Smart Home in Smart Microgrid: A Cost-Effective Energy Ecosystem with Intelligent Hierarchical Agents

Bingnan Jiang, Student Member, IEEE, and Yunsu Fei, Member, IEEE

Abstract—Smart grid is advancing power grids significantly, with higher power generation efficiency, lower energy consumption cost, and better user experience. Microgrid utilizes distributed renewable energy generation to reduce the burden on utility grids. This paper proposes an energy ecosystem; a cost-effective smart microgrid based on intelligent hierarchical agents with dynamic demand response (DR) and distributed energy resource (DER) management. With a dynamic update mechanism, DR automatically adapts to users’ preference and varying external information. The DER management coordinates operations of micro combined heat and power systems (µCHPs), and vanadium redox battery (VRB) according to DR decisions. A two-level shared cost-led µCHPs management strategy is proposed to reduce energy consumption cost further. VRB discharging is managed to be environment-adaptive. Simulations and numerical results show the proposed system is very effective in reducing the energy consumption cost while satisfying user’s preference.

Index Terms—Demand response (DR), distributed energy resources (DER), microgrid, particle swarm optimization (PSO), Q-learning, smart grid.

I. INTRODUCTION

With more electricity-consuming products coming into daily lives, such as electrical vehicles (EVs) and advanced heating, ventilation, and air conditioning systems, load demand increases dramatically and imposes significant burdens on the existing power grid. Smart grid, integrated with distributed renewable energy generation, advanced metering infrastructure, and information technologies, can cope with the impending global energy crisis and environment deterioration. To achieve high energy efficiency in smart grid, load can be shaven by demand response (DR) and distributed energy resources (DER) have to be well managed. Residential DR can be defined as reactions of users to the time-varying energy price offered by utility companies [1], where schedulable load is shifted to off-peak hours to reduce the energy consumption cost. On the other hand, DERs, including distributed generation (DG) and energy storage system, can be controlled optimally to supply power to the load directly at the distribution level. Renewable DGs will provide cheaper and cleaner energy supplement. However, the high initial investment expense, like on wind turbine installation and high capacity batteries, would prevent an ordinary household from using renewable DGs [2]. One solution is to build a microgrid for many households in a neighborhood. In the microgrid, wind turbines and energy storage systems are shared by the whole community to reduce the investment cost for each user. Meanwhile, each household also shares its residential DG with others in the community.

A. Related Work

Recent related work includes design and analysis of DR and DG management systems for smart home and smart grid. Important elements affecting DR performance have been analyzed. Challenges relating to load forecast and DR are discussed in [3]. Reference [4] shows that the unpredictable human factors can have significant influences on DR system’s performance. Existing DR systems are designed with deterministic or stochastic algorithms in centralized or decentralized ways. A stochastic dynamic programming method for electricity usage is proposed in [5], which assumes that system states’ transition probabilities, e.g., utility power price and outdoor temperature, are known in advance. A residential DR algorithm using Q-learning is presented in [6] that takes stochastic load demand, electricity price, and user’s convenience into consideration. However, Q-learning can hardly be applied to complex tasks and price models because of the low convergence speed in high-dimensional state and action space in real scenarios. Reference [7] proposes a decentralized load management control based on real-time price. Overall, most of existing methods do not consider the interaction between DR system and the user, i.e., improving decisions by observing user’s manual adjustment. Therefore, these DR systems cannot accommodate users’ preference changes and may give unsatisfactory decisions.

For DG management in microgrids, different optimization methods have been implemented. Some work is based on the static load and weather forecast [8], [9], which neglects their dynamic and stochastic characteristics in real situation. Reference [10] proposes a scheduling method for hybrid supplies considering stochastic elements. The obtained decisions are optimal for the average performance of all possible situations, but cannot adapt to an actual instance. In [11], a
robust energy management for microgrid with intermittent renewable energy resources is proposed and the worst-case transaction cost is included in the cost function. As a new type of clean DG with high energy efficiency and low emission, micro combined heat and power systems (μCHPs) have recently attracted much attention and become a promising DG in residential homes. μCHP control strategies can be categorized as heat-led, electricity-led, and cost-led [12]–[14]. In heat-led or electricity-led strategy, the μCHP generates energy whenever there is an electricity or heat demand, respectively. The cost-led strategies proposed in [13] and [14] utilize the characteristic of co-generation to achieve the minimum overall cost. Under this strategy, extra electricity will be exported to the utility grid and additional heat will be consumed by thermal energy storage, like a hot water tank. The existing strategy only focuses on the optimal operation of a single μCHP. The issue of coordination among multiple μCHPs for cost reduction in smart microgrid has not been addressed and will be explored in our proposed ecosystem.

B. Paper Overview

This paper proposes a novel cost-effective energy ecosystem in smart microgrid. The ecosystem is dedicated to a community where there are smart homes with DERs. It consists of five components: utility grid, DGs, energy storage, appliances, and users. The objective of the energy ecosystem is to achieve high-quality energy service and low consumption cost for all households in it. The energy ecosystem effectively integrates and manages diverse energy resources, including centralized renewable energy, residential DG, and energy storage (battery). Besides, the proposed energy ecosystem has several application significances. First, since the energy ecosystem is built for a group of users, the shared infrastructure will benefit the whole community and provide affordable renewable energy that standalone residential systems cannot offer. Second, DGs and batteries in the energy ecosystem supply power for load demand in the whole community. Their utilization rates are largely raised and a shorter time to return on investment can be expected. Third, by coordinating DR and DER management with optimization, the overall energy consumption cost in the community can be reduced. The cost effectiveness of the energy ecosystem will be convincing residential communities to consider and invest in such shared infrastructure, contributing to greener and smarter grids.

