduler, which generates the schedule s_i for the i-th load. The i-th HVAC consumes either an amount of electric power $P_i \in \mathbb{R}^+$ when active, no power otherwise. Hence, the power consumption over time can be modeled with the function $p : \mathbb{R}^+ \to \mathbb{R}^+$ defined in (4).

$$p_i(t) = P_i s_i(t) \tag{4}$$

Transient phases between active and inactive states are not considered in this paper.

The overall instantaneous electric power absorbed at time t is the sum of the power consumed by all the HVACs, as stated in (5).

$$w(t) = \sum_{i=1}^{n} p_i(t). \tag{5}$$

The *peak load* is the maximum value taken by $w(t)$ during the considered timespan.

The peak load minimization can be optimally achieved by a RTPS scheduler when a uni-processor scheduling algorithm, such as EDF, is able to schedule the load set. In this case, the algorithm achieves that only one load is active at any given time, and the peak load is equal to the nominal power of the most power-consuming load. The schedulability test can be

used to determine whether there exists a feasible schedule, provided that user requirements are also met.

On the other hand, if simultaneous activations can not be avoided, i.e. when a uni-processor scheduling algorithm is not able to schedule the load set, then the minimization of the peak load becomes more complex. In this case, a RTPS scheduler generates a schedule that approximates the optimal solution. Therefore, the RTPS method represents an efficient heuristic to this problem. In fact, a scheduling algorithm as EDF has complexity $O(n \log n)$ required to sort a queue upon a load activation request.

IV. FROM PHYSICAL TO TIMING PARAMETERS

In [6] it is shown by a worst-case analysis how to derive the required timing parameters, namely the period T and the utilization U, from a first order model of a load like the one expressed by (1)-(2). The external temperature is represented by a constant function $X_o(t) = X_o$. Basically, there exists a set of pair (U, T) for which the temperature is always kept within the comfort range for any possible activation pattern generated by the real-time scheduler using a given value for the (U, T) parameters. In particular, it is shown that U must be chosen within the range $\left[U^{\min}, U^{\max} \right]$, where:

$$
\begin{aligned}
U^{\min} &= \max \left\{ 0, \frac{X_g + X^{\min} - X_o}{X_g} \right\} \\
U^{\max} &= \min \left\{ \frac{X_g + X^{\max} - X_o}{X_g}, 1 \right\}
\end{aligned}
\tag{6}
$$

It is easy to show that a valid choice is $U = \frac{U^{\min} + U^{\max}}{2}$. Once the utilization U has be determined, the period T must be chosen such that both inequalities (7) hold. Following this procedure, it is possible to assign timing parameters to the electric load such that the temperature is always kept within the comfort range. In (7), $\overline{U} = 1 - U$.

$$
X_o - X^{\min} > \frac{2X_o e^{UT/\tau} - X_g e^{2UT/\tau} + X_g e^{T(1+U)/\tau}}{1 - e^{T/\tau}}
$$

$$
X_g - X_o + X^{\max} > \frac{2X_o e^{\overline{U}T/\tau} - X_g e^{2\overline{U}T/\tau} + X_g e^{T(2-\overline{U})/\tau}}{1 - e^{T/\tau}}
\tag{7}
$$

The set of possible solutions for both (6) and (7) determines the region Ω of pairs in the $U - T$ space, called *feasibility region*. An example of feasibility region is depicted in Fig. 2. As it will be clearer from Section V, the choice of a low value for U helps to obtain a lower peak load. However, lower U brings to lower T, thus generating a higher switching frequency. This is often not tolerable in practical applications such as processes driven by electric motors. Therefore, the selection of U represents a trade-off between peak load and system lifetime.

V. LOAD PARTITIONING

This paper proposes the use of classic real-time scheduling algorithms to manage the set of electric loads, such as Rate Monotonic (RM) or Earliest Deadline First (EDF) [17]. The scheduling algorithm requires the specification of T_i and C_i for every load λ_i to build a schedule. Well known real-time
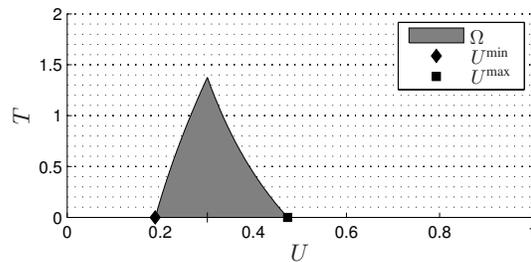

Fig. 2. Example of feasibility region in the $U - T$ space. By scheduling a load with timing parameters in the feasibility region, the achievement of user requirements is guaranteed (see Definition 2).

scheduling algorithms such as RM and EDF can generate a schedule where only one single load is active at any given time. However, this is possible only when the total utilization U^{tot} is less or equal of an upper bound $U^{\text{lub}}(\mathcal{A})$ whose value depends on the considered algorithm $\mathcal{A}$. For example, $U^{\text{lub}}(EDF) = 1$. Therefore, when $U^{\text{tot}} \leq 1$, preemptive EDF can build a schedule $\mathcal{S}$ without activating more than one load at any given time. As a consequence, the peak load $P^* = \max_i P_i$ is minimized.

