Ancillary Service Provision by Demand Side Management: A Real-Time Power Hardware-in-the-loop Co-simulation Demonstration

M. H. Syed, P. Crolla and G. M. Burt
Institute for Energy and Environment
Dept. of Electronics and Electrical Engineering
University of Strathclyde, Glasgow, Scotland
mazheruddin.syed@strath.ac.uk

Abstract—The role of demand side management in providing ancillary services to the network is an active topic of research. However, their implementation is limited due to lack of practical demonstrations and tests that can rigorously quantify their ability to support the grid’s integrity. In this paper, provision of time critical frequency control ancillary service is demonstrated by means of integrating PowerMatcher, a well discussed demand side mechanism in literature, with real-time power hardware. The co-simulation platform enables testing of demand side management techniques to provide ancillary services.

Keywords—Ancillary services, demand side management, transactive energy markets and real-time power hardware-in-the-loop simulation.

I. INTRODUCTION

Ancillary services in a power system are essential to maintain the integrity of the grid. Ensuring sufficient procurement and delivery of ancillary services is expensive. A total price of £30.72 million had been paid by National Grid (Transmission System Operator- Great Britain) to ensure the availability and provision of ancillary service during the month of August 2014 [1]. Ancillary services can be broadly classified into three: (i) Frequency Control Ancillary Services (FCAS), (ii) Network Control Ancillary Services (NCAS) and (iii) System Restart Ancillary Services (SRAS) [2].

Both, demand and generation, in theory, can participate in frequency control. Although demand is used to restore severe load generation imbalance that cannot be alleviated by fast acting generators (by interrupting blocks of load by underfrequency relays), their capability to contribute to frequency control has been underestimated in the past due to the complexity involved in their real-time monitoring and control. On the other side, with the constant increase in penetration of intermittent renewable generation and decrease in conventional generators due to the global aim of decarbonizing the power grid, frequency control by supply side is not just expensive but technically challenging. Alternative methods to provision of ancillary services are being sought by Transmission System Operators in order to reduce the overall costs involved in maintaining the security of supply. With the advancements in measuring and monitoring techniques, demand side can take an active role in control of system frequency.

The selection of the type of load that can be utilized to provide frequency support to the network is important. The best candidates for this purpose are the loads whose utility to the consumer is a function of energy consumed over a period of time rather than instantaneous power consumption [3]. Known examples of such loads are heating, cooling and pumping devices (also referred to as Thermistically Controlled Loads). The cyclic nature of these loads enables short interruptions that would be acceptable to the consumer. The methods in which these flexible loads can be controlled were classified into two, namely Direct Load Control and Indirect Load Control, in [4]. In direct load control, the power consumption of these flexible loads is directly controlled by the utility or an aggregator regardless of the consumers’ will. This method offers an easy way to control the imbalance but does not maintain acceptable levels of customer utility. On the other hand, in indirect load control, the consumers can control the consumption of loads in response to the incentives given to them, for example: real time electricity pricing signal.

Another way to classify the control mechanisms of load in demand side management (DSM) would be centralized control and decentralized control. In centralized control, the power consumption of the loads can be manipulated automatically or manually (either by utilities, aggregators or customers) in response to a signal received from a center responsible for its operation and control [5-8]. In contrast, decentralized control refers to a control in which the power consumption of the loads is manipulated automatically or manually by local measurements and settings [9-14]. Further, a mix of centralized and decentralized control is also possible where the power consumption of the devices corresponds to both the central signal and the local measurement and settings [15, 16].

Certainly, DSM has the potential to provide ancillary services and Transmission System Operators are keen to exploit it. However, there is a lack of experimental evaluation that would bolster confidence in their wide scale deployment. In order to rigorously characterize the real-world performance of DSM in provision of ancillary services, a controlled and scientific experimental environment is required [17]. The objective of this work is to provide a platform that would enable the evaluation of these mechanisms. By development of a real-time power hardware-in-the-loop co-simulation platform,
the provision of ancillary services by DSM can be evaluated. Further, the experimental evaluation of these mechanisms would provide an insight to the implications of implementing these systems on a larger scale in a real network.

II. POWERMATCHER

PowerMatcher (PM) is a DSM technology that coordinates a cluster of electricity producing and consuming devices by creating a transactive energy market [18]. It exploits the capability of multi-agent system framework to implement scalable, distributed, complex and open Information and Communication Technology infrastructures. In a multi-agent based environment, theoretically, multiple software agents interact and negotiate to reach a common system goal while preserving their local goal [19-21]. Similarly, in a PM cluster, the interest of each device, either producing or consuming electricity, is represented by a device agent. A centralized auctioneering agent in PM facilitates the devices to trade in the transactive energy market. Each device in the PM cluster sends a bid to the auctioneering agent. The bid is indicative of the priority or willingness of the device to consume or produce electricity at a given price. The auctioneering agent aggregates the bid and a market clearing price is determined from the aggregated bid. The market clearing price is then communicated back to the device agents, which depending upon their bid either start consuming or producing electricity or wait till the market price/device priority changes. The only information that is communicated from the device agents to the auctioneering agent is the bid and price from auctioneering agent to the device agents. The communication between the device agents and the auctioneer is kept to a minimum as the bids propagated are event based, i.e. a new bid is generated only if the device priority changes.

