Real-Time Data and Energy Management in Microgrids

Zhichuan Huang and Ting Zhu
University of Maryland, Baltimore County
zhihu1@umbc.edu, zt@umbc.edu

***Abstract*—Microgrids are desired in remote areas, such as islands and under developed countries. However, given the limited capacities of local energy generation and storage in such a community, it is extremely challenging for an isolated microgrid to balance the power demand and generation in real-time with dynamically changing energy demand. Meanwhile, more and more sensing devices (such as smart meters) are deployed in individual homes to monitor real-time energy data, which can be helpful for homes and microgrid to better schedule the workload and generation. However, it is still difficult to conduct real-time distributed control due to the unreliable sensing devices and communications between sensing devices and controllers. To address these issues in microgrids, we designed a novel approach for the system to i) process the collected sensing data, ii) reconstruct the missing data caused by sensing error or unreliable communication, and iii) predict the future demand for real-time distributed control with missing data in extreme situations. The control center then decides the operations of the local generator and each home decides the scheduling of the flexible workload of appliances based on the collected and predicted data. We conducted extensive experiments and simulations with real world energy consumption data from 100 homes for one year. The evaluation results show that our design can recover the missing data with more than 99% accuracy and our distributed control can balance power demand and generation in real-time and reduce the operational cost by 23%.**

I. Introduction

Microgrids play a important role in energy cyber-physical systems [1]. In a typical microgrid, it consists of local generators and energy storage (e.g., batteries) to provide power for a small community with commercial and residential buildings. Microgrids can provide power to places i) where the traditional power grid does not exist due to the poor economy or limited number of residences (e.g., islands); and ii) when the traditional power grid is temporarily not functioning due to severe weather conditions (e.g., storms). Therefore, microgrids have gained increasing attention recently [2]. Due to the very limited capacities of local energy storage and energy generation, microgrids are more difficult to maintain than traditional power grids. To ensure the stability and reliability of a microgrid, we need to conduct real-time scheduling and control the operations of local generators, batteries, and controllable workloads of appliances to offset the dynamically changing power demands of uncontrollable appliances.

Therefore, it is extremely important to collect the power consumption and generation data for distributed control in real-time, which becomes possible with the recent rapid development of smart meters. However, the sensors in smart meters and wireless communications between smart meters and controllers are not 100% reliable. The microgrid controller may encounter missing data and delayed data when conducting the scheduling for generators and distributed control operations in each home. Furthermore, to cope with the dynamically changing demand, the microgrid controller needs information to predict the future power demand to decide the operations of local generators. Meanwhile, each home also needs to predict its own future demand to schedule workloads. The existing techniques on energy consumption forecasting are mainly for long term offline forecast for large generators [3], [4]. However, to realize real-time distributed control in a microgrid, real-time data processing and short term prediction is needed. In this paper, we propose a novel data management technique for distributed control in a microgrid to process the received data, reconstruct the missing data, and predict the future power demand with existing data. The key idea is to utilize the correlation between power, voltage and frequency data of different homes in the microgrid. Then, the misses or delayed data can be reconstructed with a portion of received data or data from other homes. Because energy consumption patterns in one home are limited due to the limited number of appliances, it provides us opportunity to reconstruct the missing data and predict future data in a short term based on existing data and detected energy consumption patterns. With the reconstructed and predicted data, the controller decides the scheduling of workloads in each home and the operations of generators to maintain the stability of the microgrid.

While the workload is scheduled in each home to avoid power failures in the microgrid, the behaviors and comfort of users should not be affected. Therefore, we choose the flexible and controllable workloads of appliances for scheduling. For example, water heaters are flexible and controllable loads because we only need to make sure that there is enough hot water in water heaters when people use hot water. Our approach can be easily extended to support any other types of flexible and controllable workload (e.g., HVAC). For the local generator's scheduling, we adopt a widely used generator model and propose an optimal algorithm to minimize the operational cost. Specifically, we summarize our major contributions as follows:

- We conducted a systematic investigation on the correlation between power, voltage, and frequency in a microgrid and developed a holistic sets of correlations models (i.e., power-voltage, frequency-voltage, temporal, and spatial correlation). Through extensive experiments and simulations, we show that

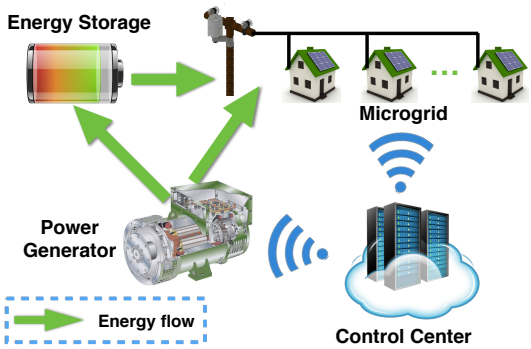


Fig. 1: Architecture of a microgrid

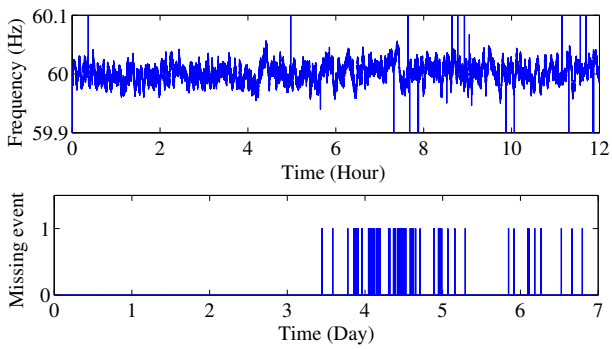


Fig. 2: Examples of data faults in energy monitoring

our design can recover the missing data with more than 99% accuracy for the short term prediction.

• To reduce the operational cost of isolated microgrids, we present holistic real-time scheduling algorithms for both local generators and controllable loads in individual homes even when there exists communication failures between control center and individual homes.

• Utilizing the empirical energy consumption data from 100 residential homes, we conducted extensive simulations. The results indicate that our proposed distributed control can reliably balance power demand and generation in real-time and reduce operational cost by 23%.

II. Background and Motivation

A microgrid is a distributed electric power system that can autonomously coordinate local generations and demands in a dynamic manner [5]. Microgrids can operate in either grid-connected mode or isolated mode, some of which are now deployed in the US, Japan and European countries [6].