On the other hand, if $U^{\text{tot}} > U^{\text{lub}}$ then the simultaneous activation of two or more loads can not be avoided, leading to a possibly larger peak power consumption $P > P^*$. The proposed solution is to partition the whole load set into m disjoint sets Λ_j, $j = 1, \ldots, m$, called *scheduling groups*. Scheduling groups are determined such that their total utilization, defined as

$$
U_{\Lambda_j} = \sum_{\lambda_i \in \Lambda_j} U_i,
\tag{8}
$$

is smaller than or equal to $U^{\text{lub}}(\mathcal{A})$. This property enables an uni-processor scheduling algorithm $\mathcal{A}$ to find a valid schedule independently for each scheduling group.

Since there is no relationship between the schedule generated within any pair of scheduling group, the maximum overall peak load will happen when the loads with the highest power are simultaneously activated in every scheduling group. Therefore, an upper bound P^{ub} on the peak load can be found considering the simultaneous activation within every group of the load with the highest power P_i, i.e.:

$$
P^{\text{ub}} = \sum_{\Lambda_j} \max_{\lambda_i \in \Lambda_j} P_i.
\tag{9}
$$

A. Level packing

The problem of partitioning the set of loads can be formalized as a *level packing* problem [18]. Level packing is a special case of the generic two-dimensional packing problem. In level packing, one or more strips are filled to accommodate a set of rectangles such that the total height is minimized. The peculiarity of level packing is that rectangles are partitioned in horizontal strips or levels. The complexity of the level packing problem is NP-hard; in fact, it can be easily reduced to a classical one-dimensional packing problem having NP-hard complexity. Approximation methods have been proposed to face the complexity issue [19]. The approximation methods
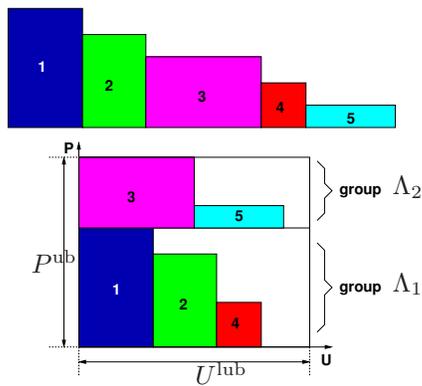
Fig. 3. Example of level packing using the FFDH algorithm. Five items are firstly ordered by non-increasing height and then packed into two levels. Note that item 3 generates a new level Λ_2 since it does not fit at the right of item 2 within level Λ_1. The load utilizations of the 5 loads are respectively $(0.28, 0.26, 0.49, 0.22, 0.30)$. Their consumed power are $(5, 4, 3, 2, 1)$. The schedule generated by this system is depicted in Fig. 1.

are built by ordering the rectangles by non-increasing heights. Rectangles are grouped to fill the strips using different strategies. In each level, items are packed from left to right by non-increasing height, similarly to the arrangement of books within a bookshelf (see Fig. 3). The *First-Fit Decreasing Height* scheme (FFDH) is conceived such that it inserts the next item X (in non-increasing height ordering) on the first level where X fits. If no level can accommodate X, a new level is created. After the packing action, the height of a level is equal to the height of the leftmost item. The interesting aspect of FFDH is that the time complexity is $O(n \log n)$. Moreover, its approximation ratio has been formally derived. In particular, it holds $\text{FFDH}(I) \leq (17/10) \cdot \text{OPT}(I) + 1$, where I is a set of items to be packed, $\text{FFDH}(I)$ is the height obtained by FFDH, and $\text{OPT}(I)$ is the height produced by the optimal algorithm. The asymptotic bound of 1.7 is proved to be tight.

B. Application to the scheduling problem

The application of the level packing to the scheduling of electric loads requires the proper modeling of the loads. Therefore, each load λ_i is represented as a rectangle having height equal to the power consumption P_i and width equal to its utilization U_i, being $U_i \leq 1$. The packing happens in a two-dimensional space where the utilization appears on the x axis, while the consumed power is on the y axis. The width of the packing space corresponds to the least upper bound on the utilization of the considered scheduling algorithm (e.g., $U^{\text{lub}} = 1$ for EDF). The goal to limit the total height of packed rectangles clearly corresponds to the goal of limiting the peak load of power consumption of the whole power system. On the other hand, fitting the items on the x axis in each level corresponds to group a set of loads whose total utilization is less than or equal to U^{lub}, thus composing a set of loads that is successfully schedulable by the considered real-time scheduling algorithm. Once all loads have been grouped into scheduling groups using the level packing, each group of loads is scheduled independently from other groups. The schedulability is guaranteed since the utilization U_{Λ_j} of the
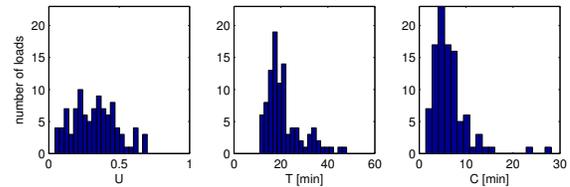