III. EXPERIMENTAL SETUP

The aim of this paper is to test the provision of FCAS by PowerMatcher. This is done by means of a real-time power hardware-in-the-loop co-simulation platform. The co-simulation platform is the product of integration of a PowerMatcher implementation in Matlab (PM simulation) with the Distribution Network and Protection (DNAP) laboratory [22] at University of Strathclyde. The experimental setup is composed of two parts: firstly, simulating a cluster of loads and generation in PM simulation; secondly the integration of PM with the DNAP laboratory. The experimental setup is explained in detail in the following sub-sections.

A. PowerMatcher Cluster Simulation

In this paper, for the purpose of experimental demonstration, the PM cluster comprises of an islanded microgrid that serves one thousand residential houses. The loads at a residential house can be typically classified into two: non-flexible and flexible loads. The term non-flexible load in this paper refers to devices that must be turned on irrespective of the incentive (for example, lighting, computers, laptops, TV’s, etc.). The term flexible load refers to the devices that can alter their states based on the incentive provided. The statistics for number of bedrooms and occupants of the houses have been taken from [23]. The capacity of heat pump (COP = 3) installed at each house has been estimated depending upon the number of occupants in the house. The parameters of PowerMatcher load cluster have been presented in Table I. The ability of the cluster to provide ancillary services depends upon the flexibility available in the cluster. The flexibility available in the cluster is dependent upon the flexible load. In this experiment the flexibility is decided by the buffer set upon the heat pump for space heating.

In order to determine the day-ahead commitment of the generators in the cluster, an estimate load profile needs to be obtained. A load profile for a single day has been obtained for this work. To obtain a realistic load profile, the number of occupants in the house and their occupancy pattern needs to be taken into consideration. The fixed load profile has been obtained by the method presented in [24], that based on a probabilistic occupancy pattern determines the load profile for a single day. The flexible load profile has been generated in PM Simulation. To generate the day-ahead estimate for the flexible loads, these loads were operated using a standard on/off temperature control, i.e. the heat pump will heat the water from 30 °C to 45 °C (temperature range of water in the radiators) and then waits until the water temperature reaches back to 30 °C before turning on again. This method has been chosen for simplicity, however, this load estimation can be obtained by any other forecasting method presented in literature [25]. At the beginning of the simulation, a temperature is randomly allotted to each heat pump in the range of 30 °C to 45 °C. The aggregated fixed and flexible load profile for one thousand houses is presented in Fig. 1.

It is crucial for sufficient randomness, in terms of customer behavior, to be present for load and generation balance to be maintained. The PM cluster has been modeled with sufficient randomness for this work (the rate at which heat is lost at each house is randomly chosen within a specific range). This is a valid assumption for modeling as each house in the network would have different insulation and heat retention property.

<table>
<thead>
<tr>
<th>Table I, PowerMatcher Load Cluster Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of</td>
</tr>
<tr>
<td>Bedrooms</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Fig. 1. Forecasted load profile.
B. PowerMatcher Integration with DNAP Laboratory

The architecture of the PM and DNAP laboratory real-time power hardware-in-the-loop co-simulation platform has been presented in Fig. 2. The architecture consists of three modules: 1) DSM module, 2) Real Time Simulator (RTS: by Applied Dynamics International) module and 3) Micro-grid module. The DSM module incorporates the PM as the DSM technology. However, PM can be replaced by any other DSM technology from the literature. The DNAP laboratory micro-grid comprises of 80kVA and 2kVA synchronous generator, 4 induction machines (2X7.5kW, 5kW and 3kW), 15kVA Triphase inverter unit and 60kW load banks (2X10kW and 1X40kW). For the co-simulation platform, only the 80kVA synchronous generator and the load banks have been used. The RTS module has 2 RTS units and 2 host personal computers (PC) for processing the measurements, controlling the micro-grid and observing the network parameters. The RTS module is the interface between the micro-grid and the outside world (DSM module and laboratory users).