However, it is challenging to integrate DR and DER management in either centralized or distributed optimization. On account of different roles and decisions in the energy ecosystem, the number of control variables is large. For centralized optimization, the computational complexity greatly increases when the system scale becomes large. In addition, the failure of a centralized control will result in malfunction of the entire system. Although distributed optimization provides a scalable solution approach, the dynamic interactions between different roles and decisions have to be simplified to suit distributed optimizations. Another difficulty is to integrate stochastic elements, such as stochastic wind power and load demand, into the optimization system. These elements are affected by various factors, such as weather, wind turbine physical characteristics, and users’ demand change. Their accurate mathematical models are thus hard to build.

This paper adopts a hierarchical optimization method to decouple the complex problem of coordinating DR and DER management. In the hierarchical optimization, information and solutions from a lower level stage will be input to a higher level. The relationships among components in the energy ecosystem are hierarchical in nature. Load demands are first requested and scheduled by users. DGs and batteries then respond to load demands by supplying power. With the hierarchical optimization design, the original unsurmountable complex problem is decomposed into several tractable subproblems which can be solved efficiently at different stages. We also use a model-free reinforcement learning method to incorporate stochastic elements in optimization. Overall, the energy ecosystem designed in this paper has four novelties.

1) Hierarchical optimization based on both centralized and distributed agents (DAs) will reduce the computational complexity and get satisfying performance.
2) Interaction between users and DR agents is enhanced by adopting users’ feedback. DR agents make decisions adaptable to user’s preference change.
3) Instead of optimizing a single μCHP operation separately for its belonging house, all μCHPs cooperate to generate power for the whole community in a shared cost-led mode based on two-level optimization.
4) Environment-adaptive battery management based on reinforcement learning considers stochastic elements in the microgrid and obtains an optimal discharging policy.

The rest of this paper is organized as follows. Section II describes the overall system scheme. Section III illustrates the system design and detailed algorithms. The simulation setup and results are given in Section IV. Section V concludes the paper.

II. System Scheme

The studied microgrid works in the grid-connected mode. As shown in Fig. 1, there are three types of flows in the microgrid: 1) electric power flow; 2) thermal power flow; and 3) information flow. The electric power demand is supplied by the utility grid, centralized wind turbines and batteries, and distributed μCHPs. The thermal power demand in each individual house is supplied by μCHP or electric heat pump. Wind turbines are shared by the whole community. Each household has its subscription rate indicating the amount of wind and battery power that can be used. μCHPs generate power according to the load demand in the microgrid. In the grid-connected microgrid, power generation and consumption are balanced. When load demand is higher than generation, extra power will be supplied by utility grid. If extra electricity is generated, it will be sold back to the utility grid. Current policies usually allow wind energy to be sold with retail rates while μCHP energy with lower avoided cost rates [15]. For general consideration, some houses in the community are installed with μCHPs, whereas others are not. For the latter, thermal energy can only be provided by electric heat pump. Extra thermal energy generated by μCHP can be either stored in the hot water tank.
or dumped. The temperature of water tank should also be maintained within a range along time. Batteries belong to the community. They discharge in peak hours to reduce cost. They also work as standby power supplies for emergent blackouts.

This paper selects the vanadium redox battery (VRB) rather than the conventional deep cycle lead-acid batteries because VRB has much longer life cycle, higher efficiency, and lower discharging cost [16]–[18]. The information flow contains utility price, wind power prediction, and insufficient capacity power price, wind power prediction, users’ input, system status, control signals from agents, etc.

The three-level hierarchical optimization is depicted in Fig. 2. Load demand and power supply in the system are decoupled in this hierarchy. The time resolutions at different optimization levels are set the same for easy synchronization. The lowest level is DR executed by the DA of each house for energy consumption cost reduction with users’ satisfaction taken into consideration. Distributed DRs can significantly reduce computational complexity without losing much optimality. In each house, DA collects relevant external information, e.g., day-ahead time-varying utility price and wind power prediction, and realizes dynamic DR at every decision time, i.e., every hour. DR results include the starting time of each schedulable task. The optimization formulation in DR considers power supply from utility grid and wind turbine (subscribed wind power of each household), but not μCHPs or VRB. This decoupling is reasonable since the cost and ability of μCHP generation do not vary along the time. The VRB discharging capability does not change much unless it is depleted. Thus, DR optimization results are not affected much when power supply from μCHPs and VRB are not considered.

At the second level, a centralized agent gets load demand of each house from DR decisions and optimizes μCHPs’ generation. At one time, some houses may have high electric load demand that cannot be supplied merely by their subscribed wind power and μCHP self generation, while others with low demand do not need μCHP generation. Thus, this paper considers the potential improvement of energy generation efficiency by coordinating distributed μCHPs during optimization and proposes the shared cost-led μCHP management strategy. In this strategy, instead of generating power for its own house, generation of all μCHPs is coordinated to minimize the cost of the whole community. The optimization agent first calculates the remaining load demand in the community after deducting the predicted wind power supply. Since μCHPs generate electric and thermal power simultaneously and have higher power output than battery, generation of μCHPs is first optimized at this level to supply the remaining load. The optimization considers DR decisions from the first level as well as utility power price and wind power prediction. VRB discharging is not considered in optimization at this level.

The last level optimization is for VRB charging and discharging. Different from μCHPs power generation, VRB can respond fast to load changes with charging/discharging. So its discharging is optimized to compensate stochastic load and insufficient μCHP power generation in the microgrid at the final stage. VRB is charged by the extra wind power if it is not full and the power selling price is low. Obtaining an optimal discharging policy is challenging, determined by the stochastic environment, i.e., load demand, utility price, and future available wind power. Building an environment model for the microgrid is complex and impractical. Therefore, we propose a reinforcement learning-based VRB discharging strategy by evaluating decisions’ immediate and subsequent effects on the ecosystem. The centralized agent gets load demand and μCHPs and wind generation information, and takes into account their possible stochastic changes during policy making. With reinforcement learning, the discharging policy can be obtained from the interaction between the agent and the environment without establishing its detailed models.