Fig. 4. Distribution of timing parameters values obtained from physical parameters in the 100-loads Monte Carlo simulation.

group Λ_j is less than or equal to the upper bound U^{lub} that achieves the schedulability of the load set. Fig. 1 shows an example of schedule of loads whose relevant parameters are listed in Fig. 3. The grouping of loads is as in Fig. 3.

The proposed technique recalls the *Rate Monotonic First-Fit Decreasing Utilization* (RM-FFDU) partitioning scheme for scheduling fixed priority real-time tasks on a multi-processor system [20], where bin-packing techniques are used to allocate tasks to processors. However, [20] does not address the optimization of the total power consumption. Moreover, the key distinction is that in this paper the ordering is made with respect to the value of load's consumed power, and utilization is not considered for this purpose.

VI. PERFORMANCE EVALUATION

This section assesses the performance of the proposed approach by means of simulation and using realistic parameters. Similarly to [9], from the viewpoint of generating different realistic operating scenarios, Monte Carlo simulations are applied in this paper. By assuming uncertainties on different variables (τ, X, X_o, X_g) that closely resemble real-life operating conditions, Monte Carlo simulations are performed through repeated sampling of uncertain variables.

Each simulation run has been initialized with pre-specified stochastic parameters, chosen with the following method: number of air conditioner – 100; internal temperature distribution – Normal distribution with mean 72 F and standard deviation 12 F, i.e. $X(0) \in N(74, 12)$ [F]; air-conditioner model parameters – $\tau \in N(64, 5)$ [min], $X_g \in N(30, 10)$ [F]; outside air temperature – uniform distribution between 75 F and 90 F $X_o \in [75, 90]$ [F]; desired temperature range – $X^{\text{min}} = 70$ [F], $X^{\text{max}} = 76$ [F]. In terms of power request, 5 air-conditioner sizes are considered – $P \in \{1.2, 2.5, 3.0, 4.5, 6.0\}$ [kW].

Given the load set, the timing parameters (U, T, C) are calculated for each load as illustrated in previous sections of this paper. In particular, it is set $U = \frac{U^{\text{max}} + U^{\text{min}}}{2}$. Then, T is set as the maximum value within the set of feasible values (Definition 2). Fig. 4 shows the resulting histograms of timing parameters. Utilizations range from 0, i.e. always off, to 1, i.e. always on. Periods are in the range $[10, 60]$ minutes.

In order to evaluate the performance of the proposed approach, the RTPS control method is compared with the traditional hysteresis control. In the hysteresis-control approach, each air-conditioner is turned on when the room inside air temperature reaches the upper thermostat set-point X^{max} and turned off whenever this temperature falls below the lower thermostat set-point X^{min}.
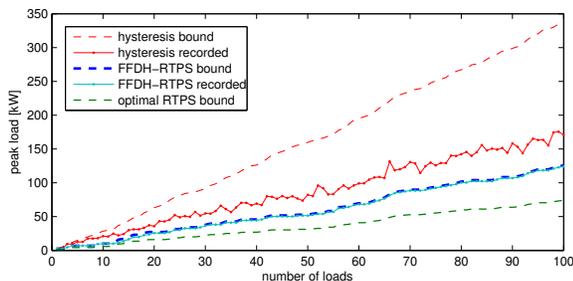
Fig. 5. Comparison between hysteresis- and RTPS-control methods. The RTPS-controlled actual behavior, during the simulation of 120 hours, reduces the peak load of the 35% in respect of hysteresis control in regular working conditions. The improvement on the theoretical bound associated with worst case working conditions is around 60%.

Fig. 5 shows the peak load as a function of the total number of loads in the system under different control strategies. With the hysteresis-based control, the worst case bound is equal to the sum of all loads power, while in the RTPS-controlled system the theoretical bound is obtained by summing the power consumed by the most power-consuming loads in each scheduling group generated by the packing algorithm. It is worth to note that, while the theoretical bound for the hysteresis-controlled method may represent a very unlikely worst-case condition, it is still a possible situation, whose likelihood increases with the system lifetime. For both control methods, the figure also shows the actual peak load recorded by simulating the system behavior over a 120 hours (5 days) time span. The actual (recorded) peak load generated by the RTPS-based control is able to reduce the peak load in average by 35% with respect to the actual (recorded) peak load in absence of coordination, which represents a normal working condition. On the other hand, the peak load is reduced by up to 60% with respect to the worst possible case (theoretical bound) of the hysteresis-controlled case. Finally, it is worth to note that the recorded peak load of the RTPS-controlled system is very close to its theoretical bound This means that the worst situation in the partitioned scheme (i.e., when the most power-consuming loads in every scheduling group are simultaneously activated) do happen almost always.

