IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 February 2024, accepted 21 March 2024, date of publication 27 March 2024, date of current version 2 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3381624

RESEARCH ARTICLE

Timeliness Study of Home Energy Management System Based on Dynamic Allocation of Bandwidth

WEI ZHANG, (Member, IEEE), AND CHENHAO HUANG
College of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China
Corresponding author: Wei Zhang (zw19830904@usst.edu.cn)

ABSTRACT Household electrical appliances have developed into a significant part of demand response as a result of the widespread use of smart meters and the popularity of home energy management systems. Load-side resources will still be wasted as a result of the transmission and queuing delays brought on by insufficient communication resources. To address the difficulties of large-scale, decentralized, and dispatchable household electricity load engaging in demand response, this study presents an approach to enhance the timeliness of dispatchable resources in home energy management systems, which may be separated into two categories. The most effective approach to allocate bandwidth is determined in the first section using historical bandwidth demand data and ARIMA models. The second section, meanwhile, makes use of offloading strategies and the cloud edge cooperation architecture to decrease queuing delays brought on by computational workloads. The results demonstrate that the suggested strategy can effectively reduce the time needed for information transmission and queuing, as well as alleviate the information interaction time between supply and demand. The timeliness of the proposed strategy is evaluated based on the number, value, transmission delay, and queuing delay of schedulable loads at each node.

INDEX TERMS Home energy management system, smart homes, cloud-edge collaboration, timeliness, demand-side resources.

I. INTRODUCTION

The uncertainty and variability associated with renewable energy will present challenges to the maintenance of adequate balancing reserves in the power system as the proportion of electricity consumption by residential users continues to rise rapidly [1], [2], and with the ongoing increase in the penetration rate of renewable energy sources. To address the aforementioned issues, home energy management systems (HEMSs) are now recognized as an effective method of controlling distributed energy and transferrable household appliance loads, resulting in a significant reduction in peak household energy demand [3], alleviating the challenges of insufficient renewable energy output. HEMS is in charge of managing household electrical equipment such as smart appliances, small-scale power generation systems, electric vehicles (EVs), and energy storage systems (ESS) [4], and ensuring the robust operation of each home device [5]. As a result, appropriate scheduling strategies must be contributed into effect to ensure household consumer participation while simultaneously accomplishing efficient scheduling of household load resources in terms of dispatchable resources and timeliness.

Demand response (DR) is an effective scheduling method to reduce peak energy consumption and improve operational efficiency and reliability [6], [7], [8]. Existing demand response methods can be categorized into two types: direct load control (DLC) and price-based techniques [9]. Dispatch centers can directly adjust the operation of certain customer equipment (e.g., air conditioning systems) through DLC mechanisms [10], as Advanced Metering Infrastructure (AMI)-driven HEMS transforms end-users from passive observers to active participants in energy management [11], [12]. Recently, many studies have focused on developing

The associate editor coordinating the review of this manuscript and approving it for publication was Maria Carmen Falvo.

new HEMS models that employ powerful optimization and management techniques to achieve cost-effectiveness for households [13]. Literature [14] performs HEMS user scheduling by exchanging energy and prices with upstream grid agents. The main objective of literature [15] and [16] is to propose a MAS-based HEMS that utilizes time-of-use (ToU) tariffs to trade with the local electricity market. In addition, the literature [17] manages energy received from distributed energy sources (DES) outside the utility grid based on the day-ahead ToU price. The Stackelberg game was developed in [18] to investigate real-time pricing scenarios for smart grids involving different retailers and residential customers. In [19], prices are transmitted and enforced by exchanging information with customers connected to the community controller, thus reducing the real-time nature of the proposed pricing scheme. Existing HEMS research focuses more on the participation of individual users in DR, but in practice, the large number of HEMS users participating in demand response generates a large amount of data when interacting with demand-side dispatchable resources [20], [21], and the large amount of data generates a large amount of bandwidth demand and computational power demand. Unreasonable bandwidth allocation will result in data transmission delays and waste of bandwidth resources, while insufficient computing power in the data center will result in queuing delays, both of which will affect the real-time data transmission [22], [23].

To address the above bandwidth issues, many researchers have proposed many approaches on bandwidth allocation as shown in Table 1. Literatures [24], [25], [26] aim at high accuracy rate and bandwidth allocation method based on joint learning can effectively handle wireless communication resources, but it has the disadvantages of long training time, high data transmission delay and insufficient communication resources [24]. Articles [27], [28], [29], [30] consider the objectives of high user bandwidth demand, low utilization and low user response rate, respectively, to achieve reasonable bandwidth allocation, but their accuracy is insufficient. Therefore, appropriate bandwidth allocation methods should be selected according to different application scenarios.