Background. In this paper, we consider a modern microgrid, illustrated in Figure 1, consists of generation technology (e.g., local electricity generators) and batteries. To ensure compatibility with the traditional power grid, we adopt the microgrid architecture, which is similar to the one used in a traditional power grid. If the microgrid is built from nothing (e.g., island, where there is no electricity grid before), the microgrid can be built the same architecture as traditional grid with a distribution network across the community of homes. If the microgrid is built from a traditional grid, we only need to add local generators, batteries and a control center into the microgrid. Within the microgrid, sensors are deployed in each home to collect and send energy related data (e.g., power, voltage and frequency) to the control center. The control center decides the workload scheduling in each home and the generations to balance the power demand and supply.

Motivation. To realize the real-time control, it is very important to collect the energy related data from homes and send back the control instructions in real-time. However, based on our more than 6 years' experiences of energy monitoring in residential homes, the data collection from homes may suffer from different types of faults: i) data point missing; ii) sensing error; iii) communication delay; and iv) communication loss. The first two faults are caused by the low reliability of sensors due to the long-term monitoring. The latter two faults are caused by unreliable wireless communication. We show some examples of faults in Figure 2. The first one is caused by sensing errors, which generate peaks but do not happen very frequently (we observe average 1.5 seconds sensing error in 12 hours). The second one is whether we receive readings in the cloud server. The Y-axis value is set to be 1 if there is a data missing event. We can see the missing events are very bursty, which means once we have a missing event, there will be high probability there would be missing events in near future.

With the demand of real-time data collection and reality of multiple different faults in monitoring, it is crucial to manage the real-time collected data and reconstruct the missing data for real-time control in a self-sustainable microgrid.

III. System Overview

To ensure the reliability of the microgrid, we propose the system design as shown in Figure 3, which includes three main components: data management, central scheduler, and local scheduler. In summary, our system works as follows: i) power meters deployed in homes monitor the power consumption, voltage and frequency in the power line, then send collected data to control center; ii) control center receives the collected data from homes and processes the data for missing data reconstruction and future data prediction; iii) the central scheduler decides the control instructions for each home and generators based on the processed data; iv) individual homes and the power generator execute the instructions from central scheduler if control instructions are received; v) if control instructions are not received by the individual homes and the power generator, local scheduler will conduct local control in these homes and the power generator will maintain the same amount of power generation.

Data Management. Due to the sensing errors and unreliable communication, it is highly possible we will miss important energy data from sensors. To reconstruct and predict the sensing data, we investigate the correlation models among energy data for recovery. The received data will be used both for data reconstruction and update for correlation model. Specifically, we investigate i) correlation between power and voltage for homes under the same transformer; ii) correlation between frequency and voltage at individual homes; iii) temporal and spatial correlation for power consumption of all homes. Based on these correlations, the missing data can be reconstructed and future data can be predicted for real-time control.

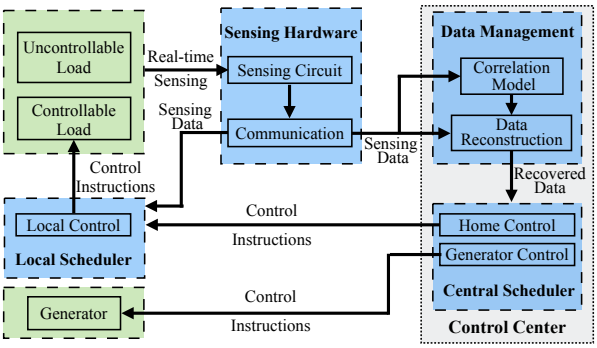


Fig. 3: System overview

Central and Local Scheduler. Based on recovered and predicted data, the central scheduler decides control instructions for both controllable workload in homes and the generation from the power generator. The key idea of workload control in each home is to turn on some appliances when power consumption is low and turn off some appliances when power consumption is high. In our paper, we use the workload of the water heater as an example because its workload is flexible and it is commonly installed in residential homes. Our design can easily be extended to support the scheduling of HVAC systems. The generation control is to decide the power supply from generators. Because the power supply of generators cannot be changed as fast as workload at homes, we schedule the generators based on prediction of future power demand in the microgrid. Batteries are used as buffer to offset the prediction errors of the future power demand. Due to the unreliable communication, control instructions may not be received at the local home or generator, then the local scheduler conducts local control based on the local sensing data.

IV. Data Management

To reconstruct and predict the sensing data, in this section, we introduce four correlation models for reconstructing the missing data and predicting the future data. While most of existing works focus on missing data reconstruction of a single time series data [7], [8], we investigate the correlations among multiple time series data (power consumption from multiple homes, voltage and frequency) in microgrids and utilize the correlation models to reconstruct the missing data.

A. Correlation Models

The most important part of data management is to build and utilize the correlation models among the collected data. Specifically, we identified and built four correlation models: i) correlation between power and voltage; ii) correlation between frequency and voltage; iii) temporal correlation of power consumption data in a single home; and iv) spatial correlation between power consumption data from multiple homes.

1) Power Voltage Correlation: Without loss of generality, we assume that N homes are connected under the same transformer (shown in Figure 4). According to the Electric Power Distribution Handbook [9], a transformer can be considered as a constant kVA device for a voltage from 100% to 105%. If the power consumption of one home increases, the total

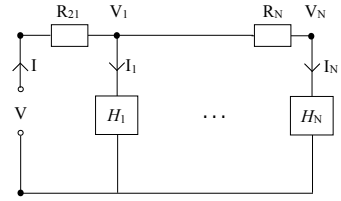


Fig. 4: The topology of homes and the transformer

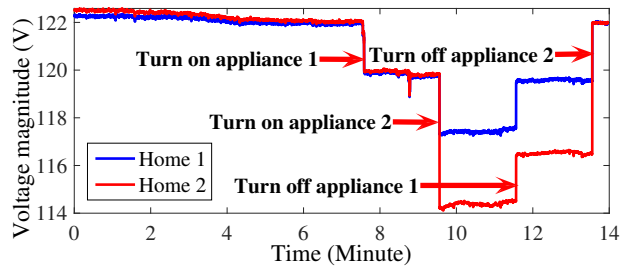


Fig. 5: Relationship between power and voltage

current I increases and voltage V drops. We find that home H_{i-1}'s voltage value depends on i) the transformer's output voltage (V); ii) the current from the transformer to H_{i-1}; and iii) resistances of the power line from the transformer to H_{i-1}. For example, H_1's voltage value only depends on the transformer's output voltage (V), the current (I_1) through H_1, and the resistance (i.e., R_1). Based on the above analysis, the voltage values at homes H_{i-1} and H_i can be calculated by using Equation (1).

$$V_{i-1} = V - \sum_{j=1}^{i} \sum_{k=j}^{N} I_{k-1} R_{j-1} \quad i = 1, 2, ..., N \tag{1}$$

To verify it, we conduct experiments with 2 homes under the same transformer and keep power consumption at home 2 stable to study the power voltage relationship. The measured voltage in both homes are both related to the power consumption in home 1 (shown in Figure 5). Thus, based on Equation (1) and evaluation results, the voltage drop from transformer to each home is in linear relationship of currents going through the power line.