VII. CONCLUSION

This paper presented an approach to coordinate the activation of a large sets of HVACs in a Demand-Side Management scenario. The proposed method is based on scheduling techniques adapted from the domain of real-time scheduling. The combination of a level packing strategy and uni-processor scheduling algorithms allows to meet both timing and physical constraints. Simulation results based on realistic parameters prove a relevant improvement of peak load reduction.

REFERENCES

[1] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," IEEE Transactions on Industrial Informatics, vol. 7, no. 3, pp. 381–388, aug. 2011.

[2] G. Koutitas and L. Tassiulas, "A delay based optimization scheme for peak load reduction in the smart grid," in Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet. ACM, 2012, p. 7.

[3] G. Buttazzo, Hard real-time computing systems: predictable scheduling algorithms and applications, ser. Real-Time Systems Series. Springer, 2005.

[4] D. Novosel, M. M. Begovic, and V. Madani, "Shedding light on blackouts," Power and Energy Magazine, IEEE, vol. 2, no. 1, pp. 32 – 43, jan-feb 2004.

[5] M. L. Della Vedova, E. Di Palma, and T. Facchinetti, "Electric loads as real-time tasks: an application of real-time physical systems," in 7th International Wireless Communications and Mobile Computing Conference (IWCMC), July 2011, pp. 1117–1123.

[6] M. Della Vedova, M. Ruggeri, and T. Facchinetti, "On real-time physical systems," in Proceedings of the 18th International Conference on Real-Time and Network Systems (RTNS), November 2010, pp. 41–49.

[7] A. Molina-Garcia, M. Kessler, J. Fuentes, and E. Gomez-Lazaro, "Probabilistic characterization of thermostatically controlled loads to model the impact of demand response programs," Power Systems, IEEE Transactions on, vol. 26, no. 1, pp. 241–251, Feb. 2011.

[8] T.-F. Lee, M.-Y. Cho, Y.-C. Hsiao, P.-J. Chao, and F.-M. Fang, "Optimization and implementation of a load control scheduler using relaxed dynamic programming for large air conditioner loads," Power Systems, IEEE Transactions on, vol. 23, no. 2, pp. 691 –702, may 2008.

[9] B. Ramanathan and V. Vittal, "A framework for evaluation of advanced direct load control with minimum disruption," IEEE Transactions on Power Systems, vol. 23, no. 4, pp. 1681–1688, November 2008.

[10] C.-M. Chu, T.-L. Jong, and Y.-W. Huang, "A direct load control of air-conditioning loads with thermal comfort control," in IEEE Power Engineering Society General Meeting, June 2005, pp. 664–669 Vol. 1.

[11] M. Pipattanasomporn, M. Kuzlu, and S. Rahman, "An algorithm for intelligent home energy management and demand response analysis," Smart Grid, IEEE Transactions on, vol. PP, no. 99, p. 1, 2012.

[12] T. Facchinetti, E. Bini, and M. Bertogna, "Reducing the peak power through real-time scheduling techniques in cyber-physical energy systems," in Proceedings of the First International Workshop on Energy Aware Design and Analysis of Cyber Physical Systems (WEA-CPS), April 2010, pp. 18–25.

[13] T. Facchinetti and M. L. Della Vedova, "Real-time modeling and control of a cyber-physical energy system," in Proceedings of the First International Workshop on Energy Aware Design and Analysis of Cyber Physical Systems (WEA-CPS), April 2010, pp. 10–17.

[14] ——, "Real-time modeling for direct load control in cyber-physical power systems," IEEE Transactions on Industrial Informatics, special issue on Information Technology in Automation, vol. 7, no. 4, pp. 689–698, November 2011.

[15] S. Shao, M. Pipattanasomporn, and S. Rahman, "Development of physical-based demand response-enabled residential load models," Power Systems, IEEE Transactions on, vol. PP, no. 99, p. 1, 2012.

[16] M. L. Della Vedova and T. Facchinetti, "Feedback scheduling of real-time physical systems with integrator dynamics," in Proceedings of 17th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), September 2012.

[17] C. L. Liu and J. W. Layland, "Scheduling algorithms for multiprogramming in a hard real-time environment," Journal of the Association for Computing Machinery, vol. 20, no. 1, pp. 46–61, January 1973.

[18] A. Lodi, S. Martello, and D. Vigo, "Models and bounds for two-dimensional level packing problems," Journal of Combinatorial Optimization, vol. 8, no. 3, pp. 363–379, 2004.

[19] A. Lodi, S. Martello, and M. Monaci, "Two-dimensional packing problems: A survey," European Journal of Operational Research, vol. 141, no. 2, pp. 241 – 252, 2002.