With the three-level hierarchical optimization, the energy consumption cost for the whole community is minimized. To guarantee fairness for all households, their utility bills need to be balanced according to their energy consumption and generation at different time.

III. SYSTEM DESIGN AND MANAGEMENT

In this section, models of DR, DG, and energy storage are designed first. Management at each level is then formulated as an optimization problem. In the end, the problem solving algorithms are given.
A. Distributed DR

We first present DR models that include different types of tasks and their historical execution records. We then formulate DR as an optimization problem.

1) DR Model: DR gives load scheduling decisions which are updated dynamically at different time according to the change of load demand and wind power prediction. Electric load demands are either schedulable or fixed energy consuming tasks. A schedulable task can be assigned to operate at different time with different user’s satisfaction. For example, the working of laundry machine and EV charging are schedulable tasks. It is also assumed that a scheduled task cannot be interrupted once it starts. On the contrary, a fixed task is time-sensitive and must be executed at designated time, such as the operation of refrigerator, watching TV programs at specific time, and turning on heating and air conditioning by house residents. DR is designed for schedulable tasks and will give the optimal scheduling solution at each decision time. A DR decision time point can be in three situations: at the beginning of each hour, when a user adds new tasks, or when a user intends to adjust the scheduling decisions. Each task is associated with a preference function, which is designed to indicate a user’s varying satisfaction dependent on the task’s starting time. Preference functions are updated dynamically according to users’ preference change due to summer/winter time switch, weather change, holiday seasons, short term change of living habit caused by irregular working agenda, etc.

At each DR decision time \( k \), the preference function \( F^k_{pr,i}(t) \) for task \( i \) is a function of task starting time \( t \). The preference function is based on \( \hat{f}^k_i(t) \), which is the estimated probability density function (PDF) of task \( i \)’s starting time \( t \). As a nonparametric density estimation method, the kernel density estimation (KDE) method has broad applications in the univariate case [19] and is suitable for estimating \( \hat{f}^k_i(t) \). Initially, \( \hat{f}^0_i(t) \) is estimated from the historical task execution record as

\[
\hat{f}^0_i(t) = \frac{1}{Nh} \sum_{n=1}^{N} K \left( \frac{t - T_n}{h} \right),
\]

where \( K \) is a symmetric PDF, e.g., Gaussian density function, called kernel function. \( T_n \) is the \( n \)th sample in the data set. \( N \) is the total number of samples in the data set. \( h \) is the smoothing parameter called bandwidth, which determines the trade-off between estimation bias and variance. Since the performances of different kernel functions are very similar, Gaussian kernel is selected in this paper with its convenient mathematical properties. \( h \) is selected to minimize the mean integrated square error (MISE) defined as

\[
MISE(\hat{f}) = E \int \left[ \hat{f}(t) - f(t) \right]^2 dt.
\]

For Gaussian kernel, the optimal bandwidth is \( h^* = 1.06\sigma N^{-1/5} \), where \( \sigma \) is the sample standard deviation. \( \hat{f}^k_i(t) \) at time point \( k \) is updated by processing new samples, i.e., tasks’ actual starting time in either a regular way or with weighted update. The main idea is to weight user’s adjustment and learn user’s preference change faster. If a task is scheduled by DR and accepted by the user, the sample is processed with an ordinary update. If the scheduling is not accepted and rescheduled by the user, it is updated with a weight \( M \). The exact value of \( M \) is determined by a tunable parameter \( \rho (\rho > 0) \) as \( M = \max(2, [\rho N]) \), where \( N \) is the size of data set. The data set has its capacity. When the data set is full and new samples come, the oldest ones will be replaced. \( F^k_{pr,i}(t) \) is set to be equal to the normalized pdf \( \hat{f}^k_i(t) = \hat{f}^k_i(t)/\max(\hat{f}^k_i(t)) \).

2) Optimization Problem Formation: The length of a DR cycle is set at 24 h for scheduling tasks. Because users desire more satisfaction with lower cost, the optimization at each decision time is to minimize the unit cost, the energy consumption cost per satisfaction, for the current and remaining time in the cycle. Monetary cost is calculated as the product of electricity price (utility price and wind price), load power demand, and load duration. After discretization with time resolution \( \tau \), the energy consumption cost at DR decision time \( k \) for all tasks \( (i = 1, ..., I) \) is formed as

\[
C(k_n) = \sum_{k=k_n}^{N_D} \left[ R_W(k)E_{LW}(k) + R_G(k)E_{LG}(k) \right]
\]

where \( E_{LW}(k) \) and \( E_{LG}(k) \) are wind and utility grid energy used in time slot \( k \), respectively. Considering both the schedulable tasks and fixed tasks, \( E_{LW}(k) \) is calculated according to user’s wind power subscription rate \( \alpha_s \) and its price is \( R_W(k) \). The extra energy demand \( E_{LG}(k) \) will be supplied by the utility grid with price \( R_G(k) \). Power consumption from wind \( P_{LW}(k) \) and grid \( P_{LG}(k) \) at time \( k \) are formulated as

\[
P_{LW}(k) = \min \left\{ \sum_{i \in I_{LW}} P_i(k) + P_F(k), \alpha_s P_W(k) \right\}
\]

\[
P_{LG}(k) = \max \left\{ 0, \sum_{i \in I_{LG}} P_i(k) + P_F(k) - \alpha_s P_W(k) \right\}
\]

where \( P_i(k) \) is the power consumption of task \( i \) at time \( k \). Each task \( i \) requires \( T_{R,i} \) time slots for operation. \( P_i(k) \) equals to the rated power \( P_{R,i} \) if time \( k \) is within the task operating time. Otherwise, it is 0. The total satisfaction one user can get at decision time \( k_n \) by scheduling tasks is

\[
U(k_n) = \sum_{i \in I_n} u_i(k_n) = \sum_{i \in I_{LW}} u_i s_i(k_n) F_{pr,i}^k(k_{sh,i}(k_n)).
\]

Other variables and parameters include the following.