2) Frequency Voltage Relationship: A typical microgrid may contain multiple transformers. Therefore, it is also important to investigate the other features in a microgrid. According to the Electric Power Distribution Handbook [9], frequency is a good indicator of the relationship between power supply and demand. If the power demand surpasses the power supply, then the frequency decreases because the generator can not generate enough power. Thus, the frequency should be related to the total power consumption of the microgrid. Because voltage is related to power supply and demand, thus, the frequency value has a linear relationship with the voltage value. This relationship can be modeled as follows:

$$\triangle F = \triangle V * \lambda_1 \tag{2}$$

To verify it, we conduct experiments with 3 homes with one month data. Two of them are under the same transformer while the other home is under a different transformer. To make the relationship clear to see, a typical example of the measured

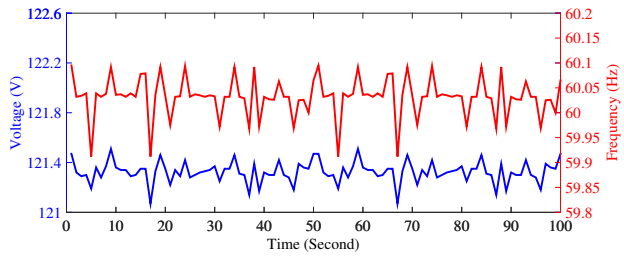


Fig. 6: Relationship between frequency and voltage

frequency and voltage relationship is shown in Figure 6. The frequency value is well synchronized with the voltage value.

Based on Equation (2) and experimental results, the frequency change in each home is in linear relationship of voltage change. Therefore, we can utilize the frequency voltage relationship to recover the missing data.

3) Temporal Correlation: In a microgrid, we need to reconstruct the missing data and predict the power consumption in the very near future (e.g., next 1 second) for real-time control. Thus the power consumption data of yesterday or last month is much less useful. To address this problem, we leverage the power consumption signatures of appliances to reconstruct the missing data and predict the short term power consumption. This is because there are limited number of power consumption signatures for different loads. To evaluate this approach, we use the empirical data collected for 2 years from a home to investigate the temporal correlation between power consumption data. We run a power consumption signature detection algorithm on the data set to find the signatures. Specifically, the similarity between two vectors is calculated by using a Euclidean distance-based function as shown below:

$$\rho_{i,j} = \frac{1}{|S_i|}\sum_{t=1}^{|S_i|}(S_i(t) - S_j(t))^2 \tag{3}$$

$|S_i|$ is the length of signature S_i. If the distance of these two vectors is small, then the similarity of two vectors is high. Then we go through the data set to find the possible signatures. To simplify the algorithm, we use a fixed length of energy consumption patterns which achieves very effective results (detailed in Section VI). As shown in Algorithm 1, the signature set S is empty initially. When $t < T$, we calculate the similarity between power consumption data and signatures of each appliances based on Equation (3). If we find the similarity between current power consumption and existing signatures is higher than the current maximum similarity, we reassign the maximum similarity and mark $index = i$. Then we compare the maximum similarity we find to the threshold of the minimum similarity ρ_{min}. If $\rho_{max} > \rho_{min}$, we then detect a new signature S_{new}, add it to the signature set S and update $t = t + |S_{new}|$. Otherwise, we update $t = t + 1$ and continue the detection process.

Based on the detection signatures, we can reconstruct the missing data and predict the near future data as:

$$P_i(t+k) = S_j(|S_j|/2+k) + \sum_{x=1}^{|S_j|/2}\frac{2\cdot(P_i(t+x-|S_j|) - S_j(x))}{|S_j|} \tag{4}$$

Algorithm 1 Signatures Detection Algorithm

```
S = ∅;
while t < T do
    ρ_max = 0, index = −1;
    for detected signature S_i do
        Calculate ρ_i(t) based on Equation (3);
        if ρ_i(t) < ρ_max then
            ρ_max = ρ_i(t), index = i,
        end if
    end for
    if ρ_max > ρ_min then
        Detect a new signature S_new and add to S;
        t = t + |S_i|;
    else
        t = t + 1;
    end if
end while
```

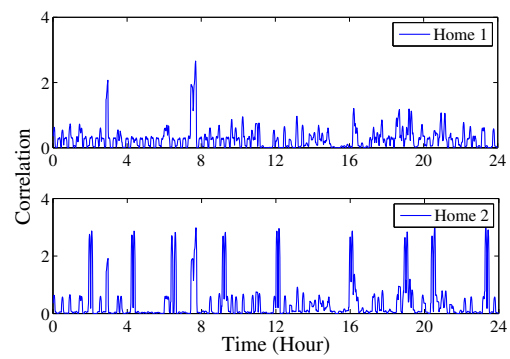


Fig. 7: Spatial correlation of different homes over time

where S_j is the detected signature that has shortest distance to $\{P_i(t - l(S_j) + 1), \cdots, P_i(t)\}$.

4) Spatial Correlation: Because homes in the same area may have the similar power consumption pattern, we can use the spatial correlation among power consumption of homes for power consumption prediction. However, different homes will have different correlations at different time. Thus, we need to keep on updating the correlations among these homes for prediction. To evaluate our idea, we use empirical power consumption data collected from 100 homes for one month to investigate the spatial correlation of power consumption among different homes. The spatial correlation between homes 1 and 3, and homes 2 and 3 is shown in Figure 7. X-axis is the time, Y-axis is the correlation between either homes 1 and 3 or homes 2 and 3. For most of the time, home 1 and home 2 are quite similar to home 3. However, from hour 2 to hour 4, home 1 is closer to home 3 while from hour 4 to 6, home 2 is closer to home 3. Thus, we build a model to predict the power consumption based on historical correlations among homes. The correlation between two homes can be calculated as:

$$c_{ij}(t) = \frac{1}{l}\sum_{x=t-l}^{t}(P_i(x) - P_j(x) - \frac{1}{l}\sum_{x=t-l}^{t}(P_i(x) - P_j(x)))^2 \tag{5}$$

From the above equation, we can predict $p_i(t)$ based on readings from other homes:

$$P_i(t) = \sum_{j=1}^{N}\frac{P_j(t) * c_{ij}(t)}{\sum_{j=1}^{N} c_{ij}(t)} \tag{6}$$

If $c_{ij}(t)$ does not exist because of missing data, then we replace $c_{ij}(t)$ with $c_{ij}(t_k)$. Where t_k is the latest time for updating the correlation between homes i and j.