To address the above queuing delays, the article [31]proposes to perform all tasks including recognition, detection and recommendation in a data center through cloud computing to reduce the computational delay. However, cloud computing cannot meet the growing number of computing tasks. Literature [32]suggests that since edge servers have the advantages of proximity, low delay, and real-time access, some computation tasks can be scheduled to edge nodes to alleviate the computation pressure in cloud centers and reduce the transmission delays [33], [34]. However, edge servers have the disadvantage of smaller data processing and storage capacity, which cannot satisfy the computing tasks with larger data size. In summary, edge computing or cloud computing alone cannot meet the demand for reduced queuing delay. Since cloud-side collaboration improves the capability of the network in terms of data processing, information interaction, and real-time feedback, it will be a viable solution for reducing queuing delays [35], [36].

Recently, several articles have begun to focus on the time delay in the demand response process. The article [37] proposes a distribution network operation strategy based on cloud-side collaboration to improve the potential of distribution network operation. Instead of studying HEMS user response with cloud-side collaboration alone, this paper considers multiple delays to reflect the real situation. The article [38] constructs a cloud terminal collaboration mechanism

TABLE 1. Literature related to bandwidth allocation.

Reference	Challenge	Solution
[24]	Wireless resource allocation for multiple simultaneous federated learning (FL) services	Distributed bandwidth allocation algorithm
[25]	Training delay in wireless FL	Introducing dynamic quantization and bandwidth adaptation
[26]	Transmission delay and excessive communication resource usage	Dynamically assigns clients to edge servers based on client mobility
[27]	Bandwidth oversubscription	Adopt the ARIMA model to make the prediction
[28]	High bandwidth utilization of datacenters or low request delay of end-users	A novel Particle Swarm Optimization (PSO) based request allocation algorithm
[29]	Bandwidth allocation	Bandwidth allocation method based on Gaussian process regression
[30]	Sensing and communications	bandwidth allocation method for a general perceptive network

TABLE 2. Research drawbacks in previous literature and contribution of this paper.

Reference	Drawback	Contribution
[31]	Consider only computational delays	Transmission delay, computation delay and queuing delay are considered
[32]	Large-scale computing tasks not considered	Tackling massive computing tasks with cloud-side collaboration
[37]	Consider only cloud-side collaboration	Additional consideration of bandwidth
[38]	Insufficient bandwidth not considered	Consider allocation with limited bandwidth resources
[39]	Incomplete types of electrical loads	Consideration of multiple electrical loads

based on 5G and edge computing, targeting air conditioning and electric vehicle charging, and provides real-time data transmission services by satisfying bandwidth demand through 5G. However, there are only two of its response terminals, and there is no queuing delay in the actual simulation. This paper takes a large number of household loads as the response object, analyzes the impact of queuing delay on the response volume of dispatchable resources, and proposes corresponding strategies. The article [39] proposes a load clustering management and control strategy based on a cloud-based cooperative framework using air conditioners and electric water heaters as the objects of study. Similarly, it ignores the impact of transmission delay on scheduling. Compared to the above articles, this paper considers both transmission delay and queuing delay in the corresponding process of DR and proposes a corresponding strategy with bandwidth dynamic delay and offloading strategy, which supplements the research gap in this part.

In summary, the goal of this paper is how to minimize the loss of schedulable resources due to time delays and increase the amount of schedulable resources that a HEMS user actually provides during scheduling. The contributions of this paper compared to previous papers are reflected in Table 2, which we will summarize below:

1) According to the data volume size of HEMS schedulable resource information, ARIMA model [5], [27] is used to realize the dynamic allocation of bandwidth, reduce the transmission delay of HEMS user data in the DR phase, and set up multiple scheduling scenarios to select appropriate scheduling intervals.
2) Based on the cloud-edge collaboration framework, the offloading strategy is used to select a reasonable offloading path to reduce the queuing delay in the offloading process of computing tasks, taking into account the transmission delays among nodes, the edge nodes, and the computing resources in the cloud center.
3) Based on the above two points, the total time delay for each node to complete a single scheduling task is calculated, and the timeliness value actually provided by the HEMS users within each node in the demand response session is analysed.

II. METHODOLOGY

This paper describes a strategy for mitigating the issue of delayed energy management caused by HEMS users' extensive participation in DR. Fig.1 depicts the proposed strategy framework, which can effectively alleviate the temporal delay during the involvement of numerous HEMS users in DR.

The proposed strategy is based on a cloud-edge collaborative framework, which is a method for dealing with delays and large data transmissions from smart devices and sensors that leverages the strengths of both cloud and edge computing [40]. The framework is constructed from three parts: the cloud center, edge nodes (ENs), and terminals [41], [42]. The terminals are considered as Business Nodes (BNs) as they collect the schedulable resource information from the HEMS users within the node and upload it to the ENs. The cloud center receives the uploaded schedulable resource information from all the ENs and performs the overall scheduling of the schedulable resources.

