Autonomous Two-Tier Cloud Based Demand Side Management Approach with Microgrid

Mohammad Hossein Yaghmae1, 2, Senior Member, IEEE, Morteza Moghaddassian2, Student Member, IEEE, Alberto Leon-Garcia2, Fellow Member, IEEE
1Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
2Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada
hyaghmae@um.ac.ir, m.moghaddassian@mail.utoronto.ca, alberto.leongarcia@utoronto.ca

Abstract—Demand Side Management (DSM) is an important application of the future Smart Grid (SG). DSM programs allow consumers to participate in the operation of the electric grid by reducing or shifting their electricity usage during peak periods. Therefore, in this paper we propose a two-tier cloud-based demand side management to control the residential load of customers equipped with local power generation and storage facilities as auxiliary sources of energy. We consider a power system consisting of multiple regions and equipped with a number of microgrids. In each region an edge cloud is utilized to find the optimal power consumption schedule for customer appliances in that region. We propose a two-level optimization algorithm with a linear multi-level cost function. At the edge cloud, the power consumption level of local storage and the amount of power being demanded from both local storage facilities and power grid are scheduled using a bi-level optimization approach. The core cloud then gathers information of the total demand from consumers in different regions and finds the optimal power consumption schedule for each microgrid in the power system. Simulation results show that the proposed model reduces consumption cost for the customers and improves the power grid in terms of peak load and peak-to-average load ratio.

Index Terms— demand side management, power consumption scheduling, cloud computing, home energy management systems, optimization

I. INTRODUCTION

The Smart Grid (SG) uses two-way communications to gather information from different parts of a power network. This information is used to monitor and control the generation, transmission and distribution equipment. Information and Communications Technology (ICT) is the foundation of many applications in the smart grid. By utilizing ICT capabilities, the smart grid improves the efficiency, reliability and sustainability of the power grid, and it delivers many benefits including: efficient transmission of electricity, quick restoration of electricity after power disturbances, reduced operations and management costs for utilities, low power costs for consumers, reduced peak demand, increased integration of large-scale renewable energy systems, better integration of customer-owner power generation systems and improved security.

Demand Response (DR) which is one of most important applications of smart grid, can be used in the future smart cities to inform consumers about their energy usage and costs. Smart consumers can make decisions autonomously about how and when to use electricity. By developing the Internet of Thing (IoT) technology, it is possible to transfer customer’s power consumption information to the cloud and develop a central demand side management program to control and schedule the customer’s appliances centrally. Without utilizing the smart grid applications, it is not possible to develop the smart cities.

As described in [1], the IoT can be used to furnish intelligent management of energy distribution and consumption in heterogeneous circumstances. By leveraging the IoT-based appliances, the smart customers can send their optimal schedule to the utility companies. In the recent years, by the growth of IoT and digital technologies, smart cities have been becoming smarter than before.

In this paper, we propose a cloud-based DSM program that schedules the power consumption by customers in different regions and in microgrids so that both customer and utility company costs are optimized. There is a noticeable confluence between our proposed approach and the smart cities and IoTs, and that is the concept of service layer abstraction. More clearly, the proposed model can be implemented as an energy management component of the smart city that provides consumer electricity consumption management as a service. This concept defines many benefits including modularity in design smart cities components and reduction the time and effort to extend the smart city services. Furthermore, it is designed to run on commodity hardware on a cloud computing platform, where the aggregation of hardware resources provides more power than any individual computing box. It can be considered as a data-driven model that can be further adjusted for any other utility management services. Cloud-based nature of the proposed model reduces the cost of computation which makes it easier for our future smart cities to deploy such services in future smart cities.

The cloud-based DSM utilizes the processing and storage resources of two-tier cloud computing consisting of the “Smart Edge” and the “Core cloud”, to develop an optimal demand side management. In our architecture, customers are classified into different regions. Each region is controlled by a “Smart Edge” cloud to provide cloud computing resources at the edge of the network precisely to meet low latency requirements as well as to reduce the volume of traffic that needs to traverse the network backbone. The core cloud performs a central optimization at the multi-region level. At the customer side, consumption information and local generation data are forwarded to an edge cloud for the region. The edge cloud runs an optimization process to find the optimal power consumption schedule for both user appliances and local storage. After the edge cloud obtains the optimal schedule for both appliances and
local storage devices, the total optimal load schedule of each region is calculated and forwarded to the core cloud. The core cloud has the load information for each region, and it also knows the total stored energy in each microgrid. It can therefore perform a centralized optimization to schedule the resource usage in the microgrids so that the total multi region cost is minimized. Our main motivations for proposing a cloud-based DSM program is as follows:

1- The single-point of failure and the Distributed Denial-of-Service (DDoS) attacks from compromised nodes are some significant concerns in the demand response programs which are based on master-slave architecture where the utility is the master and customers are slaves. Although, there are solutions that are totally auto reconfigurable and fault-tolerant [2], but DDoS attack still is a big concerns in these approaches. By utilizing the cloud computing, the proposed DSM model can decrease the negative effects of DDoS attacks. The elastic nature of cloud computing allows it to provide the required communication and computation resources, dynamically as needed especially when a DOS attack happens. As the proposed DSM is based on two-tier cloud computing, it can leverage the existing defense method to prevent possible DDoS attacks by rapidly provision resources when any attack happens. The cloud-based DSM model can utilize some popular methods against the DDoS attacks.

2- Current Energy Management Systems (EMS) that are used by utilities to perform the demand response programs suffer from limited memory and storage especially when the number of customers is increased. By increasing the number of customers in the system, to store the customer’s data and run the optimization process more computation and storage resources are needed. In the cloud-based DSM solution, as the optimization program is run at the cloud server the memory and processing resources are always available. When we need more resources (due to increase in the number of customers or the size of optimization problem), the cloud server can easily use some techniques such as auto-scaling to scale up the virtual machine and increase its resources. Increasing the number of customers in the optimization problem also increases the execution time. So we partition the problem into two parts edge and core clouds which provides high scalability and fast response time. Together these “edge clouds” and “core clouds” create a multilayer computing cloud. The motivations for these edge clouds certainly apply to smart grids, and so we explore DSM in this multilayer context. We utilizes the high processing and storage capabilities of multi-tier cloud computing to run central optimization problems.