1) Control Variables at Decision Time \( k_n \): \( s_i(k_n) \) is a binary value indicating the scheduling decision for task \( i \) at \( k_n \). “1” means task scheduling and “0” means not. It determines the set of scheduled tasks \( I_{LW} \) after decisions at time \( k_n \) are made. \( k_{sh,i}(k_n) \) is the scheduled starting time of task \( i \).

2) Power Parameters: \( P_W(k) \) and \( P_F(k) \) are predicted total wind power supply (kW) and total load of fixed tasks (kW), respectively, at time \( k \).

3) Time Parameters: \( N_D \) is the number of time slots in one DR cycle.
4) Other Parameters: $R_g(k)$ and $R_w(k)$ are utility electricity price (USD/kWh) and wind power price (USD/kWh), respectively, at time $k$. $u_i$ is the weight coefficient reflecting the importance of task $i$. $L_{kn}$ is the set of tasks have not started till the beginning of DR decision at $k_n$.

DR optimization constraints include the following. First, all scheduled tasks should be completed before the end of DR cycle. No tasks are allowed to be postponed to the next day. Second, one task may depend on the completion of another one, for example, a dryer can start to work only after the laundry machine finishes washing. Third, execution time of each task should be scheduled between current decision time and the end of DR cycle. At last, each house has its maximum allowed power restricted by the circuit breaker.

Tasks are allowed to be dropped when some constraints cannot be satisfied, such as the situation when a user has tasks with high rated power which causes the total power exceeds the maximum allowed one at any time. Task dropping penalty $P_T$ is thus introduced. The optimization is to minimize the unit cost with task drop penalty considered

$$
\min \frac{C(k_n)}{U(k_n)} + \left[ \left| L_{kn} \right| - \sum_{i \in L_{kn}} s_i(k_n) \right] P_T 
$$

s.t. DR constraints.

B. Shared Cost-Led $\mu$CHPs Management

Operations of distributed $\mu$CHPs are optimized according to the load scheduling results from the DR system. The $\mu$CHP model described in [14] and [10] is applied in this paper. It has three states: 1) idle; 2) start-up; and 3) generation. When $\mu$CHP is idle, there are no fuel consumption and power generation. In the start-up period, fuel will be consumed without power generation. After start-up, the system generates both electric and thermal power based on the adjustable fuel consumption.

1) Shared Cost-Led $\mu$CHPs Management Strategy: In view of the dynamic characteristic of the microgrid discussed above, including DR and wind power forecast, it is important to ensure fast response of $\mu$CHPs to the load and supply changes. The state transition between idle and generation takes time and consumes energy. Therefore, it is important to determine the optimal “ON/OFF” state of $\mu$CHPs in advance according to available information. Since the amount of heat dump is limited, thermal power generation should be constrained to keep desired water tank temperature.

The shared cost-led $\mu$CHPs management is formed as a two-level optimization problem. The main idea is that the coarse-grained optimization has long-term perspectives and will guide the fine-grained optimization in terms of $\mu$CHP state and thermal power generation along the time. The fine-grained and coarse-grained optimization have time resolutions $\tau$ (slot, same as DR resolution) and $T_C$ (period), respectively. $T_C$ is an integral multiple of $\tau$. The $\mu$CHP start-up time $T_S$ is also set to an integral multiple of $T_C$. The fine-grained optimization is to determine the detailed optimal $\mu$CHP fuel input stream and electric heat pump generation for each slot $\tau$ in the current period $T_C$ to minimize the energy consumption cost of the whole community. The coarse-grained one is to minimize the sum of the approximate cost of the community in next $N_F$ coarse-grained periods by determining the optimal $\mu$CHP states, the average fuel input volume, and the average electric heat pump generation in each period. The solved optimal $\mu$CHPs’ states will be used for $\mu$CHPs’ state transitions. The total thermal generation in each period will serve as a constraint for fine-grained optimization in that period.

2) Fine-Grained Optimization for Current Period: There are $N_C = T_C/\tau$ time slots in one period. The total load demand in the community $P_{SL}(n)$ in time slot $n$ is calculated according to DR decisions as $P_{SL}(n) = P_f(n) + \sum_{i \in R_u} P_i(n)$. The electric power consumption cost $C_{CE,k}(n)$ and fuel consumption cost $C_{CF,k}(n)$ of the whole community in time slot $n$ of current period $k$ are formulated as

$$
C_{CE,k}(n) = R_g(n)E_{EG}(n) + R_wE_{EW}(n)
$$

$$
C_{CF,k}(n) = R_f \tau \sum_{m \in MG} g_{F,m,k}(n)
$$

where $\tau \sum_{m \in MG} g_{F,m,k}(n)$ in (8) is the total $\mu$CHP fuel input volume within time $\tau$. The fuel consumption $C_{CF,k}$ is obtained as the product of fuel volume and fuel price. Variables and parameters in (7) and (8) include the following.

1) Control Variables: $g_{F,m,k}(n)$ is the fuel input stream of $\mu$CHP $m$ at time $n$ (Nth^3/s).

2) Energy and Power Terms: $E_{EW}(n)$ and $E_{EG}(n)$ are calculated energy consumption (kWh) from wind power supply and utility grid, respectively, at time $n$ according to the scheduled load demand $P_{SL}(n)$ and $\mu$CHP electric power generation $P_{CHPE}(n)$ in the microgrid. $P_{CHPE}(n) = \eta_{CHPE} \sum_{m \in MG} g_{F,m,k}(n)$ where $\eta_{CHPE}$ is the electric efficiency of $\mu$CHPs and $q_F$ is the heating value of fuel (kJ/Nth^3).