B. Data Reconstruction and Prediction

With the above four correlation models, we can reconstruct the missing data and predict the future data for real-time control. In order to schedule the controllable workload in each home and maintain the stability of the microgrid, the control center needs to collect the data of real-time power consumption, voltage, and frequency. Then the reconstruction process is executed as follows: i) if only part of the data is missing in one home, the power voltage relationship can be used to reconstruct the power consumption or voltage; ii) if only frequency data is collected from one home, frequency voltage correlation can be used to reconstruct the voltage and then apply power voltage relationship to recover power; iii) if no data is collected from one home, we can utilize the temporal and spatial correlation to reconstruct the data; iv) if no data is collected from any homes, only temporal correlation can be used to reconstruct the data; and v) if all the data is collected from one home, the collected data is applied to update the correlation weight for temporal and spatial correlation models.

The prediction process is the same as the scenario that no data is collected from any homes. Note the prediction with temporal correlation is only accurate for short-term data missing. Because generators can not be turned on/off very frequently and the generation control needs long-term power consumption prediction, other traditional consumption prediction algorithms can be applied in this scenario.

V. Central and Local Schedulers

With the reconstructed data and predicted future data, central and local schedulers need to schedule the future generation of generators and the controllable workload in each home. The design goal is to balance the power demand and generation with minimum operation cost in the microgrid.

A. Design Goal

Assume there are N homes in the microgrid and the power consumption of home i at time t is $g(t)$. The microgrid has M units of homogeneous local generators, each has a maximum power output capacity L. Based on a common generator model [10], we denote β as the startup cost of turning on a generator. Startup cost β typically involves the heating up cost (in order to produce high pressure gas or steam to drive the engine) and the time-amortized additional maintenance costs resulted from each startup (e.g., fatigue and possible permanent damage resulted by stresses during startups). We denote y_m as the sunk cost of maintaining a generator in its active state per unit time, and y_o as the operational cost per unit time for an active generator to output an additional unit of energy. Note that our design is not limited to any generator model. Table I summarizes the definition of parameters. Our design goal is to balance the power demand and generation while minimizing the operational cost of power generators. The problem can formulated as follows:

Notation	Definition
$b_i(t)$	Amount of energy in battery of home i at time t
B_i	Battery's capacity of home i
$d_i(t)$	Consumed power of home i at time t
β	Cost of changing output power of generator
L	Maximum power output of generator
y_m	Sunk cost of maintaining generator per time
y_o	Operational cost of generator for output power
p_u	Maximum ramping-up rate
p_d	Maximum ramping-down rate
$g(t)$	Output power of generator at time t
$g_o(t)$	On/off status of generator at time t
$n_o(t)$	Equals 1 if output of generator changes at time t
t_o	Minimum time for generator to change output power
$e_i(t)$	Power from generator to home i at time t
$b_i^u(t)$	Power discharged from battery to home i at time t
$b_i^g(t)$	Power from generator to battery at time t

TABLE I: Definition of notations

$$
\begin{aligned}
\text{Min} \quad & \sum^{M}(y_o g(t) + y_m g_o(t) + \beta n_o(t)[g(t) - g(t-1)]) \\
s.t. \quad & 0 \le b_i(1) \le B_i; \quad \forall i && (a) \\
& 0 \le b_i(t) + b_i^g(t) - b_i^u(t) \le B_i; \quad \forall i, t && (b) \\
& d_i(t) \le b_i^u(t) + e_i(t); \quad \forall i, t && (c) \\
& \sum_{i=1}^{N}(e_i(t) + b_i^g(t)) \le g(t) \le L; \quad \forall t && (d) \\
& g_o(t) - g_o(t - t_o) \le p_u; \quad \forall t && (e) \\
& g_o(t - t_o) - g_o(t) \le p_d; \quad \forall t && (f) \\
& \sum_{i=t+1}^{i=t+t_o} g_o(i) = t_o, \sum_{i=t-t_o}^{i=t-1} g_o(i) = 0; \forall t, n_o(t) = 1 && (g)
\end{aligned}
$$

Constraints (a) and (b) ensures the battery energy level is always not less than zero and not greater than the battery capacity. Constraint (c) means a home consumes less energy than the amount of energy it obtains from generator and battery. Constraint (d) limits the output power of generator. Constraints (e) and (f) mean that the speed of increasing and decreasing generator power. Constraint (g) ensures that the minimum time for generator to change output power is t_o. The object function and constraints are all linear functions, thus the problem is a mixed integer programming problem, which is NP-complete. In reality, it is not possible to obtain optimal solutions in real-time for generation and workload scheduling. Therefore, the central scheduler uses a heuristic approach to solve this problem (detailed in §V-B). Furthermore, because of the unreliable communication, the home controller may not receive the control message from the control center. Therefore, a distributed algorithm is proposed for the local scheduler to schedule workload in each home (detailed in §V-C).

B. Central Scheduler

The central scheduler has the power consumption prediction in next few seconds and next minimum on/off time of generator. Thus, central scheduler can schedule controllable workload in homes to fulfill constraint (c) and schedule generation for next minimum on/off time. Therefore, the optimization problem can be decomposed into two subproblems: i) generation

Algorithm 2 Generation Scheduling Algorithm

1: Calculate the power demand in next minimum on/off time $\triangle E(t+1, t+t_o)$ based on Equation (7);
2: **if** $\triangle E(t+1, t+t_o) > b(t+1) + g(t) \cdot t_o$ **then**
3: Increase generation $g^u(t)$ based on Equation (8);
4: **else if** $\triangle E(t+1, t+t_o) < b(t+1) + g(t) \cdot t_o - B$ **then**
5: Decrease generation $g^d(t)$ based on Equation (9);
6: **else**
7: Maintain the same generation $g(t)$.
8: **end if**

scheduling to fulfill the demand and minimize the operational cost; and ii) workload scheduling in each home to stabilize the aggregated demand.