3- In decentralized DSM, the optimization problem is solved by the Energy Consumption Scheduler (ECS) which is usually placed in the Home Energy Management System (HEMS) or smart meter which has limited computational and capacity power. In the distributed DSM approaches many iterations should be performed to find the optimum solution. For example, for the distributed DSM program given in [3] and for a power network with 100 users, when the channel bit error rate is 0.01 almost $10^6$ update messages are exchanged to converge to the optimal solution. In contrast, in the proposed cloud-based DSM, all necessary calculation is performed at the cloud servers provided by the utility companies. It means that the users do not need to spend money to buy sophisticated HEMS. They just need to participate in the cloud-based energy efficient programs provided by the utility companies or third party to optimize their energy consumption.

4- In DSM programs based on the game theory (such as [3],[4]), customers are classified in some clusters with different members. A local communications network need to be established between all customers. The assumption that the customers have knowledge about their own and the other customers pay-offs is not practical. Furthermore, techniques for solving games using mixed strategies, particularly for large pay-offs matrix, is too complicated. Unlike the decentralized DSM models, the proposed work is based on central optimization at the cloud server. It means that the customers do not need to communicate and corporate together to find the optimal solution. All operation is performed centrally at the cloud server. We just need to collect power consumption information from all the customers and then run the optimization problem. As the central server has a global view of the power system, achieving optimized solution is more feasible than the decentralized approaches which are based on local information.

5- When power consumption scheduling is performed in distributed fashion, security is a big challenge. In the distributed DSM, customers broadcast their local optimal solutions. It has been proven that data broadcasting is not secure. The hackers may accesses to the ECS data, change the users’ consumption and scheduling information, and broadcast fake data to the other users in the same cluster. Cloud computing offers a deployment architecture, with the ability to address vulnerabilities recognized in traditional information security. Cloud-based DSM can be more secure than decentralized DSM by using some approaches such as multifactor authentication, security patching, physical security and security certifications.

6- Current grid technology suffers from peak loads arise from a drop in the supply or an increase in the demand. It also limits demand response to static strategies, such as time of use pricing and day-ahead notification based on historical averages. In the proposed model we consider the PV based microgrid as an auxiliary source of energy in our model and optimize it so that the customer’s cost be minimized. Since microgrids are independent of the power grid, they can continue operating while the main grid is down. They can function as a grid resource for faster system response and recovery. Also, in the proposed model, the use of local sources of energy to serve local loads helps reduce energy losses in transmission and distribution.

The rest of this paper is organized as follows. Section 2, presents the literature review and background. In section 3, we explain the proposed model in detail. Section 4 shows the simulation results that confirm the superior performance of the proposed model. Finally, section 4 concludes the paper.

II. LITERATURE REVIEW AND BACKGROUND

During past few years much research has been devoted to
DSM programs. There is now a rich literature on using optimization techniques and game theory to manage the demand at the customer side by minimizing the cost of power generation or maximizing the customers’ utility [3-9]. Phase Change Materials (PCM) plays a significant role in the future of buildings. PCM can be used for thermal energy storage system because simply it would be possible to include it into building components such as walls. In [10] by considering price based and incentive-based demand response programs, an optimize HEMS which employs the PCM for decreasing the residential demand and cost, has been designed. As investigated in [10], when PCM is combined with HEMS, the battery usage is reduced comparing to the case without PCM. In case of using PCM, most of the battery energy is utilized at the peak hour where the energy price is high. [11] defines a decentralized demand response approach that can be used to minimize the amount of residential power consumption, by maximizing the utilization of the generated power from wind energy resources. For a power system consisting of electric vehicles and wind renewable energy generators, the Distributed W-Learning (DWL) algorithm has been utilized. Each customer device is controlled by an intelligent agent which learns how to meet multiple goals and objectives. However, this approach suffers from some problems. First, distributed DR is based on local objective function so finding the global solution especially when there are a lot of customers in the system is not possible. Second, cloud-based demand response can utilize the auto-scaling capabilities of cloud computing to adjust the necessary computation resources dynamically and provide more scalability and flexibility. The work in [12] for the customer homes with the time of use energy pricing, proposes two different demand response and scheduling approaches including centralized and decentralized. In the decentralized mode, a microprocessor with a stand-alone algorithm is added to the Smart Plug (SP), to schedule the SPs optimally. However, adding microprocessor and related software to the SPs make them expensive for the customers. In the centralized approach, a central controller located at home energy management system, gathers the necessary information from the SPs to optimally schedule the SPs inside the home. However, this model doesn’t consider the use of small power generation facilities and is not as broad as our model. [13] presents a demand response program considering with Distribution Locational Marginal Price (D-LMP) energy market. It is supposed that customers can receive D-LMP price signal through the home gateway. The customers can join to the real-time demand response program by installing some specific equipment. The proposed optimization tries to maximize the benefit of the consumers and minimizes the production cost of the producer. However, it just considers the fixed and shiftable loads and doesn’t support the optimal scheduling of the local PV and microgrid in the model.

Recently cloud computing has received attention for smart grid applications [14]-[16]. Most smart grid applications need reliable and efficient communications. This can be met by utilizing the cloud computing based on software-defined infrastructure [17]. As investigated in [16] and [18], cloud computing brings some opportunities for smart grid applications. Flexible resources and services shared in network, parallel processing and omnipresent access are some features of cloud computing that are desirable for smart grid applications. [19] presents the architecture of cloud-based demand response (CDR) which outperforms the previous work in terms of convergence speed while keeping the same messaging overhead.