3) Other Parameters: $R_f$ is the fuel gas price (USD/Nth^3). $MG$ denotes set of houses with $\mu$CHP states “generation.”

Extra generated power is sold to the utility grid with income $B_{CM,k}$. The optimization problem at the current decision time $k$ is to minimize the total cost in the period as

$$
\min \sum_{n=1}^{N_C} C_{CE,k}(n) + C_{CF,k}(n) - B_{CM,k}(n)
$$

subject to the following constraints. First, the thermal generation of each house should be equal to the value solved from coarse-grained optimization. Second, both fuel input and electric heat pump generation have their allowable ranges. Finally, with electric heat pump added, the total electric power consumption in a house cannot exceed the maximum value.

3) Coarse-Grained Optimization for Future Periods: The coarse-grained optimization has to consider $N_F$ periods. The cost function is the sum of approximated cost of the whole community in these $N_F$ periods. Control variables are each $\mu$CHPs state (binary values, 0 for idle state and 1 for generation state), its average fuel input volume (for the units at generation state), and electric heat pump power consumption for heat generation. For each time period, its approximated cost has the same formation as the fine-grained one, except that it has a larger time resolution $T_C$. This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.
In addition to the constraint of thermal generation and maximum load, other constraints include: first, in each periods, the temperature of water tanks should be maintained within in a range; second, at a designated time, the temperature of water tank should reach the set point as the desired average temperature; third, the maximum allowable heat dump is constrained.

C. VRB Discharging Management With Q-Learning

In this third-level optimization, VRB will be optimized for discharging to supply the remaining load after consuming wind and μCHP power at the first two level. This happens when the load demand is high but the wind power is low or the μCHP generates insufficient power for stochastic load demand. The efficiency of VRB is determined by its charging/discharging current and the state of charge (SOC) with nonlinear characteristics. To keep high battery efficiency, the charging/discharging current and SOC are constrained within certain ranges, in which the efficiency can be approximated to a constant value [18], [20].

The stochastic load demand and wind power can be modeled as Markov chains. VRB management is formed as a Markov decision process (MDP) with the decision time resolution τ. At decision time k, the system state space can be described as

\[ X(k) = [R_0(k), P_W(k), E_{INS}(k), S_{DOD}(k)] \]  

(10)

where \( E_{INS}(k) \) is the state of remaining load demand energy calculated by applying wind power \( P_W(k) \) and μCHP generation \( P_{CHP}(k) \) to load demand \( P_{SL}(k) \). \( E_{INS}(k) = \max[0, \tau(P_{SL}(k) - P_W(k) - P_{CHP}(k))] \). \( S_{DOD}(k) \) is the depth of discharge (DOD) state of VRB. The action is to discharge \( \mu(k) \) percent of \( E_{INS}(k) \) from VRB at decision time k. Actions are constrained by minimum and maximum VRB discharging power, as well as VRBs SOC. The reward function for action \( \mu(k) \) is designed considering both the cost saving from discharging and system stability with battery backup energy

\[ r(k) = a_1(k) - \lambda a_2(k) \]  

(11)

where

\[ a_1(k) = \frac{[R_G(k) - R_B] \mu(k) E_{INS}(k)}{(R_{G,max} - R_B) E_{D,max}} \]  

(12)

\[ a_2(k) = \frac{\Delta S_{DOD}(k)}{1 - S_{DOD}(k + 1)} \]  

(13)

\( a_1(k) \) is the normalized cost reduction for the microgrid with VRB discharging. \( R_G \) is the VRB discharging cost (USD/kWh). \( R_{G,max} \) is the maximum value of time-varying utility price. \( E_{D,max} \) is the maximum VRB discharging energy in a decision period. \( a_2(k) \) is the normalized battery DOD change weighted by battery SOC state at time \( k + 1 \). \( \lambda \) is a positive weight for \( a_2 \). When \( a_1(k) \) is larger, which indicates more cost saving is achieved, the action is considered as cost effective and denoted with larger reward. On the other hand, larger \( a_2(k) \) means more energy is discharged from VRB and therefore less energy is available as backup. In that case, the reward is reduced.

The MDP will find the optimal policy \( h^* \) and action \( u_t = h^*(X(k)) \) to maximize the total reward that discounts the future rewards with a factor \( \gamma \)

\[ R = \sum_{n=0}^{\infty} \gamma^n r(k+n). \]  

(14)

D. Problem-Solving Algorithms

The DR is formulated as a nonlinear integer programming problem. For the shared cost-led μCHP management, the fine-grained optimization for the current period is a linear programming problem and the coarse-grained optimization for the future periods is a mixed integer nonlinear programming. These nonlinear programming problems are nonconvex and finding their global optimal solutions is NP-hard. Therefore, in DR and μCHP management, local optimal solutions are solved by particle swarm optimization (PSO) algorithm in real-time for practical operation [21]. References [22]–[25] have evaluated PSO on different benchmarks and shown good solution qualities. PSO has many advantages over other evolutionary algorithms, like genetic algorithm [26] [27]. First, PSO has more effective memory capacity and better diversity for optimal solution search. Second, PSO has faster search speed which is important for highly dynamic systems, such as the DR and shared cost-led μCHP management.