The above is the scheduling model in the ideal state (no time delay). However, under time delay, each layer will generate corresponding resource allocation schemes for the time delay that may be caused by the scheduling mode in order to improve the timeliness of the schedulable resources for HEMS users. First, before the start of the next scheduling phase, the cloud center aggregates the bandwidth resources required by each node in the previous phase per unit of scheduling time and predicts the bandwidth resources required by each node per unit of time in the next scheduling phase using the ARIMA algorithm, and the bandwidth allocation scheme is distributed to each service node through the edge nodes to alleviate the transmission delays of each node in the next scheduling phase. Second, the computing resources of the edge nodes may not be able to satisfy the computing resources required by the business nodes under their jurisdiction during the scheduling period, so the cloud center will count the remaining computing resources of the cloud center and each edge node during the scheduling period, and determine an offloading strategy for the computing tasks based on the computing resources required by each computing task and the transmission delay between the communication links between the nodes to alleviate the queuing delay. Finally, the timeliness of the schedulable resources of each node is calculated based on the delay generated by the above two steps.

In summary, the proposed strategy aims to achieve timely scheduling of resources in HEMS by leveraging a cloud-edge collaborative framework and implementing a two-part implementation strategy to optimize data transmission and offload computation tasks.

A. HOME ENERGY MANAGEMENT SYSTEM

The HEMS that is proposed in this section can track all sorts of household appliances in real time. Each household appliance can be categorized into four categories: a temperature-controlled load (TCL), an interruptible load (IL), a transferable load (TL), or a fixed load (FL) [43]. An intelligent socket, as shown in Fig.2, is in charge of gathering information on electricity usage and dispensing electricity to home appliances. The information processing module connects the intelligent socket, system management center, and smart meter. The customer sends instructions for power consumption to the system management center, which uploads dispatchable resources. Based on the uploaded data, the higher-level system issues dispatch instructions. The smart meter determines the load scheme and obtains power from the distributed grid after receiving power consumption instructions from the information processing module.

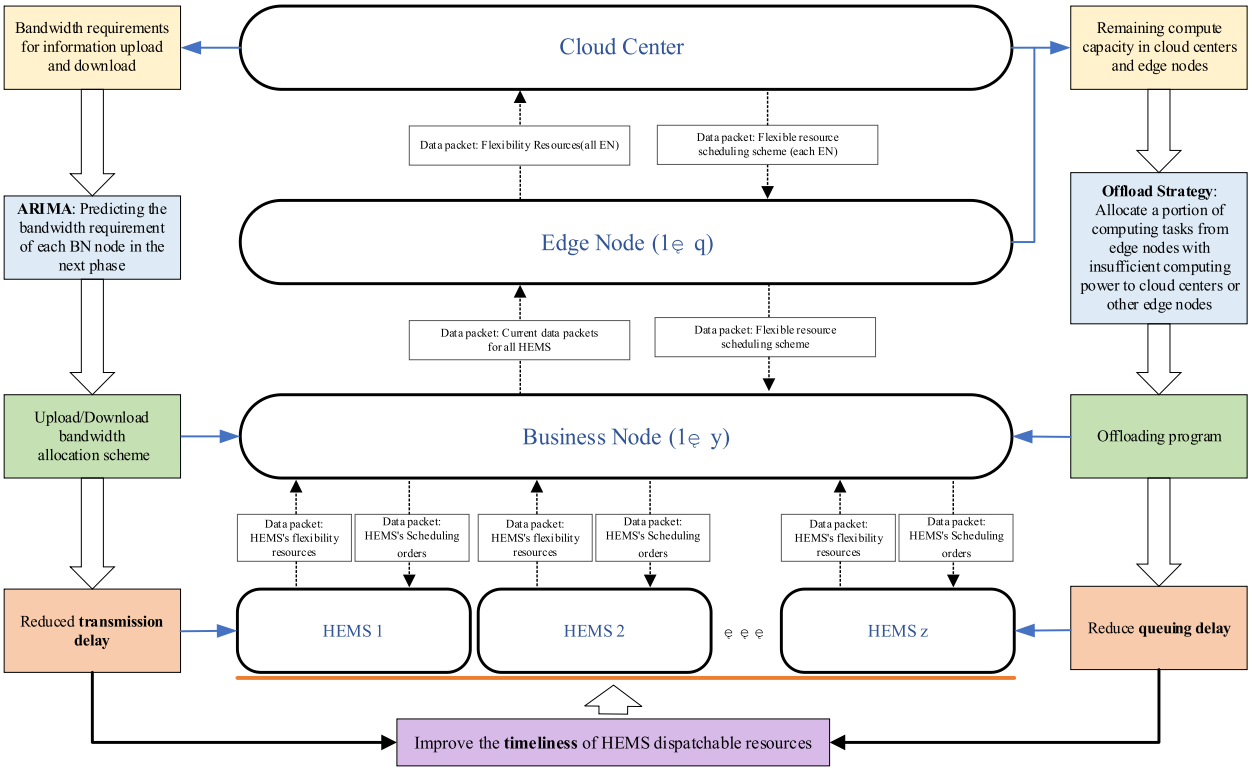

FIGURE 1. The proposed strategy aims to improve the timeliness of resource flexibility in HEMS.