III. PROPOSED MODEL

In this section we present our proposed power system model and the cost function.

A. System Model

We assume there are $R$ different regions in the system. In each region $r$, there are $m_r$, $r \in \{1, \ldots, R\}$ customers which are connected to the grid and consume energy. There are $M$ distinct microgrids in the system. Each microgrid consists of Distributed Generation (DG) and Distributed Storage (DS) units. Without loss of generality, we consider Photovoltaic (PV) energy generation in each microgrid. We consider a two-tier cloud consisting of edge and core clouds. The edge cloud gathers the consumption (and also generation) information from all customers in each region, and finds the optimal power consumption schedule for the customers so that the total energy consumption cost is minimized. Using existing data networks, the optimal consumption schedule is transferred to the HEMS. After all the edge clouds have calculated the optimal power consumption schedule of all customers in all regions, the total scheduled load information is transferred to the core cloud. The core cloud computes the total scheduled load gathered from the all edge clouds. Based on the total hourly load in the system, the core cloud schedules the optimal power consumption in each microgrid so that the Peak to Average (PAR) ratio is decreased.

B. Power Consumers

We define three different types of power consumers:

- **Type 0**: These are traditional power consumers. Type 0 consumers neither have local generation and storage, nor a home energy management system. It is not possible to schedule the power consumption for these consumers. These consumers do not require a data connection to the cloud. We consider these consumers as a variable hourly load in the power system which is not shiftable.

- **Type 1**: These consumers have a home energy management system, and two types of appliances, shiftable and non-shiftable. These consumers have a data connection to the cloud using HEMS. In order to minimize power consumption cost and also reduce the peak to average ratio, the power consumption of shiftable appliances is scheduled.

- **Type 2**: These are sophisticated consumers that are equipped with local generation and storage as well as home energy management system. When the main power grid is disconnected which usually happens due to outages and the other source of blackout, the consumers can use the local energy generation and storage systems. These consumers generate part
of their required energy using PV systems. We assume that the generated power is stored in the local storage (battery) and is consumed by local appliances.

Let $m^0_r$, $m^1_r$, and $m^2_r$ denote the number of Type 0, Type 1 and Type 2 consumers in region $r$, $r \in \{1, ..., R\}$, respectively. As Type 0 consumers are not controllable, in the rest of this paper we consider just Type 1 and Type 2 consumers. For any customer $n$ in region $r$, $n \in \{1, ..., m_r\}$, $r \in \{1, ..., R\}$, let $A_{r,n}$ denote the set of appliances of this customer. We define $X_{r,n}$ as the energy consumption scheduling vector, where $a \in A_{r,n}$ denotes the appliance number and $H$ is the scheduling horizon that indicates the number of hours ahead which are taken into account for decision making in energy consumption scheduling. We define $X_{r,n}$ vector as energy consumption scheduling vector for all appliances of customer $n$ in region $r$. For Type 2 consumers we define $P_{r,n} = \{p^1_{r,n}, p^2_{r,n}, ..., p^6_{r,n}, ..., p^H_{r,n}\}$ as the total power generation vector where, $p^h_{r,n}$ denotes the amount of energy generated by the local PV system of customer $n$ in region $r$ at time $h$. In the next subsection, we will describe the proposed local generation model in detail.

The State of the Charge (SOC) is one of the most important parameters for batteries which is defined as the ratio of its current capacity to the nominal capacity. The nominal capacity represents the maximum amount of charge that can be stored in the battery. The Coulomb counting method has been applied in order to estimate SOC [20]. Suppose $SOC_{r,n}^h$ and $B_{r,n}^F$ represent the SOC and the nominal battery capacity of customer $n$ in region $r$ at time $h$, respectively. The current value of SOC ($SOC_{r,n}^h$) based on its previous value ($SOC_{r,n}^{h-1}$) and the charging/discharging current, $I_r(t)$ is estimated as:

$$SOC_{r,n}^h = SOC_{r,n}^{h-1} + \frac{I_r(t)}{B_{r,n}^F} \Delta t \quad (1)$$

We define $B_{r,n} = \{b^1_{r,n}, b^2_{r,n}, ..., b^6_{r,n}, ..., b^H_{r,n}\}$, as the battery state vector where $b^h_{r,n}$ denotes the amount of energy stored in the battery of customer $n$ in region $r$ at time $h$. Due to limited capacity of local storages, the following condition should be satisfied:

$$0 \leq b^h_{r,n} \leq B_{r,n}^F \quad (2)$$

The value of $b^H_{r,n}$ is obtained by the following equation:

$$b^H_{r,n} = SOC_{r,n}^h, B_{r,n}^F \quad (3)$$

Suppose $g_{r,n}^h$ and $y_{r,n}^h$ denote the amount of energy generation and consumption of Type 2 customer $n$ in region $r$ at time $h$, respectively. We assume that the customer consumes its available power in local storage first and then demands power from the power grid, if needed. We define $G_{r,n} = \{g^1_{r,n}, g^2_{r,n}, ..., g^6_{r,n}, ..., g^H_{r,n}\}$ and $Y_{r,n} = \{y^1_{r,n}, y^2_{r,n}, ..., y^6_{r,n}, ..., y^H_{r,n}\}$ as the power generation and consumption vector from the PV system, respectively. As the battery is charged by solar energy ($g_{r,n}^h$) and is discharged by local consumption ($y_{r,n}^h$), by combining equ.(3) in equ. (1) we have:

$$b^h_{r,n} = b^{h-1}_{r,n} + g^h_{r,n} - y^h_{r,n}, h = 2, ..., H \quad (4)$$

We define $v^h_{r,n} = X^h_{r,n} - y^h_{r,n}$ as the total hourly household energy consumption at each upcoming hour $h$.