1) Generation Scheduling: The key idea of the generation scheduling is to turn on the generator when power demand is low and turn off the generator when the power demand is high. Note that it is not able to change the output power of generator due to the minimum on/off time. Thus, we should decide whether to change the output power of generator based on power demand in the next minimum on/off time. At time t, the power demand from time $t+1$ to $t+t_o$ can be calculated as

$$\triangle E(t+1, t+t_o) = \sum_{k=t+1}^{k=t+t_o} \sum_{i=1}^{N} d_i(k) \tag{7}$$

With the given demand, we need first to ensure the power generation and energy storage in battery is higher than the power demand. Assume the power generation at time t is $g(t)$, we have three options: i) increasing generation with $g^u(t)$; ii) decreasing generation with $g^d(t)$; and iii) maintaining the same generation $g(t)$ in next minimum on/off time. The algorithm of decision making for generation is shown in Algorithm 2. If power demand is higher than energy storage in battery and energy generation $g(t)$ in next t_o, we have to increase the power generation to avoid power outage (Lines 1-3). To minimize the operation cost. The amount of generation increase can be calculated as

$$g^u(t) = \frac{\triangle E(t+1, t+t_o) - b(t+1) - g(t) \cdot t_0}{t_o} \tag{8}$$

Otherwise, we can either decrease power generation or maintain the same generation. Note that there is extra cost for changing output power of generator, thus, we only decrease the power generation when the power demand is too low that the battery can not store the extra power generation (Lines 4-5). The amount of generation decrease can be calculated as

$$g^d(t) = \frac{b(t+1) + g(t) \cdot t_o - B - \triangle E(t+1, t+t_o)}{t_o} \tag{9}$$

Otherwise, we maintain the same power generation $g(t)$ (Lines 6-8).

2) Workload Scheduling: The goal of workload scheduling is to avoid power outage and minimize the extra cost for changing power generation. The key idea is to turn off the controllable workload when aggregated power demand is high to avoid power outage and turn on the controllable workload when power demand is low to maintain stable power demand. Note that the demand in each home can be divided by controllable workload $d_i^c(t)$ and uncontrollable workload $d_i^u(t)$:

$$d_i(t) = d_i^u(t) + d_i^c(t) \tag{10}$$

Based on the prediction of short time power consumption in next time slot, our central scheduler decides the scheduling the controllable workload demand in each home. At time t, we can calculate power demand at time $t+1$ as

$$\triangle E(t+1) = \sum_{i=1}^{N} d_i(t+1) \tag{11}$$

If power demand is higher than power generation and energy storage in battery $\triangle E(t+1) > b(t+1) + g(t+1)$, we need to turn off some controllable workload. Otherwise, if $\triangle E(t+1) < b(t+1) + g(t+1)$, we need to turn on some controllable workload. In our simulation, we utilize water heater as controllable workload. Thus, If $\triangle E(t+1) > b(t+1) + g(t+1)$, we turn off some water heaters in homes with less hot water demand until $\triangle E(t+1) = b(t+1) + g(t+1)$. If $\triangle E(t+1) < b(t+1) + g(t+1)$, we turn on some water heaters in homes with most hot water demand until $\triangle E(t+1) = b(t+1) + g(t+1)$.

C. Local Scheduler

However, if the communications between the control center and homes are unreliable, the instructions for each home may be lost or arrive late. Thus, we also provide a distributed control when the control instructions from central controller are not available. The key idea is to schedule controllable workload based on power-voltage model proposed in §IV-A1 because the local voltage in each home can be used to infer the aggregated power demand in the microgrid. However, the problem is that every home may decide to turn on/off controllable workload simultaneously to balance power supply and demand because they can not communicate with each other. Then it may cause an endless loop for each home to turn on/off controllable workload at the same time, which is not helpful to balance power generation and demand.

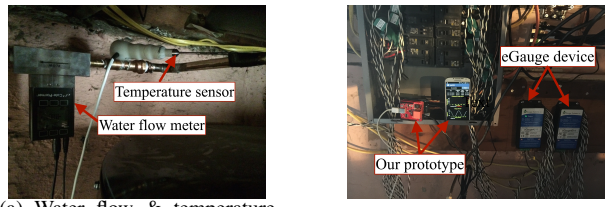
In our design, we take an adaptive feedback control to enable homes to stabilize the power demand cooperatively. When each home detects the power demand change (power generation is relative stable since there exist minimum on/off time for generators), it does not turn on/off controllable workload immediately but with some backoff time. The detailed algorithm is shown in Algorithm 3. At time t, each home calculates the consumption of its controllable workload and estimate the change of power demand $\triangle d$ in the microgrid by utilizing power voltage correlation shown in § IV-A1 (Line 1). When the power demand is stable, each home keeps controllable workload with previous state. When it detects either high or low demand and there is no backoff timer, each home calculates backoff time t_i^b (Lines 2-3). t_i^b can be calculated as

$$t_i^b = \frac{\triangle d}{N \cdot d_i^c(t)} \tag{12}$$

If $t_i^b \neq 0$, home updates the timer (Lines 4-5). Then each home check the timer again, if the timer expires, local controller

Algorithm 3 Local Scheduler Algorithm

1: Calculate $\triangle V_i$, $d_i^c(t)$;
2: **if** $\triangle V_i \neq 0$ & $t_i^b = 0$ **then**
3: Calculate t_i^b based on Equation 12;
4: **else if** $t_i^b \neq 0$ **then**
5: $t_i^b = t_i^b - 1$;
6: **if** $\triangle V_i > 0$ **then**
7: Increase controllable workload;
8: **else if** $\triangle V_i < 0$ **then**
9: Decrease controllable workload.
10: **end if**
11: **end if**



(a) Water flow & temperature monitor

(b) Energy monitoring

Fig. 8: Experiment setup in residential homes

immediately turns on the controllable workload when demand is low and turn off the controllable workload when demand is high (Lines 6-10).

VI. Experimental Evaluations

In this section, we evaluate the performance of our design. We collect empirical data of i) total power consumption and water heater power consumption from 100 homes; ii) hot water usage from three homes; and iii) voltage and frequency data from three homes. Note that we only have control access of three homes (in Binghamton, New York), in which we collected hot water usage, power consumption of water heaters, voltage and frequency data and all the experiments are conducted in these three homes. For the rest 97 homes (in Austin, Texas), we collected the power consumption data for simulations. Because we do not have the hot water usage, voltage and frequency data from these 97 homes, we use water heater power consumption to generate the hot water usage data and apply correlations among power, voltage and frequency obtained from the experiments to generate voltage and frequency data for simulations. The hot water usage and power consumption are measured by water flow sensors and eGauge sensors every second. The experiment setup of one home is shown in Figures 8(a) and 8(b). The power consumption in one year and water flow data in two months are shown in 9 and Figure 10, respectively.