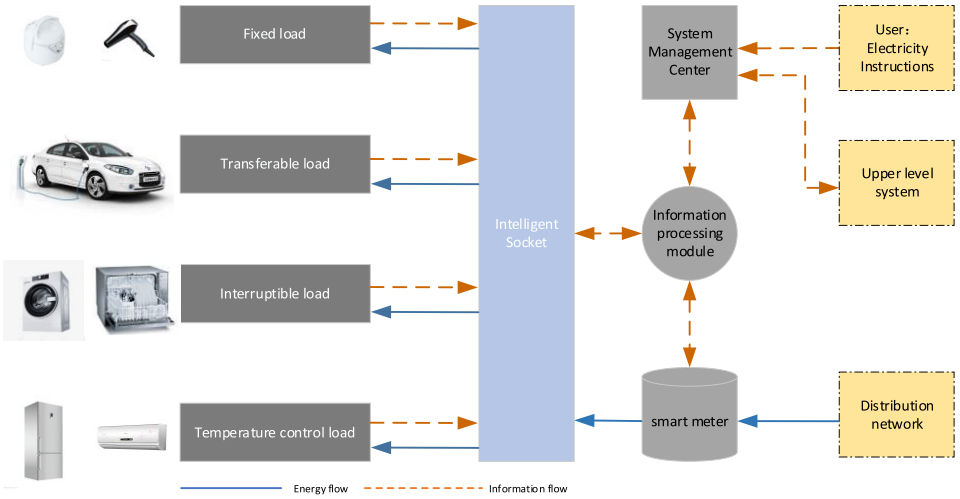

FIGURE 2. Management chart of home energy management system.

The objective function of the real-time phase domestic energy management problem is shown in Equation (1). In this case, washing machines, air conditioners, electric cars, and must-run services are denoted by the letters IL, TCL, TL, and FL. The goal of the HEMS within each node is to minimize its combined metrics through real-time scheduling. The objective of the HEMS is to maximize its dispatchable energy through real-time scheduling. The first part of the objective function denotes the cost of electricity purchased by the customer at time t. The second and third parts denote the impact of IL and TL devices being dispatched on the user's electricity experience and the impact of TCL devices being dispatched on the user's comfort, respectively, where N, M and K are the corresponding number of domestic TLs, ILs and TCLs, respectively. It is important to note that charging and discharging electric vehicle batteries frequently may decrease their lifespan [43]. As a consequence, in this paper, electric vehicles are regarded as TL rather than energy

TABLE 3. Notations.

T^{sch}	Scheduling time period
M	Interruptible load
N	Transferable load
K	Temperature controlled load
γ	Maximum delayed completion time of the device
Y	Business Nodes
Q	Edge Nodes
P	Cloud Center
P_t^{total}	Sum of dispatchable resources scheduled by HEMS at time t
$P_t^{n/m/k}$	Sum of dispatchable resources of TL/IL/TCL at time t
f	Price of electricity
$1_t^{n/m/k'}$	Switching status of TL/IL/TCL during the T^{sch}
$P_{\min}^{n/m/k}$	Minimum device power for TL/IL/TCL
$P_{\max}^{n/m/k}$	Maximum device power for TL/IL/TCL
1_t	Start/stop status of equipment at time t
$t_{\text{start/end}}$	User expectation on/state off
t_{total}	Rated duration of the device
1^{Γ}	Switching status of the device after scheduling
θ_t	Temperature at time t
θ_{en}	Ambient temperature at time t
D_k	Thermal resistance coefficient
G_k	Heat capacity
t_{on}	Length of running period
t_{off}	Length of downtime period
ω	Energy efficiency ratio
$\theta_{\min/\max}$	Set the upper and lower temperature limits
$Tem_k(t)$	Percentage of room temperature deviation from set point
t_{pre}	Predicted start time
B_{history}^{t}	Historical data before the start of the forecast
$B_{\text{next}}^{t_{pre}}$	Predicted bandwidth data
L	Lag operator
ϕ	The parameters of AR part of the model
η	The parameters of MA part
ε_t	An error term
ϑ	The number of parameters in the estimated ARIMA model
l	The maximum likehood for the model
$x_{\text{LE}}, x_{\text{OE}}, x_{\text{CD}}$	Binary variables
w_y	BN's task calculation volume，measured by CPU cycles
$f_{\min,y}$	Minimum computing power required for BN tasks
cpu_y	CPU resources required for BN tasks
mem_y	Memory resources needed for BN tasks
$data_y$	The amount of data to be transferred when BN tasks are offloaded
$t_{\max,y}$	Maximum time delay allowed to complete BN tasks

TABLE 3. *(Continued.)* Notations.