[20] Y. Oh and S. Son, "Allocating fixed-priority periodic tasks on multiprocessor systems," Real-Time Systems, vol. 9, no. 3, pp. 207–239, 1995.

Smart Meters Security Assessment in the Brazilian Scenario

Rafael Cividanes, Nelson Uto, Bruno Botelho,
Sergio Ribeiro, Christiane Cuculo
Information Security Management
CPqD Foundation
Campinas, São Paulo, Brazil
{rafaelsc, uto, bpereira, sribeiro, ccuculo}@cpqd.com.br

Danilo Suiama, Heron Fontana
Management of Metering, Losses and Technology
ELEKTRO
Campinas, São Paulo, Brazil
{danilo.suiama, heron.fontana}@elektro.com.br

Abstract – Smart meters play an important role in smart grid architectures and technologies. Preliminary research has already indicated vulnerabilities as well as attacks likely to happen on the new meters. The Brazilian scenario combines two main perspectives: one is the growing rate of fraud in the electric energy measurement and the other is a resolution that establishes the replacement of all electromechanical meters with the new smart meters. This paper presents a threat modeling in smart meters and some possible attacks related to the mentioned scenario. It also discusses the ongoing Brazilian R&D project that focuses on security assessment of these new devices and the creation of a laboratory for smart meter vulnerability research and pentest execution.

Keywords-smart meter cyber-security; vulnerabilities; attacks; security laboratory for smart meter pentest

I. INTRODUCTION

The Smart Grid technology brings several benefits to the utilities, the customer and the environment, such as: increased resilience after service disruption, flexible distribution, energy efficiency (less waste), enhanced usage monitoring, variable pricing structure, reliability and operational efficiencies, renewable sources of energy, etc. However, these benefits come together with some security concerns, including many types of vulnerabilities related to each asset, system or communication channel in the Smart Grid architecture. In fact, each Smart Grid component can be the target of an attack: communications, meters, transmission substations, distribution substations, and corporate network. Each element should have integrity mechanisms working properly [1]. The Smart Grid security issue is being analyzed and some publications have been issued [2] [3].

The implementation of Smart Grids in Brazil has an important goal: the reduction of non-technical losses. In the current scenario, the energy theft in Brazil represents a loss of US$ 1.4 billion, not including taxes. For example, the city of Rio de Janeiro has about 20% of energy loss. The new electronic meters will help detect and reduce this huge loss. However, if not properly designed and tested, the use of these new meters may have a huge impact on the entire Advanced Metering Infrastructure (AMI).

Almost all residential and business customers (low voltage) in Brazil are using electromechanical energy meters and the fraud techniques related to these current meters are well known and even obsolete. Violations are restricted to physical attacks. On the other hand, the new electronic smart meters open the door for vulnerabilities to be exploited. Firmware analysis, data extraction, access to cryptographic keys, bus sniffing, denial of service, etc., are examples of actions that could be performed to compromise a smart meter device. In this case, not only physical but also logical attacks are feasible.

In August 2012, ANEEL (Brazilian Electricity Regulatory Agency) approved a resolution which states that energy distributors will have to install electronic meters for all consumers who choose to be billed in differentiated tariffs. It is the first step by the Brazilian Government to replace the 67 million electromechanical meters by January 2014. This is a unique moment in the history of the Brazilian energy sector when all meters will be replaced. However, the choice of smart meters with security weaknesses or severe vulnerabilities may cause serious damage to Brazilian economy.

It is in this context that the research and development Project entitled *Security Assessment for Smart Meters* was designed. The objective is to investigate different brands and types of smart meters available in the market, then run tests on them for checking security requirements, assess potential impacts, and finally build the Smart Meter Cyber Security Laboratory specialized in security evaluation of smart meters. The results will help the entire energy sector in Brazil and the development of additional security regulations and standards to be included to metrologically relevant requirements currently in use by Inmetro (Brazilian agency responsible for running tests on the meters).

The remaining part of this paper is structured as follows: Section II describes a threat modeling applied to smart meter devices, presenting what could be done if these assets were designed without considering a security baseline. Section III presents some possible attacks, how they are executed and the concepts related to the existing vulnerabilities. Section IV presents a synthesis of the Brazilian project, addressing the objectives, methodology of work, mechanisms of security tests and the establishment of a Laboratory. Finally, the conclusions are exposed in Section V.

II. THREAT ASSESSMENT IN SMART METERS

In the field of Smart Grid security, one of the most important components are the smart meter devices. They are the front door of several types of attacks on security. Besides being the common hub in the Smart Grid architecture, they are in the wild (outside the physical boundaries of the utility

company's property), which increases the probability of vulnerability exploitation. The physical access to these devices, which can be easily obtained by consumers and adversaries, facilitates the execution of an attack. In order to study and understand the smart meters' characteristics and mechanisms of protection (if any), any individual can, for example, buy a device on eBay.