A. Basic Evaluation Results

In this section, we evaluate the effectiveness of our system, which includes three metrics: i) data reconstruction accuracy; ii) total operation cost in the microgrid; and iii) the impact on homeowners' hot water usage. All results are simulated with the seven days' empirical data of hot water usage and power consumption. Because our design goal is to minimize the operation cost of generator, we refer our design as MOC in the latter description. The baseline we compared with is original

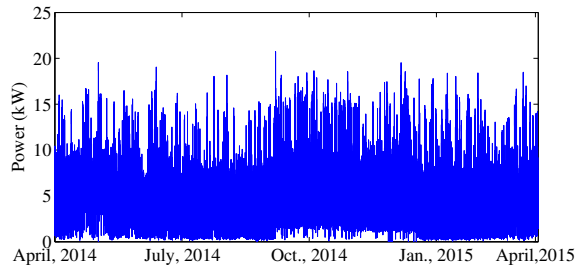


Fig. 9: Power consumption in one year

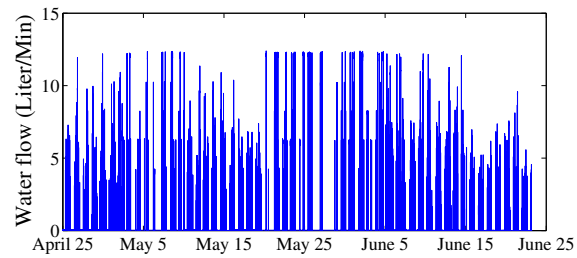


Fig. 10: Water flow in two month

power consumption in individual homes without workload scheduling. In simulations of MOC, we also use baseline in the first day since the correlation model needs to be trained based on historical data. Thus we mainly compare the performance of MOC and baseline for the rest of six days.

1) Reconstruction Accuracy: We run the detection algorithm with one month data and predict the missing data for the next month. The results are shown in Figure 11. The prediction matches well with the ground truth. The maximum error of prediction we observe is $0.498kW$ and the average error of prediction is $0.0289kW$.

We run spatial reconstruction algorithm to recover one home's energy data from other 99 homes. The results are shown in Figure 13. The prediction overall is very close to ground truth. The maximum error of prediction and average error of prediction are $3.836kW$ and $0.131kW$, respectively.

2) Generation and Consumption: To clearly illustrate the difference between baseline and our design, we only show the power consumption of baseline and our design for one day in Figure 12. For power consumption without water heater, the peak demand is mainly from 8am to 12am and from 6pm to 8pm. In the mean time, hot water usage is also during the

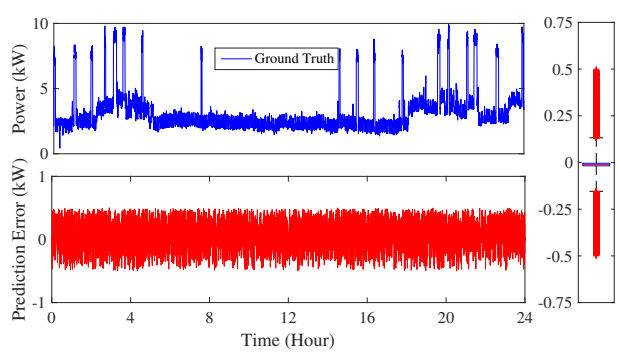


Fig. 11: Prediction accuracy with temporal correlation model, the right figure shows the average, 25 percentile and 75 percentile of the prediction error

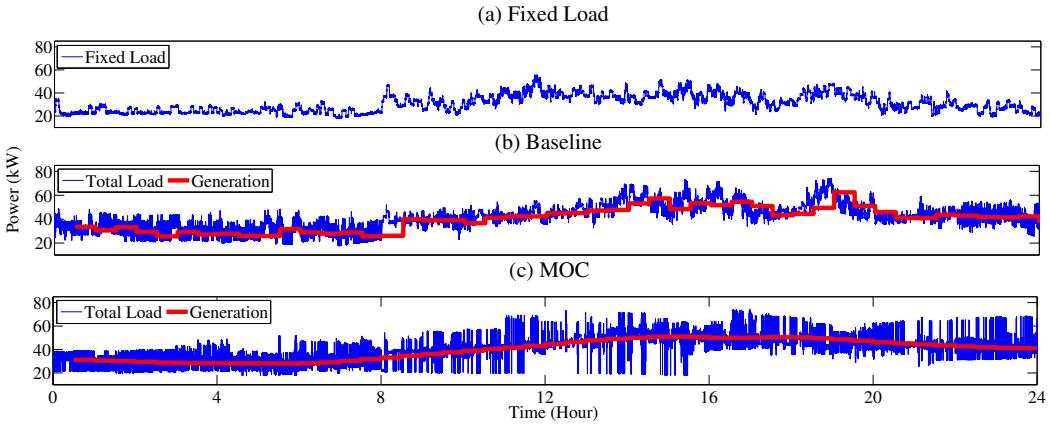


Fig. 12: Power consumption and generation for one day

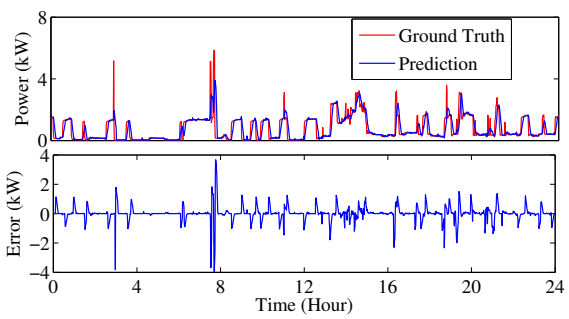


Fig. 13: Prediction accuracy with spatial correlation model

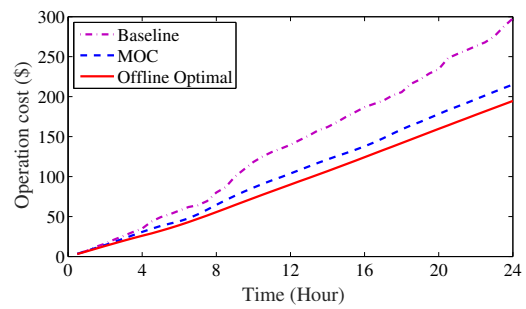


Fig. 14: Total operation cost for generators

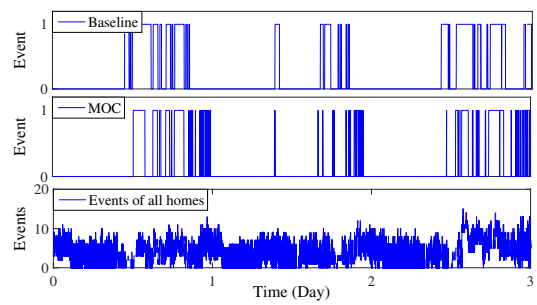


Fig. 15: Water heater events over three days

similar time. For baseline, when it detects hot water usage, it turns on water heater immediately. Furthermore, because different homes are highly possible to use hot water in similar time of evening, the peak demand rises from $50.31kW$ to $75.45kW$ (at around 7pm and 11pm). The power of the water heater in our simulation is $5.29kW$, thus at least 5 homes turn on water heaters at the same time. For the generation, because it does not predict the future power consumption to smooth the generation, the hourly generation changes very quickly, which introduces more operation cost. For MOC, because it predicts short-term and long-term power consumption in future, the hourly generation is stable.