### C. Consumer Price Model

The proposed price model is designed by combining the Real Time Pricing (RTP) and Inclining or Increasing Block Rates (IBR) and considering $N$ different blocks shown in Fig. 1. Our main objective of proposing this pricing model is to charge a higher rate per kWh at higher levels of energy usage, and a lower rate at lower usage levels. The total cost of the customers who consume low energy (lower blocks), is less than those who consume more energy (higher blocks). The first block of $B_1$ kilowatt-hours (kWh) would cost $c_1$ per kWh, the second block of $B_2$ kWh would cost $c_2$ per kWh, and so on. Note that $c_1 > c_{i-1} > 0$, $i \in \{2, ..., N\}$, which means that the proposed pricing model charges a higher rate for each incremental block of consumption. To support real time pricing, the price factors $c_i$ are time-dependent and may be changed hourly. For example suppose $N = 7$ and $B_1 = 1$ KWh. Based on the proposed pricing model the total power consumption cost for a customer with 6.5KWh power consumption is computed as $2 \times c_1 + 0.5 \times c_2$. Unlike the ToU and flat pricing, the introduction of IBR leads to energy savings. Another advantage of the proposed pricing is that it is straightforward and easy to understand by households. In the proposed pricing model, the cost of each block can be changed in real time. For each customer $n$ in region $r$ at time $h$, the power consumption cost $c_{r,n}^h (v^h_{r,n})$ is calculated as follows:

$$c_{r,n}^h (v^h_{r,n}) = \begin{cases} c_1^h, & \text{if } B_0 = 0 \leq v^h_{r,n} \leq B_1 \\ \sum_{j=2}^{N-1} c_j^h (B_{j-1} - B_{j-2}) + c_j^h (v^h_{r,n} - B_{j-1}) & \text{if } B_{j-1} < v^h_{r,n} \leq B_j \\ \sum_{j=1}^{N-1} c_j^h (B_{j-1} - B_{j-2}) + c_j^h (v^h_{r,n} - B_{j-1}) & \text{if } B_{N-1} < v^h_{r,n} \end{cases} \quad (5)$$

It can be seen that the proposed cost function is increasing and strictly convex. It means that:

$$c_{r,n}^h (v^h_{r,n}) < c_{r,n}^h (v^h_{r,n}), \quad \forall v^h_{r,n} < v^h_{r,n} \quad (6)$$

$$c_{r,n}^h (v^h_{r,n} + (1 - \varepsilon) v^h_{r,n}) < c_{r,n}^h (v^h_{r,n}) + (1 - \varepsilon) c_{r,n}^h (v^h_{r,n}) \quad (7)$$

![Fig. 1. The proposed N levels Inclining Block Rate (IBR) pricing model](image)

As the proposed power system model is based on the hierarchical model which consists of different regions, we can define different price functions for different regions. This is because each region is responsible for the power consumption optimization of its customers. So, the proposed region-based optimization, allows us to take the geolocation of customers
into account which help to prevent possible congestion and peak loading within particular regions. For instance, when a power overload occurs at a particular region, the utilities may increase the power price at that particular time at the specific region. Thus customers are encouraged to decrease their power consumption or shift it to the non-peak hours.

**D. Distributed Power Generation Model**

We suppose that PV systems are used for distributed generation by Type 2 consumers and in microgrids. It has been proven that the total energy generation by PV systems depends on parameters such as: temperature, total solar panel area, solar panel yield, performance of the installation including all losses (inverter losses, temperature losses, DC and AC cables losses, shadings, losses weak radiation, losses due to dust and snow).

Previous studies [21]-[24] show that the power output of a PV module depend linearly on the operating temperature. The electrical performance is primarily influenced by the type of PV used. A typical PV module converts 6-20% of the incident solar radiation into electricity, depending upon the type of solar cells and climatic conditions. The rest of the incident solar radiation is converted into heat, which significantly increases the temperature of the PV module and reduces the PV efficiency of the module. To consider the effect of temperature in our PV generation model, based on work given in [21],[24] we propose the following equation:

$$G_{pv} = A, \mu, \theta, \tau_g (1 + \alpha(T_m - 25))$$  \hspace{1cm} (8)

where $G_{pv}$ is the total annually energy generation (KWh), $A$ is total solar panel area (m²), $\mu$ is solar panel yield (default value %15), $\theta$ is annual average irradiation on tilted panels which is changed regionally (between 500 to 2500 KWh/m².an) , $\tau_g$ is performance ratio and coefficient for losses (default value %75) , $T_m$ is the temperature (in centigrade) and $\alpha$ is the temperature coefficient for power of the PV module, which is $-0.20\%/\text{°C}$.

Suppose $G_{h,d}$ represents the solar cell energy generation at each hour $h$ in day $d$ when the sky is clear. When the sky is cloudy the energy generation is a portion of $G_{h,d}$ depending on the cloud coverage. It is clear that the cloud coverage is related to the time and the day of year. As it has been investigated in [25], at each hour $h$ in day $d$, the amount of solar energy generation $G_{h,d}$, is calculated as follows:

$$G_{h,d} = K . f(d) . g(h) . (1 - 0.75F_{h,d})^3$$  \hspace{1cm} (9)

where $F_{h,d}$ is the fraction of sky cloud cover on a scale from 0 (no clouds) to 1 (complete coverage) at time $h$ of day $d$. $K$ is a normalization constant and $f(d), g(h)$ are two functions which indicate how much of the sun’s power can be captured in each hour for a particular day of the year. $f(d)$ is given as follows[26]:

$$f(d) = \frac{\sin(90 - \varphi + \phi(d))}{\sin(90 - \varphi + \phi(d) + \theta)}$$  \hspace{1cm} (10)

where $\varphi$ and $\theta$ are latitude and the tilt angle of the module measured from the horizontal, respectively. $\phi(d)$ is declination angle calculated as[26]:

$$\phi(d) = 23.45\cdot \sin \left( \frac{360}{365} (284 + d) \right)$$  \hspace{1cm} (11)