Smart meter devices are made of electronic components and encompass different types of technologies, protocols, and embedded systems. The risks related to these new devices come from the fact that they are, almost never, built with security requirements in mind. In this way, it is expected that they would fail if exposed to unexpected situations, creating opportunities for security violations. Thus, to mitigate these risks, it is essential to determine the attacker's perspective, which generally includes the opportunity for financial gain, the opportunity for mischief, and the opportunity for chaos [4]. These perspectives can result in different kinds of threats, depending on each country's scenario.

In the Brazilian scenario, it is possible to identify four main threats: (1) energy usage frauds; (2) user privacy violation; (3) propagation of malicious code to others meters through the AMI; and (4) malicious interruption of electricity supply. Each threat outlined above can result from different security vulnerabilities being exploited, culminating in success of the attack. However, in all the cases, the threats exist because of the smart meters' technological characteristics, i.e., the existence of a firmware (maybe the most common target of the attacks), the physical interfaces that provide access to the device (serial, parallel, and infrared/optical ports), the electronic components that store data (EEPROM, Flash, RAM), the buses that pass data between components (parallel and serial buses), the wireless communication protocols, and the two way communication.

In the case of the first threat, the energy usage fraud, besides the existence of many frauds related to electromechanical meters currently in use in Brazil, there are also record of frauds related to others new electronic devices, for instance pay TV [5]. From this context, it is possible to infer that the new smart meter devices to be used in Brazil will further increase the current level of fraud. Hacker attacks against the new meters are likely to occur. This kind of threat also includes attacks executed only for fun, to show how easy it is to change the energy consumption data in the meters.

User privacy violation, the second threat, is another important issue when analyzing smart meters security [6]. The metering data may reveal some customer behaviors, determining when they are at home, at work, or traveling. When at home, even specific activities may be deduced. It was recently approved in Brazil a law addressing Internet Privacy and it is also being studied the Privacy issue in the Smart Grid environment as a part of another R&D project, funded by ANEEL. Thus, the metering data stored in these smart meters are also the target of attacks, being necessary, therefore, to fully test the encryption mechanism used to protect them.

Considered perhaps the most dangerous threat, the propagation of malware over the AMI, could cause irreparable loss to the entire energy sector. In this case, an analogy can be made with the spread of a computer virus over the Internet. The recent outbreak of Stuxnet worm is a real example in the nuclear facilities [7]. Indeed, the overarching goal of an attacker is to try to identify vulnerabilities that allow expanding the control of a single device to other devices with limited or no physical access. The aggravating factor related to the Brazilian context is that, in a short period of time, all old meters will be replaced, increasing the risk of adopting vulnerable smart meters if the process occurs without care.

At last, but not least, the existence in the smart meters of a functionality for remotely interrupting the energy supply can be maliciously used as part of a physical attack. Consider for instance that a thief wants to invade a store, protected by an alarm system connected to an uninterruptible power supply. In order to commit the robbery, the burglar could simply cause a power cut and wait until the batteries get drained and the protection system is disabled. A worst scenario to be considered consists of simultaneously and persistently replicating the aforementioned attack to every single building of a city or a major area, which would quickly lead to total chaos.

Although some research activities are being conducted, there are still several threats related to vulnerability exploitation in smart meters. This section addressed the four main threat cases in the Brazilian scenario. Various types of possible attacks can result in the materialization of these threats. For this reason it will be necessary to map and study them, so that tests can be performed on these devices just as they are tested in the metrological aspects (already performed in Brazil). The following section discusses some possible attacks against smart meters, related to this R&D project.

III. POSSIBLE ATTACKS

Before presenting the possible attacks against smart meters, it is important to understand their architecture and main components as illustrated in Figure 1. Generally, there are two processors: one for calculating, with the aid of the current and voltage sensors, the energy consumption, which is showed on the LCD, and the other for processing and transmitting the collected data, besides implementing additional functionalities such as access control and firmware management, for example. The smart meter firmware and configuration parameters are stored in the EEPROM and/or flash memory modules and can be remotely updated if necessary. Several modules are responsible by providing the communication layer, using technologies such as PLC, GSM/GPRS, ZigBee, and WiFi. Examples of buses used to interconnect these components include SPI, I2C, and SMbus. Maintenance can occur using the aforementioned network protocols or through an optical port compliant with the ANSI C12.18 specification.
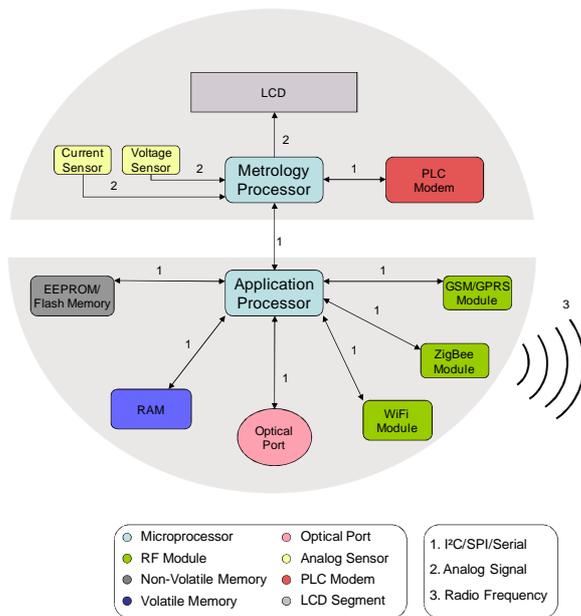
Figure 1. Smart Meter Anatomy