$T_y^{\text{LE/OE/CD}}$	BN task transfer to attributed ENs/offload to other ENs/cloud center node
$T_{\text{BN}\to\text{LE}}^{\text{td}}$	Transmission delay of BN task transmission to the attributed EN
$T_{\text{LE}}^{\text{cd}}$	Computational delay of affiliated EN q
f_q	Maximum computing power of EN q
cpu_q	Remaining available resources of EN q
mem_q	Available memory resources of EN q
$R_{q,y}$	Transmission rate between EN q and BN y
$T_{\text{BN}\to\text{OE}}^{\text{td}}$	Transmission delay of BN offload to other ENs
$T_{\text{OE}}^{\text{cd}}$	Offloading the computational delay of ENs
T^{queue}	Queuing delay
T_{front}^{cd}	The EN itself governs the computational delay of the BN
T_{front}^{td}	The EN itself governs the transmission delay of the BN
T_{behind}^{td}	Transmission delay of BNs offloaded to ENs
f_p	Maximum computing capacity of the cloud center
R_p	Data transfer speed between cloud center p and BN y
$T_{\text{Ed}\to\text{CD}}^{\text{td}}$	Transmission delay between cloud center p and BN y
T_{CD}^{cd}	Computational delay in the cloud center p
x_{yqp}	Binary variables of BN-EN-cloud
ϖ	Value factor of dispatchable resources
$\eta_{1,2,3}$	Weighting factor

storage devices. Specifically, FLs are household appliances that are essential to daily life and are exempt from scheduling.

$$\begin{aligned}\text{Max } C_z = \eta_1 \sum_{t=1}^{\xi} f_t \Big(\sum_{n=1}^{N} P_{t,z}^{n} \ell_t^n + \sum_{m=1}^{M} P_{t,z}^{m} \ell_t^m + \sum_{k=1}^{K} P_{t,z}^{k} \ell_{t'}^k\Big) \\ + \eta_2 \sum_{h=1}^{H} \frac{\sum_{t=1}^{\xi} \ell_{h,z}(t) T}{t_{\text{end}} - t_{\text{start}}} + \eta_3 \frac{\sum_{t=1}^{\xi} \left(\frac{e^{Tem_{k,z}(t)}-1}{e-1}\right) \ell_{t',z}^k}{\sum_{t=1}^{\xi} \ell_{t',z}^k}, \\ \forall t \in T^{sch}, H = N + M, \quad \forall z \in Z \end{aligned} \tag{1}$$

The following are the corresponding constraints of the three types (TL, IL, TCL) of domestic load. Equation (2) represents the maximum and minimum limitations of the energy of the three types of domestic load during the start time and end time of the device set by the home user.

$$\begin{cases} P_{\min}^{n} \le P_t^{n} \le P_{\max}^{n} \\ P_{\min}^{m} \le P_t^{m} \le P_{\max}^{m} \\ P_{\min}^{k} \le P_t^{k} \le P_{\max}^{k}, \end{cases} \quad \forall t \in [t_{\text{start}}, t_{\text{end}}] \tag{2}$$

The operating state of domestic appliance (TL, IL) is expressed by Equation (3). Note that ℓ_t is a binary variable.

If ℓ_t is equal to 1, the household appliances start working, if ℓ_t is equal to 0, the household appliances exit working.

$$\ell_t^{n/m} = \begin{cases} 0, & \text{other case} \\ 1, & t \in [t_{\text{start}}, t_{\text{end}}] \text{ and } t \in T^{sch} \end{cases} \tag{3}$$

The constraints governing the initiation of TL and IL are denoted by Equation (4). This expression stipulates that the temporal differential between the actual initiation of TL and IL and the user's power command time must not exceed the maximum delay time for device completion γ set by the user. Should the differential exceed the value specified by the user, this would imply that the power command cannot be executed within the user-specified timeframe.

$$\sum_{\Gamma=t_{\text{start}}}^{t_{\text{start}}+\gamma} \ell_{n/m}^{\Gamma} \geq 1, \quad 0 \leq \gamma \leq t_{\text{end}} - t_{\text{start}} - t_{\text{total}} \tag{4}$$

Since the IL can be turned off freely during operation without significant impact on the user's usage, the operational status during dispatching is as shown in Equation (5).

$$\sum_{t=\Gamma+1}^{\Gamma+T^{sch}} \ell_t^m \geq t_{\text{total}}, \quad \Gamma \in [t_{\text{start}} - 1, t_{\text{end}} - t_{\text{total}}] \tag{5}$$

Given that TL cannot be interrupted during operation, its running time must be continuous, and the total running time cannot be less than the rated working time of TL. The running state during scheduling is shown as Equation (6).

$$\sum_{t=\Gamma+1}^{\Gamma+T^{sch}} \ell_t^n \geq t_{\text{total}}(\ell_n^{\Gamma+1} - \ell_n^{\Gamma}), \ \Gamma \in [t_{\text{start}} - 1, t_{\text{end}} - t_{\text{total}}] \tag{6}$$

The operational status of TCL household appliances is determined by monitoring temperature changes in their service targets, as shown in Equation (7).The trend of temperature change is affected by ambient temperature $\theta_{en,t}$, the thermal capacity G_k and thermal resistance parameters D_k of the home, as well as the temperature of the device θ_t in the previous moment.