The tilt angle has a major impact on the solar radiation incident on a surface. For a fixed tilt angle, the maximum power over the course of a year is obtained when the tilt angle is equal to the latitude of the location ($\varphi = \theta$). We propose the flowing Gaussian function for $g(h)$:

$$g(h) = e^{-\frac{(h-h_0)^2}{2\sigma^2}}, \quad h_0 = (t_{\text{sunset}} + t_{\text{sunrise}})/2$$  \hspace{1cm} (12)

where $h_0$ and $\sigma^2$ are, the center and the variance of the Gaussian function. $t_{\text{sunset}}$ and $t_{\text{sunrise}}$ are the sunset and sunrise time, respectively. By substituting equs. (10)-(12) into equ. (9), $G_{h,d}$ is obtained as follows:

$$G_{h,d} = K \cdot \frac{\sin(90 - \varphi + \phi(d))}{\sin(90 - \varphi + \phi(d) + \theta)} e^{-\frac{(h-h_0)^2}{2\sigma^2}} (1 - 0.75F_{h,d})^3$$  \hspace{1cm} (13)

As the total energy generation during all hours of all days in a year should be equal to annual energy generation, the constant parameter $K$ is obtained as follows:

$$K = \frac{\frac{G_{pe}}{\sqrt{365 \cdot \gamma_{24}}} e^{-\frac{(h-h_0)^2}{2\sigma^2}} (1 - 0.75F_{h,d})^3}{\frac{\gamma_{24} - \frac{\sin(90 - \varphi + \phi(d))}{\sin(90 - \varphi + \phi(d) + \theta)} e^{-\frac{(h-h_0)^2}{2\sigma^2}}}{\sqrt{365 \cdot \gamma_{24}}}}$$  \hspace{1cm} (14)

To investigate the accuracy of PV generation model, we compare the output of PV generation model with that of the 3.15 kW PV system located in West Hobart, TAS, Australia [27], and for date July 1, 2016. The real PV system generations were measured and compared with the output of the proposed model. The sunrise time, sunset time and altitude of West Hobart, TAS were set to 8 am, 17 pm and 24°S, respectively. The results given in Table 1, confirm that the proposed PV model has almost 0.13 Mean Absolute Error (MAE).

**E. Cloud Cover Prediction**

Since the proposed model is based on a day-ahead optimization, we have to know the amount of energy generation of the Type 2 consumers and the microgrids in each hour of the next day. For this purpose, we use equ.(13), to evaluate the amount of energy generation of PV systems for the next day. We also need to estimate the cloud coverage $F_{h,d}$. The Normalized Least Mean Square (NLMS) [28] predictor is the one providing the best trade-off between complexity, accuracy and responsiveness. We used the historical cloud coverage data given in [29] to tune the filter parameters. The NLMS predictor needs the configuration of two parameters: the order $p$ and the step size $\mu$. These parameters should be set correctly so that the best performance with minimum error is obtained. In the case of the $\mu$, it is relevant to note that one of the main advantage of using NLMS is that it is less sensitive to the step size with respect to other linear predictor. In Fig. 2, at each hour of a day
for the period of Jan.1, 2015-Jan.1, 2016 and for the Toronto city, the real cloud cover and its prediction is plotted versus each hour of the day in the year (totally 24*365=8760 samples).

F. Edge Cloud Cost Optimization

As mentioned earlier, in each region there is an edge cloud that gathers all the customers’ consumption information to optimize the power consumption schedule. The optimal schedule is obtained by the use of two-level optimization approach with the predefined cost function as follows.

- Level 1 optimization (for both Type 1 & Type 2 consumers): In this stage the consumption pattern of all customer’s shiftable appliances in each region is scheduled using the following optimization problem.

\[
\text{minimize } \sum_{h=1}^{H} c_{r,n}^{h} \left( \sum_{n=1}^{m_r} \sum_{a \in A_{r,n}} x_{r,n,a}^{h} \right) \quad (15)
\]

Subject to:
\[
\alpha_a \leq x_{r,n,a}^{h} \leq \beta_a \\
\sum_{n=1}^{m_r} \sum_{a \in A_{r,n}} x_{r,n,a}^{h} = E_{r,n,a} \\
y_{r,n}^{min} \leq \sum_{a \in A_{r,n}} x_{r,n,a}^{h} \leq y_{r,n}^{max}
\]

Output: \(x_{r,n,a}^{h}\)

where \(E_{r,n,a}\) denotes the total energy needed for the operation of appliance \(a\) of customer \(n\) in region \(r\). \(\alpha_a, \beta_a\) are the beginning and end of a time interval within which the energy consumption for appliance \(a\) is valid, \(y_{r,n}^{min}\) and \(y_{r,n}^{max}\) are the minimum and maximum power levels denoted of home appliances. After optimization and thus rescheduling of shiftable appliances, for each shiftable appliance \(a\) of customer \(n\) in region \(r\), the optimized energy consumption scheduling vector \(X_{r,n,a}^{h}\) is obtained. By considering the non-shiftable and shiftable appliances the optimal energy consumption schedule vector \(X_{r,n}^{h}\) for all appliances of customer \(n\) in region \(r\) is obtained.