Recently, discussion has been growing about security attacks on smart meters. As a matter of fact, some research was done focusing on the "smart meter pentest" subject. The initial stage of this present R&D work covered the research of known attacks that are related directly to the mapped threats in the Brazilian scenario. Four works are distinguished: in [8], some general weaknesses are listed, including some hacking methods used on the devices; in [9], some interesting attacks against smart meters are explained; in [10], a methodology of attack on AMI is addressed; and in [11], a pentest model for smart grid is proposed with focus on smart meters.

Attacks on smart meters generally start by the physical interaction with the device usually performed through an optical port. In general, the first obstacle in accessing the meter is to bypass the authentication mechanism. However, some meters still do not have password or use the default manufacturer password. When using robust passwords, two possible attacks are the attempt to extract data from hardware components and the use of brute-force methods. The first one depends on the comprehension of the smart meter anatomy, as shown in Figure 1.

When accessing the device, after the successful authentication, smart meters can be connected to laptops through optical probes. This method is generally used by the company's field technicians who perform maintenance and conduct configuration changes in meters' setup. Attackers can use the same method to start their attacks. A slight change in the meters' setup data is considered an attack, as this can increase or decrease the energy consumption data. Furthermore, these changes may allow non-authorized communication with other meters, aggregators or non-AMI networks.

Another type of attack includes tampering hardware components in order to damage the device. Some smart meters contain tamper-protection mechanisms, although this type of control may fail after massive exploration. This kind of approach is initiated by gathering the available documentation such as components datasheets, operation guides and schematic diagrams. After the initial evaluation, the next important stage is to identify the weaknesses in the electronic components, for instance, the traffic of confidential data (e.g.: passwords or encryption keys) that runs between the components in a non-secure way. In this case, logical analyzers and oscilloscopes are used to help establish the type of traffic and evaluate the reading possibility. Another more advanced attack technique, related to this issue, is the use of hardware reverse-engineering techniques that makes possible the introduction of tampered components into smart meters. This type of attack is classified as hardware hacking.

Another category of attack is the attempt to access the data stored in the meter's components such as RAM, Flash memory and EEPROM chips. For the EEPROM, access is performed while the device is deactivated using the following tools: total phase beagle sniffer, bus pirate, syringe probes and JTAG programmers. Based on the dumped data, search for relevant information is performed focusing mainly on the following: (1) encryption keys and (2) firmware. Other executable codes, configuration data and files, in addition to the meter's authentication passwords or IDs, if obtained, can be useful to the attacker to run other attacks.

The attempt to find encryption keys can be performed through a simple search for basic strings (obvious keys) and entropy analysis techniques. The access to the smart meter's cryptographic keys can result, for instance, in the impersonation of the victim's smart meter, by the attacker, within the Neighborhood Area Network (NAN). Another aspect is the access to other meters that share the same symmetric key that the one which has been discovered. In any case, the impact can be enormous.

Software design flaws are the most common source of security weaknesses. Memory corruption vulnerabilities like stack overflows, heap overflows, format string, use-after-free, and off-by-one overwrite, can be exploited aiming information disclosure and arbitrary code execution. Code injection vulnerabilities like SQL and XML injection can allow improper access to user private information, device and network configuration parameters, enterprise private data, etc.

Attackers usually target firmware image recovery, as from their perspective it may drive many other attacks, like those related to software flaws. When obtaining the firmware source code, search for vulnerabilities can be performed through static analysis. When this is not possible, the alternative is to perform reverse-engineer of the firmware binary. This technique can be used to identify hard-coded strings like encryption keys and device passwords. Furthermore, firmware reverse engineering can be used with fuzzing tests [12] to identify software implementation flaws as well. The impact of attacks that explore firmware flaws goes from denial of service to total device compromise. It is important to emphasize that the binary code analysis, if demanded, is time-intensive and requires professionals with highly specialized skills.

Another attacking approach addressing the firmware consists of replacing it by a malicious version, using the

remote update functionality present in most smart meters. Vulnerabilities in the way they verify the authenticity of the code to be installed could be used by such an attack. In the worst case, one can consider devices that do not authenticate the new firmware at all. A much more advanced possibility, however, involves breaking the code authentication mechanism employed by the meter. A real example of the latter, in the context of operating systems, is the technique used by the flame malware [13] to camouflage as a valid Windows software update. In this case, it was necessary to improve Stevens's cryptanalytic attack against the MD5 hash function [14], in order to make the attack possible.