In PSO, a swarm \( S \) is a set of particles \( S = \{x_1, x_2, \ldots, x_N\} \) where \( x_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,M}]^T \) indicating its position in a \( M \)-dimension as the solution to minimize the cost function \( F(x_i) \). The dimension of each particle depends on the number of control variables. Each particle also has its velocity as \( v_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,M}]^T \) as the shift of position in each iteration. The swarm of particles will update their velocities and positions, in each iteration \( k \), toward target solution (with minimum cost) by utilizing both individual best position \( p_i(k) = [p_{i,1}(k), p_{i,2}(k), \ldots, p_{i,M}(k)]^T \) and global historical best position \( p_g(k) = \arg \min F(p_i(k)) \). The update is realized according to the following equations:

\[ v_{i,j}(k+1) = \omega v_{i,j}(k) + c_1 r_1 [p_{i,j}(k) - x_{i,j}(k)] + c_2 r_2 [p_g(k) - x_{i,j}(k)] \]  

(15)

\[ x_{i,j}(k+1) = x_{i,j}(k) + v_{i,j}(k+1) \]

where \( r_1 \) and \( r_2 \) are random variables with uniform distribution in \([0, 1]\). \( c_1 \) and \( c_2 \) are acceleration constants. \( \omega \) is the inertial weight, a value decreasing with time. To prevent swarm divergences, the velocity of \( j \)th component \( v_{i,j}(k+1) \) is clamped as \( |v_{i,j}(k+1)| \leq \max_j = (b_j - a_j)/2 \) as a common selection, where \([a_j, b_j]\) is the feasible region of \( x_{i,j} \). For the constraints of optimization, the method of using penalty function and preserving feasibility of solution during initialization is adopted [24], [28]. For integral variables with a discrete search space, discrete particle swarm optimization (DPSO) with rounding techniques proposed in [25] is used. For binary variables, a binary version of DPSO with sigmoid function is used [29].

As a model-free reinforcement learning technique, Q-learning [30] is used to obtain optimal VRB discharging
Algorithm 1 Q-Learning Algorithm With $\epsilon$-Exploitation

1: Q value initialization;
2: Initial state measurement;
3: for each step $k$ do
4: $u_k \leftarrow \begin{cases} \text{random action selection with probability } \epsilon_k & \text{if } \epsilon_k \geq \alpha \arg\max_{u'} Q_k(x_k, u') \text{ otherwise} \\ 
5: \text{Taking action } u_k, \text{ observe } x_{k+1} \text{ and } r_{k+1}; 
6: Q_{k+1}(x_k, u_k) \leftarrow Q_k(x_k, u_k) + \alpha [r_{k+1} + \gamma \max_{u'} Q_k(x_{k+1}, u') - Q_k(x_k, u_k)] 
7: \text{end for}

E. Bill Balancing Algorithm

To ensure fair energy usage for all households in the ecosystem, their bills need to be balanced according to their energy consumption and generation along the time. Wind and battery energy are allocated to a household with its subscription rate. In bill balancing, it is assumed a household utilizes all of its subscribed wind and battery energy, supplying its load and selling the extra energy to the utility grid. The bill balancing for $\mu$CHP energy generation and consumption is more complex. At one time, a household performs as either a contributor ($i \in M_C$) or a beneficiary ($i \in M_B$). A contributor exports a part of its $\mu$CHP electric energy to the microgrid or reaches a balance between generation and demand without export. On the contrary, a beneficiary consumes electric energy from other $\mu$CHPs in the microgrid. Their relationship in the microgrid is shown in Fig. 3. In each time slot, a household $i$ equipped with $\mu$CHP consumes fuel $F_{\text{CHP},i}$ and generate electric energy $E_{\text{CHP},i}$ and thermal energy $E_{\text{CHP,th},i}$. $E_{\text{CHP},i}$ first supplies its own electric load $E_{\text{L},i}$, which is the remaining load of household $i$ after utilizing its subscribed wind and battery energy. $E_{\text{L,CHP},i}$ is the part of electric energy supply from $\mu$CHPs. For a contributor $i$, $E_{\text{CHP},i} \geq E_{\text{L,CHP},i}$, and $E_{\text{CHP,mg},i}^\text{out}$ is exported to microgrid required by other households. The remaining generation is sold to the utility grid as $E_{\text{CHP,mg},i}^\text{out}$. For a beneficiary $j$, $E_{\text{CHP,mg},i}^\text{in}$ is imported from other $\mu$CHPs to supply its demand which is larger than $E_{\text{L,CHP},j} = E_{\text{CHP,mg},j}^\text{in} + E_{\text{CHP,j}}$. The total $\mu$CHP electric energy import matches export inside the microgrid. With fairness consideration, $E_{\text{CHP,mg},i}^\text{in}$ drawn from the microgrid should be proportional to the household's load $E_{\text{L,j}}'$. The net benefit a household obtains from $\mu$CHP generation in the microgrid is determined by three parts: 1) cost saving from self-generation; 2) cost saving from importing energy from microgrid; and 3) fuel consumption cost. It is fair for a household to get the full benefit from self-generation. Cost saving from energy exporting/importing is achieved by both contributors and beneficiaries. Contributors also consume more fuels to generate energy for beneficiaries. Therefore, the first two parts need to be balanced among households.