- Level 2 optimization (only for Type 2 consumers): Type 2 consumers generate part of their energy consumption using the local generation such as the PV system. This energy is stored in the batteries and consumed at the proper time. The question is that to minimize the energy cost, what is the best time for consuming this stored energy? To answer this question, the following optimization problem is defined:

\[
\text{minimize } \sum_{h=1}^{H} \sum_{a \in A_{r,n}} c_{r,n}^{h} \left( x_{r,n,a}^{h} - y_{r,n}^{h} \right) \quad (16)
\]

Subject to:
\[
\sum_{a \in A_{r,n}} x_{r,n,a}^{h} \leq x_{r,n,a}^{h}, \\
\sum_{a \in A_{r,n}} x_{r,n,a}^{h} = E_{r,n,a} \\
y_{r,n}^{min} \leq \sum_{a \in A_{r,n}} x_{r,n,a}^{h} \leq y_{r,n}^{max}, \\
0 \leq b_{r,n}^{h} \leq g_{r,n}^{h}, \\
\sum_{h=1}^{H} g_{r,n}^{h} = \sum_{h=1}^{H} y_{r,n}^{h}, \\
y_{r,n}^{i,i+1} \leq b_{r,n}^{i} - b_{r,n}^{i-1}, \quad i = 2, \ldots, H
\]

Output: \(y_{r,n}^{*}\)

where \(y_{r,n}^{*}\) is the optimal energy consumption obtained by the first optimization and is known. After the above optimization problem is solved the optimal power consumption vector \(Y_{r,n}^{*}\) for local storages is obtained. The consumption load from the grid is obtained as \(l_{r,n}^{*} = X_{r,n}^{*} - Y_{r,n}^{*}\). The optimization process in the edge cloud involves two important entities: smart meter and edge cloud server. The smart meter is involved for the purpose of power consumption information communication while the edge cloud server is involved for running the optimization problem. The smart meter communication with the edge cloud server is explained in Algorithm 1. This algorithm is performed by smart meters for just smart consumers (Type 1 and Type 2 consumers).

The algorithm for both level 1 and level 2 optimization in the edge cloud server is explained in detail in Algorithm 2. This algorithm is performed by each regional edge cloud for its own Type 1 and Type 2 consumers.

---

**Algorithm 1:** Executed by each smart meter of Type 1 and Type 2 consumers

1: Randomly initialize \(X_{r,n,a}^{h}\) for Type 1 consumers and \(G_{r,n}^{h}\), \(B_{r,n}\) and \(Y_{r,n}^{h}\) vectors for Type 2 consumers
2: Repeat
3: At random time instances Do
4: Send vectors \(X_{r,n,a}^{h}\) for Type 1 and Type 2 consumers and \(G_{r,n}^{h}\), \(B_{r,n}\) and \(Y_{r,n}^{h}\) vectors for Type 2 consumers to the regional edge cloud server
5: if changes happen to the vectors as a result of re/scheduling Then
6: Receive optimized vectors \(X_{r,n,a}^{*}\) for Type 1 and Type 2 consumers and \(Y_{r,n}^{*}\) just for Type 2 consumers from the regional edge cloud server
7: Build vector \(X_{r,n}^{*}\) and Update the state of the consumer by building \(l_{r,n}^{*}\) vector
8: End
9: End
10: Until the meter is in service

**Algorithm 2:** Executed by each edge cloud

1: Repeat
2: For any Type 1 and Type 2 consumer in the region Do
3: Receive vectors \( x_{r,a} \) from any Type 1 and Type 2 consumers and \( G, B_j \) vectors from any Type 2 consumers in the region
4: Solve linear problem (15) using IPM (interior-point method)
5: if changes happen to vector \( x_{r,a} \) Then
6: Send each vector \( x_{r,a} \) to its corresponding consumers
7: End
8: Solve linear problem (16) using IPM (interior-point method)
9: if changes happen to vector \( y_{r,a} \) Then
10: Send each vector \( y_{r,a} \) to its corresponding consumers
11: End
12: Build \( l^h_s \) vector and Store it in the local memory
13: End
14: Build the vector \( l^h_s = y_{r,a} \) and Send it to the core cloud
15: Until there is at least one edge cloud server or one microgrid in service

G. Core Cloud Cost Optimization

After optimization process is completed by each edge cloud, it sends its total hourly demand load vector to the core cloud. Each microgrid \( j \in \{1, 2, ..., M\} \) also sends vectors \( G, B_j \) and \( Z_j \) to the core cloud which represents the prediction of hourly power generation by microgrid, the remaining energy in the batteries and the hourly power consumption level of each microgrid \( j \in \{1, 2, ..., M\} \), respectively. Let \( L^h = (L^h - z^h) \), denote the total hourly load in the power system where \( L^h \), the total hourly demand load from all regions and \( z^h \), the total hourly power consumption of all microgrids in the system are calculated as follows:

\[
L^h = \sum_{r=1}^{R} \sum_{m=1}^{m_r} l_{r,m}^h
\]

\[
z^h = \sum_{j=1}^{J} z_j^h
\]

We consider the same linear multi-level cost function from equ. (5) but with different cost coefficients. The following optimization problem is defined at the core cloud:

\[
\text{minimize} \sum_{h=1}^{H} C_h(L_h) = \sum_{h=1}^{H} C_h(L^h - z^h)
\]

Note that in the above optimization problem the value for \( L^h \) is fixed and is already calculated by edge clouds in the grid. Therefore after running the optimization process at the core cloud, the optimized hourly consumption \( z^h \) from microgrid storages is obtained. The optimized total hourly load \( L^h = L^h - z^h \) which indicates the power consumption level from power grid, is obtained. The optimization algorithm is explained in detail in Algorithm 3. This algorithm is performed by the core cloud for all the entire grid after collecting the supporting data from all regional edge clouds and microgrids.

Algorithm 3: Executed by the core cloud

1: Repeat
2: For any microgrid \( j \in M \) Do
3: Receive vectors \( G, B_j \) and \( Z_j \)
4: End
5: Build vector \( Z^h \) from the state vectors that have been sent by each microgrid
6: Receive vector \( l^h \) from each edge cloud server
7: Solve linear problem (19) using IPM (interior-point method)
8: If changes happen to the optimization vectors in problem (19) Then
9: Build vectors \( Z^h \) and \( L^h \) and Update the state of the grid
10: End
11: Until there is at least one edge cloud server or one microgrid in service

The operation of whole system is described by the flowchart given in Fig. 3.