IV. R&D BRAZILIAN PROJECT: METERS SECURITY

This Brazilian R&D project is a 24-month program that includes the following phases: Phase (1) the state-of-the-art analysis of smart meters security: includes the research process on the already identified main vulnerabilities, attacks related to device intrusions both on the software and the hardware; Phase (2) elaboration of a methodology able to evaluate meter security and to build the Smart Meter Cyber Security Laboratory: covers the development of a methodology for testing, which will be used in the replication of future smart meter evaluations, and also includes the deployment of two laboratories, one for vulnerability research and discovery, located at CPqD, and another twin laboratory for running security tests, located at Elektro; Phase (3) execution of the hardware and software security tests in smart meters: considered the core of this project, addresses the smart meter pentests, the prospecting of new attack methods and the specification of minimum requirements for security taking into consideration the Brazilian scenario; and Phase (4) transfer of knowledge: deals with the transfer of knowledge to Elektro's team and the presentation of final results to the R&D committee and to the regulatory agency.

After the first phase of this work, it was prepared a survey addressing the main vulnerabilities and most relevant attacks on smart meter devices that were discussed during the latest security conferences worldwide. Considering the threats in the Brazilian scenario, it was possible to preliminarily ratify the elevated risk facing the replacement of all meters in Brazil. Considering the meters strength, in terms of security, which changes depending on the device, one can conclude that their conception is not based on requirements regarding information security.

Some laboratories in Brazil are already accredited by Inmetro to perform tests on meters. However, the information security tests are in initial stage in the country. The Smart Meter Cyber Security Laboratory will be the first of its kind in Brazil and will focus mainly on security tests. This laboratory can be used as the basis for the deployment of other similar laboratories elsewhere. Based on the results obtained from this R&D project, a set of minimum requirements for software integrity and security will be proposed, in addition to the procedures to check the compliance with these requirements. In the future, these items could be incorporated to the current Inmetro requirements. Currently, the project is in the phase of structuring the laboratories, acquiring equipment and testing tools.

V. CONCLUSIONS

By the beginning of 2014, power supply companies in Brazil will have to be ready to replace all their current meters. Considering the preliminary analysis, based on scientific papers and presentations that took place during the main security conferences, the replacement of meters may cause irreparable damage to the Brazilian energy sector. For this reason, research and development investments in the evaluation of software security for smart meters is crucial, mainly in the current Brazilian scenario. The results of this program will help the entire energy sector in Brazil through the creation of the first laboratory in the country, concentrating on information security tests including the specification of all features, tools and procedures to check the smart meter robustness for security baseline.

REFERENCES

[1] S. Fries, "Securing the Smart Grid", The First International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies - ENERGY 2011, Keynote Speaker, Presentation, May 2011.

[2] Technical briefing: "Attacking the Smart Grid," Insights on governance, risk, and compliance - Ernst & Young, December 2011.

[3] H. Khurana, "Smart-Grid Security Issues", Security & Privacy, IEEE, Volume: 8 , Issue: 1, pp. 81-85, February 2010.

[4] W. Sikora, M. Carpenter, and J. Wright, "Smart Grid and AMI Security Concerns", InGuardians & Industrial Defender, Presentation, July 2009.

[5] Art Presse, Pirataria em TV por assinatura chega a 11,44 milhões de lares na América Latina, October 2012.

[6] B. Murrill, E. Liu, and R. Thompson II, "Smart Meter Data: Privacy and Cybersecurity", Congressional Research Service, February 2012.

[7] N. Falliere, L. Murchu, and E. Chien, "W32.Stuxnet Dossier", Symantec Security Response, Version 1.4, February 2011.

[8] M. Davis (IOActive), "SmartGrid Device Security: Adventures in a new medium", Black Hat USA, July 2009.

[9] D. Weber (InGuardians), "Looking into the eye of the meter", DEFCON 2012, July 2012.

[10] InGuardians, Inc., "Advanced Metering Infrastructure Attack Methodology", March 2011.

[11] J. Searle (UtiliSec), "Dissecting Smart Meters", Black Hat Europe 2012, March 2012.

[12] American National Standard Protocol Specification for ANSI Type 2 Optical Port, ANSI C12.18-1996, April 1996.

[13] sKyWIper Analysis Team, "sKyWIper (a.k.a. Flame a.k.a. Flamer): A complex malware for targeted attacks", May 2012.

[14] M. Stevens, A. Sotirov, J. Appelbaum, A. Lenstra, D. Molnar, D. A. Osvik, and B. de Weger, "Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate", in CRYPTO 2009, Lecture Notes in Computer Science, vol. 5677, pp. 55-69, August 2009.