In time slot $k$, the balanced bill of a household $i$ has the following formulation:

$$BB_i(k) = C'_{G,i}(k) + C_{W,i}(k) + \beta_i C_F(k) - \mu_i B_{\text{CHP,share}}(k) - \mu_i B_{\text{CHP,sel},i}(k)$$

where

$$C'_{G,i}(k) = R_g(k) \left[ E_{L,i}(k) - \alpha_{s,i} (E_W(k) + E_B(k)) \right]$$

$$C_{W,i}(k) = R_W \alpha_{s,i} E_W(k)$$

$$C_F(k) = R_F \sum_{j \in M_{CHP}} F_{\text{CHP},j}(k)$$

$$B_{\text{CHP,share}}(k) = R_g(k) \sum_{j \in M_C} E_{\text{CHP,mg},j}^\text{out}(k)$$

$$B_{\text{CHP,sel},i}(k) = R_g(k) \min\{E_{\text{CHP,mg},i}^\text{in}, E_{\text{L,CHP},i}\} + R_A(k) E_{\text{CHP,mg},i}^\text{in}(k).$$

$C'_{G,i}(k)$ is energy consumption cost of household $i$ when only wind, battery, and utility grid energy are considered as supply. $C'_{G,i}(k)$ can be negative, which means subscribed wind and battery energy is larger than its demand and the extra energy is sold to the utility grid. $C_{W,i}(k)$ is the charge of wind turbine maintenance. $C_F(k)$ is the total fuel consumption cost for $\mu$CHP generation in the microgrid. $B_{\text{CHP,share}}(k)$ is the total cost saving achieved by exporting/importing $\mu$CHP electric energy inside the microgrid. $B_{\text{CHP,sel},i}(k)$ is the cost saving achieved by household from self-$\mu$CHP generation. $M_{CHP}$ is the set of households with $\mu$CHPs. $E_{L,i}(k)$ is the electric energy demand of household $i$ in time slot $k$. $E_W(k)$ and $E_B(k)$ are total wind energy generation and battery energy discharging, respectively. $\alpha_{s,i}$ is the wind and battery energy subscription rate of household $i$. Different from renewable energy, $\mu$CHP energy is sold with an avoided cost rate $R_A(k)$, which is lower than the retail rate $R_0(k)$. To fairly balance $C_F(k)$ and $B_{\text{CHP,share}}(k)$ for each household, ratios $\beta_i$ and $\mu_i$ should be well designed. It is fair for a household with larger $E_{\text{L,CHP},i}, E_{\text{L,CHP,th},i}$, and $E_{\text{CHP,mg},i}^\text{in}$ to pay more for fuel consumption. $B_{\text{CHP,share}}(k)$ is achieved by $\mu$CHP energy sharing inside the microgrid, which should be balanced according to
$E_{\text{CHPE, MG}, i}^{\text{out}}$ and $E_{\text{CHPE, MG}, i}^{\text{in}}$ of each household. Thus, two metrics $U_{\text{CF}, i}$ and $U_{\text{share}, i}$ are designed to describe the fairness of balancing $C_F(k)$ and $B_{\text{CHP, share}}(k)$, respectively, as

$$U_{\text{CF}, i} = \frac{\beta_i C_F(k)}{E_{\text{CHPE, use}, i}}$$

$$U_{\text{share}, i} = \left\{ \begin{array}{ll}
\mu_i B_{\text{CHP, share}}(k) / E_{\text{CHPE, MG}, i}^{\text{out}} & i \in M_C \\
\mu_i B_{\text{CHP, share}}(k) / E_{\text{CHPE, MG}, i}^{\text{in}} & i \in M_B 
\end{array} \right.$$  (18)

where

$$E_{\text{CHPE, use}, i} = (E_{L, \text{CHPE}, i} + E_{L, \text{CHPH}, i}^{\text{out}}) / \eta_e + E_{L, \text{CHPH}, i}^{\text{use}} / \eta_h.$$  

$U_{\text{CF}, i}$ is the unit fuel cost per $\mu$CHP energy usage, i.e., supplying load and selling to utility grid, in which energy are weighted by $\mu$CHP electric efficiency $\eta_e$ and thermal efficiency $\eta_h$. $E_{L, \text{CHPH}, i}$ is the part of $E_{\text{CHPH}, i}$ for water tank heating. $U_{\text{share}, i}$ is the unit cost saving per $\mu$CHP electric energy export/import inside the microgrid. $\beta_i$ and $\mu_i$ are designed following two rules. First, in consideration of fairness, $U_{\text{CF}, i}$, as well as $U_{\text{share}, i}$, of each household should be equal. Second, $\sum_{i \in M} \beta_i = \sum_{i \in M} \mu_i = 1$, where $M$ is the set of all households in the ecosystem. Thus, $\beta_i$ and $\mu_i$ can be selected as

$$\beta_i = E_{\text{CHPE, use}, i} / \sum_{j \in M} E_{\text{CHPE, use}, j}$$  (19)

$$\mu_i = \left\{ \begin{array}{ll} E_{\text{CHPE, MG}, i}^{\text{out}} / (2 \sum_{j \in M} E_{\text{CHPE, MG}, j}^{\text{out}}) & i \in M_C \\
E_{\text{CHPE, MG}, i}^{\text{in}} / (2 \sum_{j \in M} E_{\text{CHPE, MG}, j}^{\text{in}}) & i \in M_B 
\end{array} \right.$$  (20)

After $\beta_i$ and $\mu_i$ are determined, the balanced bill for each household can be calculated according to (16).

IV. SIMULATION AND RESULTS

A. Simulation Configuration

The simulation platform is implemented with Java. The community is configured with ten houses and four $\mu$CHPs. Suppose residents leave home at 8 a.m., and each day starts at 8 a.m. and ends at 8 a.m. of the next day. Each house is configured with its own fixed load, schedulable tasks (EV charging, laundry machine, dryer, PC downloading, etc.) and preferred execution time periods. The simulation for one week is first evaluated.

The wind velocity is generated according to Rayleigh distribution with an average speed 20 m/s. It is assumed that there is 0%–30% variance between each hourly updated wind forecast. There is also 0%–20% variance between the forecasted and actual wind power generation. The wind turbine has 20 kW rated power output. The actual wind power generation is shown in Fig. 4.