$$\begin{cases} \theta_{t+1} = \theta_{en,t} - P_t^k D_k - (\theta_{en,t} \\ -P_t^k D_k - \theta_t) e^{-(t_{\text{on}}+t_{\text{off}})/(D_k G_k)}, \forall t \in [t_{\text{start}}, t_{\text{end}}] \\ t_{\text{off}} = D_k G_k \ln\left(\dfrac{\theta_{en,t} - \theta_{\min}}{\theta_{en,t} - \theta_{\max}}\right) \\ t_{\text{on}} = D_k G_k \ln\left(\dfrac{D_k P_t^k \omega - \theta_{en,t} + \theta_{\max}}{D_k P_t^k \omega - \theta_{en,t} + \theta_{\min}}\right) \end{cases} \tag{7}$$

After determining the operational and downtime lengths of TCL devices during each time interval throughout the scheduling period, the overall operational state of the devices can be obtained as shown in Equation (8).

$$\ell_{t'}^k = \begin{cases} 0, & \text{other case} \\ 1, & t' \in [t, t + t_{on}], t \in [t_{\text{start}}, t_{\text{end}}] \cap T^{sch} \end{cases} \tag{8}$$

Specifically, Equation (9) expresses the temperature constraint of the TCL, and $\theta_{\min}$ and $\theta_{\max}$ are set by the user according to his own preferences.

$$\theta_{\min} \leq \theta_t \leq \theta_{\max}, \quad \forall t \in [t_{\text{start}}, t_{\text{end}}] \tag{9}$$

Equation (10) reflects the user's comfort level with respect to temperature, which is determined by the relationship between the actual temperature in the house and the temperature set point.

$$Tem_k(t) = \frac{2\left|\theta_t - \frac{\theta_{\max}+\theta_{\max}}{2}\right|}{\theta_{\max} - \theta_{\max}} \tag{10}$$

B. DYNAMIC BANDWIDTH ALLOCATION STRATEGY

The previous section described a scheduling model for three types of loads within a single household. However, each HEMS user's request sent during the scheduling period consumes one unit of bandwidth resources [28], i.e. $P_t^n, P_t^m, P_t^k =$ 1Mbps. We assume that the electricity usage data of each hems can be uploaded successfully. When a large number of users participate in DR, it will result in a significant demand for bandwidth, especially when bandwidth resources are limited, causing transmission delays. Therefore, considering bandwidth constraints is crucial when designing DR schemes for HEMS users. In this section, we focus on the bandwidth issue in dispatchable resource scheduling. Based on historical bandwidth demand data, we propose a time-series-based bandwidth prediction method to minimize transmission delay under limited bandwidth resources, which is of great significance.

Before formulating the bandwidth allocation model, the sampling interval T should be quantitatively analyzed. We periodically sample the bandwidth sequence at different time intervals for the time sequence $\{t_0, t_1, t_3, \ldots, t_\zeta\}$ where $t_\zeta - t_{\zeta-1} = T$, especially $T = \kappa\,\Delta t$, and κ is the length of historical Bandwidth Sequence B_{history}^t or predicted bandwidth demand values for the next phase B_{history}^t. Predicting the next bandwidth value $B_{\text{next}}^{t_{\text{pre}}} = \{b_{\text{next}}^{t_{\text{pre}}+\Delta t}, b_{\text{next}}^{t_{\text{pre}}+2\Delta t}, \ldots, b_{\text{next}}^{t_{\text{pre}}+k\Delta t}\}$ from the received historical bandwidth sequence $B_{\text{history}}^t = \{b_{\text{history}}^{t-(\kappa-1)\Delta t}, b_{\text{history}}^{t-(\kappa-2)\Delta t}, \ldots, b_{\text{history}}^t\}$, e.g. for prediction starting from the moment t_0, the solid line indicates the historical bandwidth sequence $\left[b_{\text{history}}^{t_0-(\kappa-1)\Delta t}, b_{\text{history}}^{t_0}\right]$ and the dashed line indicates the prediction sequence $\left[b_{\text{history}}^{t_0}, b_{\text{next}}^{t_0-\kappa\Delta t}\right]$. Forecasting for each time point during the scheduling period, we transform the time sequence $\{t_0, t_1, t_3, \ldots, t_\zeta\}$ into $\{t_0, t_0 + T, , \ldots, t_0 + (\zeta - 1)T\}$.

The ARIMA model was adopted to predict the bandwidth demand of node during the next scheduling period. The ARIMA model is denoted as ARIMA (AR_P, d, MA_q), ARP is the number of autoregressive (AR) terms, MA_q is the number of moving-average (MA) terms, and d is the number of differences made when the time series becomes smooth.