PM simulation runs on a PM host PC. One of the important aspects of integrating PM with the DNAP laboratory is setting up the communication between PM and the RTS module. For the purpose of a closed loop co-simulation, two values need to be communicated between PM and the RTS: 1) the power generation available, from the RTS module to the PM and 2) load power, from PM to the RTS module. User Datagram Protocol (UDP) is chosen for transmission of message from PM host PC to RTS units. TCP/IP has not been chosen for this purpose as the overheads involved in establishing the connection might not be suitable for real-time applications.

In the real-time power hardware-in-the-loop co-simulation, at each time step the forecasted/estimated load profile from the previous step is read by the RTS host PC. As the physical capacity of the micro-grid is limited to 60kW by the load banks, a scaling factor is used. The scaled generation profile is then sent from the RTS to the synchronous generator. The measurements from the micro-grid are sent to the RTS for processing. The measured value of generation is then scaled back and sent as an input to the PM. The PM then decides the amount of load that needs to be on in that particular time step depending upon the generation available. The aggregated amount of load is then sent back to the RTS units. The aggregated load is scaled down and sent to the load banks proportionately according to their capacity. All the important measurements of the network are available for observation at the RTS host PC.

The graphical user interface (GUI) at the RTS host PC provides us with an important functionality to alter the inputs of the micro-grid control in real-time. The value of generation sent as an input to the synchronous generator can be modified in real time with the GUI. This functionality allows us to emulate any unplanned disturbance in the network. The modification by the GUI is reflected in the PM as the input to PM is not directly from the RTS host PC but is the measured power from the micro-grid.

As demand side management simulations are usually performed for a larger period of time that may involve either one full day or sometimes one full year, the time step for simulation plays an important role in real-time power hardware-in-the-loop co-simulation. Performing a co-simulation for the period of a day can be expedited by choosing an appropriate ratio between real-time and virtual simulation time. However, care must be taken as a very small ratio might affect the integrity of the co-simulation. Depending upon the
application for which the co-simulation is performed, a suitable ratio can be chosen. In applications where system dynamics play an important role, it is better to simulate parts of the day that are of interest. For the purpose of this work, it was necessary to allow sufficient time for the synchronous generator to respond to events that would enable us to capture its frequency response. A ratio of 6 has been chosen.

IV. EXPERIMENTAL EVALUATION AND RESULTS

With the integration of PM with the DNAP laboratory and development of the co-simulation platform, it is necessary to first validate its performance. In this section, first, the co-simulation platform will be validated. Second, the provision of FCAS provision by PM will be tested.

A. Validation

For the purpose of validation of the co-simulation platform, the forecasted/estimated generation profile from the PM simulation (fixed+flexible load as shown in Fig. 3(a)) is given as an input to the RTS host PC. The flexible loads in the PM cluster are now controlled by PM. In order to avoid frequent switching of the flexible loads, a delay between each turn on/off has been used, i.e. once the heat pump is either turned on or off, it continues to stay on or off for a duration of two minutes. This is done to eliminate the possibility of an oscillating response from the heat pumps and to ensure a long operational life of the equipment. The minimum and maximum price of electricity has been set as 5 and 30 Euro cents. The results for the validation have been presented in Fig. 3. As can be observed from Fig. 3 (b), the load is in good conformity with the generation in the micro-grid. There might be small difference between the load and the generation due to the fixed loads in the micro-grid that cannot be controlled. This is evident from the frequency response of the network shown in Fig. 3 (d). The frequency of the micro-grid is within the operational limits of 49.8 -50.2 Hz. There are a few peaks that extend beyond the 50.2 Hz but are for a very small duration of time and do not exceed 50.225 at any time during the simulation.

The temperature profile of the heat pumps has been presented in Fig. 3 (c). At the beginning of the simulation, a temperature is allocated to each heat pump randomly. The temperature of each heat pump is well within the temperature range that had been set. Further, the PM controller maintains a higher level of comfort for the same amount of generation available compared to the no control heat pump. However, the performance of PM, in terms of temperature comfort, has not been evaluated with any other controller as it is not the focus of this work.

The price of the electricity for the duration of simulation has been presented in Fig. 3 (e). The price of the electricity is higher at the beginning of the simulation as the generation in the network is limited. Further, as the heat pumps start at a random temperature, it takes some time for them to reach temperature equilibrium. In real-time application of DSM, price can be used as an actual incentive, where the regulations allow for a dynamic pricing, or can be used as a control signal. The number of heat pumps on at every time step have been presented in Fig. 3 (f). It is interesting to note that at no point the total number of heat pumps on is more than 260 out of
B. FCAS Provision Testing

In order to test the provision of FCAS, the generation input to the micro-grid is modified by the GUI of RTS host PC. Once the generation input is read by the RTS host PC, any modification can be made in real time by the GUI and is sent to the synchronous generator. The results for the provision of FCAS have been presented in Fig. 4. As can be observed from Fig. 4 (a), two disturbances have been simulated for the duration of the day. First, a generation loss of 300kW is emulated by modifying the input from the GUI at 07:00 hours. The generation loss is simulated for duration of 2.5 hours. At 19:30 hours an excess of 300kW generation is simulated. This can correspond to an increase in generation from a wind turbine.