$\mu$CHPs are modeled with electric efficiency 0.27, thermal efficiency 0.63. $g_{\text{fmin}} = 0.0013 \text{MJ}^3 / \text{s}$, $g_{\text{fmax}} = 0.009 \text{MJ}^3 / \text{s}$. The hot water tank has temperature set point 65°C with allowable range $\pm 3\degree C$. VRB is first set with capacity $E_{\text{cap}} = 10 \text{kWh}$ for overall system evaluation. Its discharging power is constrained with $P_{D, \text{min}} = 0.5 \text{kW}$ and $P_{D, \text{max}} = 4 \text{kW}$. Its charging/discharging round-trip efficiency is set to be 0.8. The time-varying utility price for simulation is generated based on the critical peak pricing (CPP) model [31] and shown in Fig. 5. In the hierarchical optimization, time resolution is set to be $\tau = 6 \text{ min}$. For $\mu$CHPs management, parameters are selected with $T_C = T_S = 30 \text{ min}$ and $N_P = 6$. In PSO, the number of particles in a swarm is selected as 100 and 500 for DR and $\mu$CHPs generation optimization, respectively. The maximum number of iterations is selected as 1000. $w$ is selected with the initial value 0.9 and $c_1 = c_2 = 1$. Q-learning parameters are selected as $\alpha = 0.1$, $\gamma = 0.9$, and $u \in [0, 100\%]$.

B. Result Analysis

The proposed energy ecosystem is first compared with a conventional distribution system which is configured with DR and $\mu$CHPs but without the wind turbine and VRB. In the conventional system, $\mu$CHPs are not interconnect and each of them generates power according to the heat demand of its own house. The capital cost for a 20 kW wind turbine is about 70,000 USD (20 years life-span) with approximated maintenance cost 1.5% of the investment cost per year [32], [33]. The VRB discharging cost can be approximated as 0.1 USD/kWh [34]. Even including the cost of the wind turbine and VRB, results in Fig. 6 shows that large cost reduction can be achieved in the ecosystem. Performance of the hierarchical optimization is further evaluated. For easy analysis, the wind turbine investment cost will not be included in the following analysis.

1) Distributed DR Results and Analysis: The update of preference function $F_{pr}$ is affected by the weight parameter $\rho$. When $\rho$ is small, $F_{pr}$ changes slowly. If $\rho$ is set too large, $F_{pr}$ is over sensitive even to a single adjustment and therefore forms bumps, which is inaccurate and will result in DR searching solutions in some local regions. $\rho = 0.1$ is selected
for the following simulations by observing its good trade-off between the learning speed and accuracy. The energy consumption cost with and without DR are compared. A randomly selected house is evaluated for the performance of DR. Results are shown in Fig. 7. The normalized satisfaction $U$ is used to present the influence of DR on users’ satisfaction. Without DR, tasks start at users’ most preferred time with $U = 1$. $U$ with DR for the house is shown in Fig. 8. Results show that with DR, the energy consumption cost of the house in each day has reduced up to 43% while a high satisfaction is still achieved.

2) Centralized Shared Cost-Led $\mu$CHP Management Results and Analysis: The two-level shared cost-led $\mu$CHP management is compared with the heat-led management strategy. DR is applied in both strategies. To evaluate the performance, one day is selected randomly and its electric and thermal load demand of the community are shown in Fig. 9. The total $\mu$CHP total power generation and electric heat pump thermal power generation are shown in Fig. 10. The hot water tank temperature of one house is regulated within the preset region shown in Fig. 11. At the end of each $N\rho$ coarse-grained period (every 3 h), the desired temperature setpoint is achieved. The energy consumption cost of the whole community in each day compared with heat-led management strategy is shown in Fig. 12. Results show that the shared cost-led $\mu$CHP management can reduce the energy consumption cost of the whole community up to 19%. The exact cost reduction depends on wind power generation, utility power price, and load variance of the community.

3) Centralized VRB Management Results and Analysis: The performance of VRB management based on Q-learning is compared with the strategy in which VRB discharges whenever the wind and $\mu$CHP power are insufficient to supply the total load of the community. The total cost reduction of the community from VRB discharging in an extended two-week simulation is shown in Fig. 13. In direct discharging mode, VRB discharges to supply the extra load whenever the utility
The capacity is large, e.g., Figs. 12 and 13. Energy consumption cost of the community with VRB discharging.

As \( \lambda \) increases, more weight is given to energy reservation than load shaving.

V. Conclusion

This paper proposes an energy ecosystem facilitated by DR and DER management with hierarchical agents and optimization. In DR, updated external information and user’s task preference are considered in making decisions. Preference functions, modeled with KDE, are used to describe user’s task preference. With two-level shared cost-led management, \( \mu \)CHPs are fully utilized to reduce the energy consumption cost of the whole community. At last, VRB management with Q-learning obtains the optimal discharging policy considering the utility price and stochastic elements of wind power and load demand. Simulation results show the great effectiveness of this management system on the energy consumption cost reduction.

REFERENCES


Bingnan Jiang (S’13) received the B.S. and M.S. degrees in automation science from Beihang University, Beijing, China, in 2007 and 2010, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering from Northeastern University, Boston, MA, USA.

His current research interests include smart grid, optimization, data mining, and energy-efficient cyber-physical systems.

Yunsi Fei (S’01–M’04) received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1997 and 1999, respectively, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2004.

She was a Faculty Member at the University of Connecticut, Storrs, CT, USA, until 2011, and then joined the Faculty of Northeastern University, Boston, MA, USA, where she is currently an Associate Professor with the Electrical and Computer Engineering Department. Her current research interests include energy-efficient embedded system design and optimization, hardware-oriented security and trust, side-channel attacks analysis and countermeasures, computer architecture, and adaptive networking for underwater sensor networks.

Dr. Fei was the recipient of the National Science Foundation (NSF) CAREER Award. Her research has been primarily supported by the NSF and the Office of Naval Research. She currently serves as an Associate Editor for the ACM Transaction on Embedded Computing Systems, IEEE EMBEDDED SYSTEM LETTER, and the Journal of Low Power Electronics.