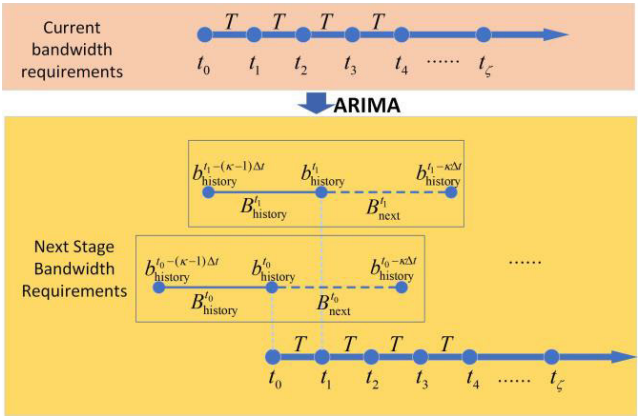

FIGURE 3. Flow of dynamic allocation of bandwidth resources.

The specific form of the ARIMA formula is as follows (11):

$$(1-\sum_{i=1}^{AR_p}\delta_i L_i)(1-L)^d X_t = \phi + (1+\sum_{i=1}^{MA_q}\varphi_i L_i)\lambda_t \tag{11}$$

The steps for solving the AR_P, d, MA_q parameters in the whole ARIMA process are shown below:

First of all, since the bandwidth data is transformed from the electricity consumption data of the users, the stability of the time series data should be checked, and this paper adopts the test of normality using the Kolmogorov-Smirnov (K-S) test, and the smoothness using the ADF test. If the examined data is non-stationary then the non-stationary data is transformed by means of differencing and this data series is examined again and if this data is still non-stationary then differencing is performed until the time series is transformed into a stationary series. When the input bandwidth data is transformed into smooth data, the number of differencing is taken as the value of d.

The next step is to explore the correlation between predicted and historical values using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) with values of [−1,1], when the value is 1 it means positive correlation and conversely it is negative correlation. With the already determined parameters d, the Akaike Information Criterion (AIC) is used to determine the most appropriate values of the parameters AR_P and MA_q. The process is shown in Equation (12):

$$\begin{cases} \mathrm{AIC} = 2\vartheta + \ln(l) \\ l \propto (\delta^2)^{-\frac{n}{2}} \exp\{-\dfrac{1}{2\delta^2}(\varepsilon_t)^2\} \\ \varepsilon_t = y_t - \sum_{i=1}^{p}\phi_i y_{t-1} - \sum_{i=1}^{q}\theta_i \varepsilon_{t-i} \end{cases} \tag{12}$$

Third, to verify the accuracy of the bandwidth predicted values, we apply a statistically based measurement average absolute percentage error (MAPE) to represent the deviation between the predicted and actual values. The process is known as Formula (13):

$$\mathrm{MAPE} = \frac{1}{\kappa}\sum_{t=1}^{\kappa}\left|\frac{B_{real}^t - B_{next}^t}{B_{real}^t}\right|, \quad \forall t \in T^{sch} \tag{13}$$

C. CLOUD-EDGE COLLABORATION-BASED OFFLOAD STRATEGY

In this section, we discuss the queuing delay caused by the lack of EN computational capacity, and to mitigate the queuing delay, we determine the offloading scheme by using the transmission delay between nodes determined in Section B. Compared with the cloud-edge collaboration framework, the offloading strategy based on cloud-edge collaboration is determined based on the task information sent by the BN to the EN, such as the size of data to be uploaded, the number of required CPU clock cycles, and the physical address of the task, as well as the resource information of the EN and the cloud center, and the result is sent to the BN. The purpose of using a cloud-edge collaborative offloading strategy is to reduce the flexibility loss caused by queuing delay, so energy optimization is not considered in this article, and the optimization goal is queuing delay. There are Y BNs that generate Y pending flexibility resource scheduling tasks $Task = \{BN_1, BN_2, \cdots, BN_y\}$, $Y = \{1,2,3,\cdots, y\}$, each with six basic attributes:

$$BN_y = \{w_y, f_{\min,y}, cpu_y, mem_y, data_y, t_{\max,y}\} \tag{14}$$

Since each EN and cloud center has limited computing resources, each flexibility resource scheduling task may be executed on the EN to which it is assigned or offloaded to other ENs or cloud center. The dispatchable resource scheduling task requests from BNs must be offloaded to the ENs or the cloud center in order to achieve the minimum total delay under the cloud-edge collaborative offloading policy. The offloading problem is more specifically expressed as Equation (15).

$$\begin{aligned} &\min \sum_{y=1}^{Y} x_{\mathrm{LE}} T_{y,t}^{\mathrm{LE}} + x_{\mathrm{OE}} T_{y,t}^{\mathrm{OE}} + x_{\mathrm{CD}} T_{y,t}^{\mathrm{CD}}, \forall t \in T^{sch} \\ &\text{s.t. } x_{\mathrm{LE}} + x_{\mathrm{OE}} + x_{\mathrm{CD}} = 1 \; x_{\mathrm{LE}}, x_{\mathrm{OE}}, x_{\mathrm{CD}} \in \{0, 1\} \end{aligned} \tag{15}$$