IV. SIMULATION RESULTS

To compare the performance of the proposed model with that of existing work, we consider a central version of work given in [3] without any Type 2 consumers and microgrids. We name it Reference Model. In the simulation model, we consider a power system that includes 5 regions with 2000 customers in each region. At each region, the percentage of Type 0, Type 1 and Type 2 consumers is 50, 30 and 20, respectively. For each customer we also consider 5 to 10 appliances with shiftable and non-shiftable operations. The initial parameters of all appliances are set using data given in [30]. We also consider that each Type 2 customer is equipped with the average of 30 square meters of photovoltaic cells. Similarly, we consider 5 microgrids with 10000 square meters of solar cells for each. We suppose that the weather temperature for all solar cells is equal to 25° Centigrade. We consider three different power usage intervals (12 AM to 8 AM, 8 AM to 5 PM and 5 PM to 12 AM) that correspond to off-peak, mid-peak and high-peak hours of the day, respectively. The price of each KW power consumption at off-peak, mid-peak and high-peak hours are supposed to be equal to 0.2, 0.3 and 0.5 cent, respectively. Then for each hour of the day we consider a seven-level pricing system that corresponds to the seven levels of load as defined before by the cost function. We consider the latitude of Toronto, Canada and set \( \varphi = 43^\circ \text{North} \). For maximum energy observation we set the tilt angle of the module to the latitude as \( \theta = \varphi = 43^\circ \). We consider 31 days simulation interval between July 1 to July 31, 2015 \((d \in [182, 212])\). The cloud coverage data given in [29] is used for training the proposed predictor. We also assume a daily time granularity (H=24). This means that the overall cloud solves the optimization problem for the next 24 hours.

In the first scenario, we investigate the effects of number of smart consumers (Type 1 & 2) on the performance of the power system. The results given in Fig. 4 confirm that by increasing the number of smart consumers, the Peak to Average Ratio (PAR) is decreased. For example, when there are 50% Type 2 and 30% Type 1 consumers, we can achieve 0.52 decrease in PAR performance comparing with the case that all consumers are Type 0. This results from the higher number of sophisticated participants as well as the higher number of distributed generation resources available.

In Fig. 5, for a Type 2 customer and at day 182, the hourly local power generation and usage from the battery (both Unoptimized and Optimized scenarios) are depicted. It can be seen that almost 70% of power consumption of the total generation of local PV system is shifted to the peak hours where the price of power is high. Therefore the total cost and amount of power consumption from the power grid will be reduced and the grid will experience lower PAR.

In Fig. 6, for three candidates Type 0, Type 1 and Type 2 in a given region and at day 182, the hourly power consumption and the power cost are depicted. It can be seen that Type 2 consumers save 53.8% and 32.9% in power consumption cost in comparison to Type 0 and Type 1 consumers, respectively.
In Fig. 7, for a candidate region with 2000 customers, the hourly regional power demand level and costs are depicted. The figure illustrates a significant improvement from the No Optimization case and the Reference Model [3]. It can be seen that in case of both level 1 and level 2 optimization (Proposed Model), the total daily cost can be reduced to 26.4% in comparison with the No Optimization case and 6.25% in comparison with the Reference Model [3], with regards to the percentage of Type 1 & Type 2 customers in the region.

In Fig. 8, for a power system with 5 microgrids, and for day 182, the amount of power generation in microgrids and power consumption schedule for the microgrid storage are depicted hourly. As the power usage from microgrids is stored in the battery and consumed at peak hours where the price is high, the results confirm that the optimized power consumption scheduling of the microgrid resources save almost 1.04$ per customer daily, in comparison with unoptimized consumption. This helps to supply part of the customers’ demands for electricity without using grid resources at peak hours. Note that the black bar in Fig.8, show how much energy is generated by the microgrids at a particular day while the dash bar in Fig.8 shows when and how much of this stored energy is consumed by the customers. This confirms that we store energy in the microgrid’s battery during the daytime when the solar energy is high and then consume it at the peak hours when the demand and the energy prices are both high.

In Fig.9, for two different cases of the proposed model (Optimized and Unoptimized) the performance of total grid is
evaluated. Optimized refers to the case that microgrid resources are used in optimal way (after core cloud Algorithm 3 optimization) while Unoptimized refers to using microgrid resources without any optimal scheduling. The results confirm that using microgrid with optimized scheduling (Proposed Model, Optimized) can significantly improve grid demand level and cost efficiency. For the example, we can save 17.9% and 10.7% in total cost in comparison with No Optimization and the Reference Model [3], respectively. We note that there are 5 microgrids in the whole grid that are assisted with 10000 square meters of photovoltaic cells which are capable of generating almost 45KW of electricity power during a day. By increasing the number of microgrids or the area microgrid solar cells, we can save more in cost and demand.

Fig. 7. Hourly power demand level and cost comparison for a candidate region

Fig. 8. The effects of optimization on the hourly power consumption level in microgrids

Fig. 9. Hourly grid demand level and costs comparison between No Optimization, Reference Model [3] and the Proposed Model

In Fig. 10, for four different approaches (No Optimization, Reference Model [3] and Optimized and Unoptimized of the Proposed Model), the total PAR in the grid is plotted at different simulation days (from July 1 to 31). As in Fig.9, Optimized refers to the case where microgrid resources are used in optimally (core cloud optimization), and Unoptimized refers to

<table>
<thead>
<tr>
<th>Region Number</th>
<th>All Type 0</th>
<th>Reference Model [3] (50% Type1)</th>
<th>Proposed Model (30% Type1, 20% Type2)</th>
<th>Proposed Model (30% Type1, 50% Type2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.82</td>
<td>1.35</td>
<td>1.34</td>
<td>1.2853</td>
</tr>
<tr>
<td>2</td>
<td>1.83</td>
<td>1.36</td>
<td>1.35</td>
<td>1.2860</td>
</tr>
<tr>
<td>3</td>
<td>1.86</td>
<td>1.36</td>
<td>1.35</td>
<td>1.2944</td>
</tr>
<tr>
<td>4</td>
<td>1.85</td>
<td>1.35</td>
<td>1.34</td>
<td>1.2880</td>
</tr>
<tr>
<td>5</td>
<td>1.83</td>
<td>1.33</td>
<td>1.32</td>
<td>1.2846</td>
</tr>
</tbody>
</table>
use of microgrid resources without any optimal scheduling. It can be seen that using more sophisticated customers and microgrids with optimized power scheduling can significantly reduce the total PAR in the whole grid.