3) Total Cost: The total operation cost for generators for a typical day are shown in Figure 14. In the beginning, because the power consumption is low, the cost for different approaches is similar. However, from 8am, power demand increases quickly because most of the people wake up. Thus, the operation cost increases quickly in baseline and the operation cost for MOC and offline optimal still increases linearly. For six days' simulations, the average daily operation cost in baseline is \$292.5 while daily operation cost for MOC is \$224.6, which is 23% lower.

4) Water Heater Scheduling: To better understand how homes schedule water heater event, we show detailed water heater energy consumption events of 15 homes in Figure 15. To show the detailed difference between baseline and MOC, we show the water heater events for 3 days. The upper figure is the water heater events of the baseline in one home. For the baseline, most of the water heater energy consumption events last for longer periods. This is because the water heaters are turned on right after hot water usage. After people take a shower or bath for ten minutes, the water heater will be turned on for around 1 hour to reach the high temperature threshold. The middle figure shows the water heater events of MOC in one home. Compared to baseline, the water heater events are more sparsely distributed to reduce the overlap of events from different homes. The bottom figure shows the total events of MOC in all 15 homes.

5) Hot Water Temperature: Though MOC allows homes to turn on water heater earlier or later, MOC can also fulfill users' hot water usage efficiently. In Figure 17, we show the distribution of difference between the targeted temperature and the hot water temperature when there exists hot water usage events. The targeted temperature of hot water in our experiment is set as $50°C$. For the baseline, it turns on water heater immediately after hot water usage, then the temperature of hot water is always a little lower than the targeted hot water

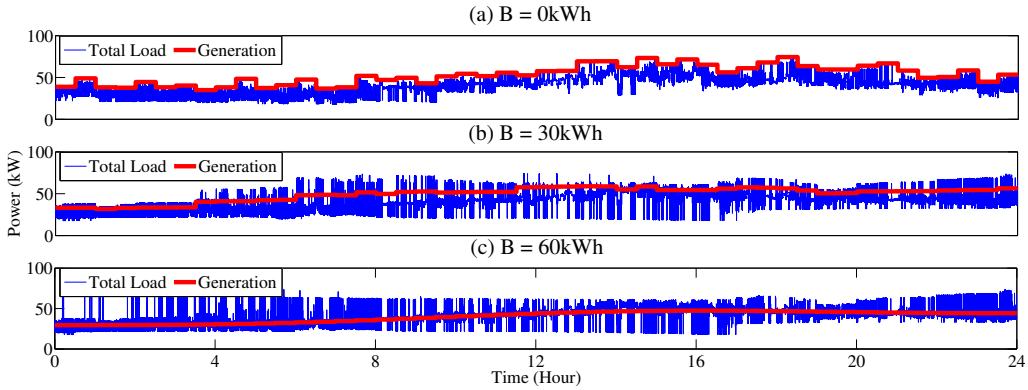


Fig. 16: Power consumption and generation for different battery capacities

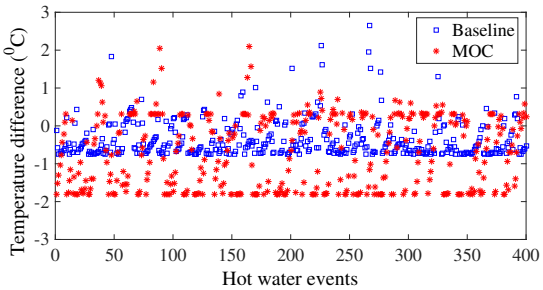


Fig. 17: Temperature distribution of hot water events

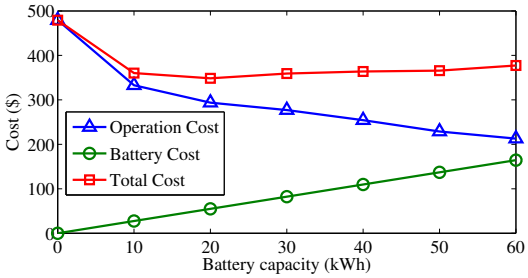


Fig. 18: Operation, battery and total cost for different battery capacities

usage (mainly $1°C$ lower than the targeted temperature in Figure 17). For MOC, when a home predicts future hot water usage, it can turn on water heater earlier to better fulfill the hot water usage, thus the hot water temperature can be higher than the targeted temperature for some time. In Figure 17, the hot water temperature in the water tank is at most $2°C$ lower than the targeted temperature. Thus the impact of our design on people's hot water usage is very low.

B. Advanced Evaluations

Because our system is designed for severe environments, such as islands, it is crucial to investigate the system's sensitivity under diverse settings.

1) Impact of Battery Capacity: Because battery is expensive and has limited life-time, it is important to investigate the system benefit with different capacities of batteries. We show the power consumption and generation with three different battery capacities in Figure 16. The top figure shows the scenario with no battery in the system. Because there is no battery, the generators needs to generate the maximum power

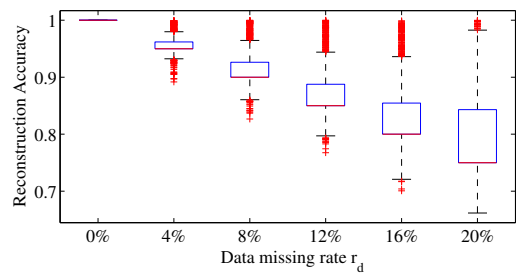


Fig. 19: Reconstruction accuracy with different r_d

in its working period to avoid power outage, which introduces high energy waste. The middle figure shows the results with $30kWh$ battery capacity. With the battery, the difference between generation and consumption can be offset, thus the overall generation is reduced. For the bottom figure, with very high battery capacity, the generator can only generate the average power consumption in the next hour, thus the overall generation is minimized. The operation, battery and total cost for different battery capacities are shown in Figure 18. With higher battery capacity, the operation cost decreases especially from no battery to $10kWh$ battery. However, the decrease slows down with higher battery capacity. For the total cost, we find that microgrid with $20kWh$ battery performs best since it balances cost between generation and batteries.