The modified profile given as input to the synchronous generator can be seen in Fig. 4 (a). As can be seen from Fig. 4(b), at the instant the generation is lost, PM is able to match the loads with the available generation. When the generation is lost, the price of the electricity starts to increase (Fig. 4 (e)). The heat pumps respond to increase in price by turning off (as shown in Fig. 4 (f)). PM is able to match the load till the point the temperature of the heat pumps is above the minimum set level. Once the temperature reaches the minimum level, the heat pumps have to turn on to maintain the level of comfort of the customers. As can be observed from Fig. 4, once the temperature reaches 30 °C, the load no longer is able to follow the generation profile. At this point, although the price of electricity is maximum, it is no longer able to control the heat pumps as the heat pumps are ready to pay even a higher price to maintain the comfort of the customer. The effect of load not being able to follow the generation can be observed in the frequency of the network that falls below the operational limit of 49.8 Hz.

When there is an increase in generation, the PM is able to maintain the balance between the generation and the load. The price of electricity decreases. The heat pumps respond to the decreasing price by turning on. As can be seen from the temperature profile, the temperature of the heat pumps starts to increase. Once the temperature reaches a maximum set level of 45 °C, PM can no longer maintain the balance between the generation and load although the price of electricity is minimum. The mismatch in load and generation is reflected in the frequency of the network that is no longer within the operational limits of 50.2 Hz.

The capability of PM to maintain a balance between generation and load is highly dependent upon the flexibility available in the cluster. The flexibility in this case is the room available for increasing or decreasing the temperature of the heat pumps. However, the focus of this paper remains to test the ability of PM to provide FCAS. As has been presented in the results, the instantaneous disturbances in the micro-grid had been effectively mitigated by PM. The requirement of FCAS is to be able to respond to changes in generation as soon as possible. In both cases presented (increase and decrease in generation), PM was successfully able to respond to the changes in a fast manner as shown in Fig. 4(b) (zoomed). The
frequency returned to the nominal level within 2.5 s (real-time) in both cases. Although an acceleration factor was used for PM simulation, the calculation of the load for each instant is done within 150 ms of receiving the measured power. This is evident from the load being able to follow the generation profile. The response of the generator is not affected by the acceleration of the PM simulation. From the above results, it can be said that PM starts to respond to a change in generation within 150 ms and was capable to restore the frequency within 2.5 seconds. The duration of the disturbance has been chosen to be able to show the saturation effect of flexibility in the network.

It is worthy to mention that the results presented in this paper correspond to the system under study. Further, it has been assumed that an IP-based communication infrastructure is available for implementation of PM.

V. CONCLUSIONS AND FUTURE WORK

This paper has introduced a new extended real-time power hardware-in-the-loop co-simulation platform achieved by the integration of PowerMatcher (PM) with the University of Strathclyde’s DNAP micro-grid. The co-simulation platform has been validated by the conduct of an initial experiment comprising a micro-grid hosting 1000 residential houses controlled by PM. For the first time, the implementation of PM with a physical micro-grid has been demonstrated. The provision of Frequency Control Ancillary Services (FCAS) by PM is tested. The experimental results validate the capability of PM to provide frequency control. In cases of increase and decrease in generation, PM was able to successfully restore the system frequency within 2.5 s. The performance of PM at its limits has been demonstrated, where PM can no longer maintain a balance of generation and load. This is due to the fact that customer utility is a priority in PM operation. These experiments prove the potential for the co-simulation platform to evaluate and characterize the real-world performance of DSM techniques in providing grid ancillary services. Further, the design of the real-time interface module, presented in this paper, supports the evaluation of the impact of communication, bandwidth and scalability on the performance of DSM in ancillary service provision.

The contribution of this paper is to open up a series of stepping stones towards the wide scale deployment of demand side management in provision of ancillary services. The development of real-time power hardware-in-the-loop co-simulation platform serves as the basis for testing ancillary service provision by demand side management. Future work involves testing the performance of PM when integrated in the network. PM’s performance under different scenarios of ancillary services provision will be experimentally tested and verified. Further, the interactions of two or more instances of PM, operating independently, will be investigated.

ACKNOWLEDGMENT

The authors would like to thank Mr. R. Munro for his continuous support. The work in this paper has partially been financed by TNO under the ETP Program “Adaptive Multi-Sensor Networks” and the ERP Program “Energy Storage and Conversion”.

REFERENCES