Finally, we investigate the scalability of the proposed model in terms of the convergence time of the optimization process time for different numbers of customers and regions. The execution time of the whole optimization process (including edge and core clouds) is measured. The proposed optimization is based on 3 different phases including collecting the information, running the optimization and transferring the optimal schedule to the customers and microgrids. The optimization processing time directly depends on the performance of the cloud server and the number of customers and regions. When customers are spread in different regions, the required time for gathering information and running the optimization process decreases. By increasing the number of customers in the system, more computation and storage are needed to store the customer’s data and run the optimization process. The simulation results show that increasing the number of customers in the optimization problem also increases the execution time. To avoid high execution time, we propose to partition the entire system into distinct regions and solve the optimization problem at each region, separately. Thus, high scalability and fast response time will be achieved. At the end of optimization process, the optimal schedule is sent to the customers and microgrids. The results given in Table III confirm that 1) when the number of regions is fixed, by increasing the number of customers in each region the convergence time is increased linearly, 2) when total number of customers in whole power system is fixed, by increasing the number of regions and spreading the customers in different regions, the convergence time is decreased significantly.

![Fig. 10. Investigating the effects of using more sophisticated customers and using microgrids’ power generation and storage facilities on the grid PAR](image)

**Table III. The Processing Time Unit for Running the Whole Optimization Process at Different Combination of Customers and Regions**

<table>
<thead>
<tr>
<th>Number of Customers in Each Region</th>
<th>Number of Regions</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>NA</td>
<td>89.5062</td>
<td>NA</td>
<td>90.7032</td>
<td>91.2865</td>
</tr>
<tr>
<td>1000</td>
<td>NA</td>
<td>179.0042</td>
<td>179.6313</td>
<td>186.3749</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>NA</td>
<td>359.1821</td>
<td>360.4404</td>
<td>373.9718</td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td>NA</td>
<td>1764.2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>20000</td>
<td>NA</td>
<td>3517.8148</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

**V. CONCLUSION**

Demand side management needs reliable and efficient communications which can be met by utilizing the multi-tier cloud computing based on software-defined infrastructure. In this architecture, edge cloud provides cloud computing resources at the edge of the network. The benefit of such architecture is that it can provide a high level of scalability and reliability. In this paper we proposed a two-tier cloud-based model for the autonomous demand side management in the future smart grid in which the customer's power consumption and microgrid resources are scheduled by the use of regional edge and core clouds, respectively to reduce the cost and improve the power grid performance. It has been shown that spreading the customers in different regions, reduces the convergence time and improves the scalability. As the proposed approach is able to provide online access to all customer power consumption information and microgrid resources, so it can enable dynamic demand response optimization of the power consumption and energy cost of the customers. Simulation results confirmed that the proposed model reduces the cost for the customers and improves the power grid in terms of peak load and peak-to-average load ratio.

**REFERENCES**


Mohammad Hossein Yaghmaee Moghaddam (SM’12) received his B.S. degree from Sharif University of Technology, Tehran, Iran in 1993, and M.S. and Ph.D degrees in communication engineering from Tehran Polytechnic (Amirkabir) University of Technology in 1995 and 2000, respectively. November 1998 to July 1999, he was with Network Technology Group (NTG), C&C Media research labs., NEC corporation, Tokyo, Japan, as visiting research scholar. September 2007 to August 2008, he was with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, USA as the visiting associate professor. July 2015 to September 2016, he was with the electrical and computer engineering department of the University of Toronto (UoT) as the visiting professor. Currently, he is a full professor at the Computer Engineering Department, Ferdowsi University of Mashhad (FUM). His research interests are in Smart Grid Communication, Software Defined Networking (SDN) and Network Function Virtualization (NFV).

Morteza Moghaddassian (S’15) received his M.S. degree in Computer Engineering from Ferdowsi University of Mashhad (FUM), Iran in 2015. Since 2016, he is a Ph.D. student of Electrical and Computer Engineering at University of Toronto, ON, CA where he is pursuing his doctoral studies on Communications System. He is also the recipient of the Edward S. Rogers Sr. departmental graduate fellowship. Prior to this, he was a Research Assistant at the Center of Excellence on Soft Computing & Intelligent Information Processing (SCIIIP), FUM where he successfully managed a research team to develop a social energy network application for the purpose of metering with built-in demand-response management modules, called “SocioGrid”. His research interests include Cloud Computing, Edge Computing, Resource Monitoring and Management, Smart Grid, Optimization of Communication Networks and Open Networking.

Alberto Leon-Garcia (F’99) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1973, 1974, and 1976, respectively. He was founder and CTO of AcceleLight Networks, Ottawa, ON, Canada, from 1999 to 2002, which developed an all-optical fabric multiliterabit, switch. He is currently a Professor in Electrical and Computer Engineering at the University of Toronto, ON, Canada. He is the author of the leading textbooks Probability and Random Processes for Electrical Engineering and Communication Networks: Fundamental Concepts and Key Architecture. His current research interests include application-oriented networking and autonomic resources management with a focus on enabling pervasive smart infrastructure. Prof. Leon-Garcia is a Fellow of the Engineering Institute of Canada. He received the 2006 Thomas Eadie Medal from the Royal Society of Canada and the 2010 IEEE Canada A. G. L. McNaughton Gold Medal for his contributions to the area of communications.