2) Impact of Data Missing Rate: With different environments, the data missing rate can be quite different. Thus, it is important to study the reconstruction accuracy of our data management design under different scenarios. In these sets of simulations, we simulated the accurate data to generate missing data with different missing rate from 4% to 20%. The results of reconstruction accuracy are shown in Figure 19. With the increase of data missing rate, reconstruction accuracy decreases slowly. Even with 20% data missing rate, the average of reconstruction accuracy is above 80%. Thus, our design is robust in situations with high data missing rate.

3) Impact of Instruction Missing Rate: In the mean time, the instruction missing rate can also be quite different under different environments. Thus, we study the performance of centralized control (instruction missing rate r_i is 0%) and distributed design (instruction missing rate r_i is 100%). We simulated the different instruction missing rate for two extreme

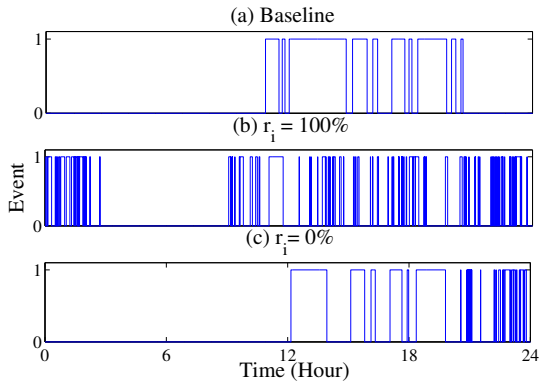


Fig. 20: Water heater events with different r_i

cases (shown in Figure 19). The water heater is turned on and off more frequently when r_i is 100%. That is because each home lacks the knowledge of behaviors of other homes, thus they may collide to turn on/off water heater and then immediately turn off/on water heater. The frequent on/off operations will decrease the lifetime of water heater. However, total cost for different instruction missing rate is similar, which is not included due to the limited space.

VII. Related Work

Our work is related to the following areas of previous work:

Energy Management. Different techniques are proposed for energy management in either demand side or generation [11], [12]. In [13], a decentralized optimal load control mechanism is proposed to provide contingency reserve in the presence of sudden demand-supply mismatch. In [14], a model predictive control algorithm is proposed to co-schedule HVAC control, EV scheduling and battery usage to reduce the building energy consumption. In [15], stochastic and robust optimization are applied for real-time price based demand response management. In [16], an optimal solution is provided to trade off between quantity and quality of variable renewable energy source in smart grid. Different from existing work, our paper presents a holistic approach of real-time scheduling for both demand in individual homes and generation of the local generators in microgrids.

Missing Data Management. There have been various works in the research community to investigate how to manage missing data in cyber-physical systems [17]. Traditionally the approach to obtaining missing values for linear time series has involved the use of curve fitting [18]. Autoregressive integrated moving average (ARIMA) model is fitted to time series data to predict future points in the time series data [7]. Maximum likelihood based approach is applied to estimate the missing data [8]. Different from existing work, we investigate the correlation between different energy data in microgrids and utilize the correlation models to cope with unreliable sensors and wireless communication.

VIII. Summary

The biggest challenge of maintaining a self-sustainable microgrid is to balance the power demand and generation in real-time with dynamically changing power demand. Furthermore, the unreliable data collection and communication between homes and the control center in a microgrid makes the real-time control even harder. To address these issues, we propose a novel data management technique to process the collected data, reconstruct the missing data caused by sensing error or unreliable communication, and predict the future demand for real-time control with missing data in extreme situations. The control center then decides the scheduling of the workload of appliances in each home and the operations of the local generator based on the collected and predicted data. Through extensive experiments and simulations, we show that our design can recover the missing data with 99% accuracy and our distributed control can balance power demand and generation and reduce operation cost by 23%.

IX. Acknowledgment

This project is supported by NSF grant CNS-1503590 and UMBC COEIT Strategic Plan Implementation Grant.

References

[1] C. A. Macana, N. Quijano, and E. Mojica-Nava, "A survey on cyber physical energy systems and their applications on smart grids," in *IEEE PES Conference on Innovative Smart Grid Technologies*, 2011.

[2] W. Su and J. Wang, "Energy management systems in microgrid operations," *The Electricity Journal*, vol. 25, no. 8, pp. 45–60, 2012.

[3] S. Aman, Y. Simmhan, and V. K. Prasanna, "Holistic measures for evaluating prediction models in smart grids," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 2, pp. 475–488, 2015.

[4] S. Haben, J. Ward, D. V. Greetham, C. Singleton, and P. Grindrod, "A new error measure for forecasts of household-level, high resolution electrical energy consumption," *International Journal of Forecasting*, vol. 30, no. 2, pp. 246 – 256, 2014.

[5] R. H. Lasseter and P. Paigi, "Microgrid: a conceptual solution," in *IEEE Power Electronics Specialists Conference*, 2004.

[6] "Microgrid deployment tracker 4Q15," *Navigant Research*, 2015.

[7] G. P. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[8] R. J. Little and D. B. Rubin, *Statistical analysis with missing data*. John Wiley & Sons, 2014.

[9] T. Short, *Electric Power Distribution Handbook*. Taylor & Francis, 2003.

[10] S. A. Kazarlis, A. Bakirtzis, and V. Petridis, "A genetic algorithm solution to the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 11, no. 1, pp. 83–92, 1996.

[11] T. Wei, Q. Zhu, and N. Yu, "Proactive demand participation of smart buildings in smart grid," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1392–1406, 2016.

[12] L. Lu, J. Tu, C.-K. Chau, M. Chen, and X. Lin, "Online energy generation scheduling for microgrids with intermittent energy sources and co-generation," in *Proceedings of the ACM SIGMETRICS*, 2013.

[13] C. Zhao, U. Topcu, and S. Low, "Optimal load control via frequency measurement and neighborhood area communication," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3576–3587, 2013.

[14] T. Wei, Q. Zhu, and M. Maasoumy, "Co-scheduling of hvac control, ev charging and battery usage for building energy efficiency," in *Proceedings of the 2014 IEEE/ACM International Conference on Computer-Aided Design*, pp. 191–196, IEEE Press, 2014.

[15] Z. Chen, L. Wu, and Y. Fu, "Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1822–1831, 2012.

[16] F. Kong, C. Dong, X. Liu, and H. Zeng, "Quantity versus quality: Optimal harvesting wind power for the smart grid," *Proceedings of the IEEE*, vol. 102, no. 11, pp. 1762–1776, 2014.

[17] L. Kong, M. Xia, X.-Y. Liu, M.-Y. Wu, and X. Liu, "Data loss and reconstruction in sensor networks," in *INFOCOM, 2013 Proceedings IEEE*, pp. 1654–1662, IEEE, 2013.

[18] C. Chatfield, *The analysis of time series: an introduction*. CRC press, 2016.