In the cloud-edge cooperative offloading strategy, the set of Q edge servers is $Edge = \{Ed_1, Ed_2, \cdots, Ed_q\}$, $Q = \{1, 2, 3, \cdots, q\}$, and the basic attributes of each edge server are shown in Equation (16).

$$EN_q = \{f_q, cpu_q, mem_q, R_{q,y}\} \tag{16}$$

When the dispatchable resource scheduling task is uploaded to the EN q to which it belongs, task y is executed at this EN if the EN's current remaining computing resources meet the computing resources required for the current task computation and the overall time delay T^{LE} is as shown in Equation (17).

$$T_t^{\mathrm{LE}} = T_{\mathrm{BN}\to\mathrm{LE},t}^{\mathrm{td}} + T_{\mathrm{LE},t}^{\mathrm{cd}}, \forall t \in T^{sch}$$

$$\text{s.t.} \begin{cases} cpu_{y,t} \le cpu_{q,t} \\ mem_{y,t} \le mem_{q,t} \\ f_{q,t} > f_{\min,y,t} \\ T_t^{\text{LE}} < t_{\max,y,t}, \end{cases} \quad \forall \text{y} \in \text{Y}, \forall q \in Q, \forall t \in T^{sch} \quad (17)$$

Equation (18) depicts the transmission delay from the BN to the EN to which it belongs. Each BN selects the link with the shortest transmission delay between the BN and the EN.

$$T^{\text{td}}_{\text{BN}\to\text{LE},t} = \frac{data_{y,t}}{R_{q,y,t}}, \quad \forall \text{y} \in \text{Y}, \forall q \in Q, \ \forall t \in T^{sch} \quad (18)$$

Transmission speed $R_{q,y}$ is obtained from Shannon's formula as Equation (18), where $B^t_{next,q,y}$ is the bandwidth between channels, p_l denotes the transmit power of the mobile device, n_0 denotes the noise parameter.

$$R_{q,y,t} = B^t_{next,q,y} \log_2 (1+\frac{p_l}{n_0 B^t_{next,q,y}}), \quad \forall \text{y} \in \text{Y}, \forall q \in Q, \ \forall t \in T^{sch} \quad (19)$$

The computational delay of the EN is shown in Equation (20).

$$T^{\text{cd}}_{\text{LE},t} = \frac{w_{y,t}}{f'_{q,t}}, \quad \forall q \in Q, \ \forall t \in T^{sch} \quad (20)$$

where the computing power of the ENs f'_q will decrease with the number of BN tasks processed.

$$f'_{q,t} = f_{q,t} - f_{\min,y,t}, \quad \forall \text{y} \in \text{Y}, \forall q \in Q, \ \forall t \in T^{sch} \quad (21)$$

When the dispatchable resource scheduling task y arrives at EN q, if the current remaining resources of the EN it belongs to are less than the resources used for task computation. Task q will be offloaded to other ENs and will not participate in the computation until the other ENs have executed the task with higher priority, and its completion time is given by Equation (22). The established task execution matrix can be used to calculate the queuing time Equation (23).

$$T_t^{\text{OE}} = T^{\text{td}}_{\text{BN}\to\text{OE},t} + T^{\text{cd}}_{\text{OE},t} + T_t^{queue}, \quad \forall t \in T^{sch} \quad (22)$$

$$T_t^{queue} = T^{\text{cd}}_{front,t} + \left| T^{\text{td}}_{front,t} - T^{\text{td}}_{behind,t} \right|, \quad \forall t \in T^{sch} \quad (23)$$

where the principle of the two parts $T^{\text{td}}_{\text{BN}\to\text{OE}}$, $T^{\text{cd}}_{\text{OE}}$ in Equation (22) is consistent with $T^{\text{td}}_{\text{BN}\to\text{LE}}$, $T^{\text{cd}}_{\text{LE}}$ in Equation (23). The queuing delay has two components: the first is the BN that was initially contained in the EN, and the second is the priority between the tasks of the initial BN and the tasks of the offloaded BN.

The cloud-edge collaborative offload policy includes a remote cloud center, and because the cloud center's CPU and memory resources are much greater than those of the ENs, only two cloud center attributes are considered in this paper, as shown in the following Equation (24).

$$C_p = \{f_p, R_p\}, \quad \forall k \in P \quad (24)$$

When all ENs' computing resources don't really encounter the computing requirements of the dispatchable resource scheduling task to be offloaded, the delay of task offloading to the cloud center is Equation (25), Equation (26) represents the time of task offloading from the mobile device to the cloud center, and Equation (27) represents the execution time of task in the cloud center.

$$\begin{aligned} T_t^{\text{CD}} &= T^{\text{td}}_{\text{EN}\to\text{CD},t} + T^{\text{cd}}_{\text{CD},t} + T^{\text{td}}_{\text{BN}\to\text{LE},t}, \\ EN \in Edge &= \{Ed_1, Ed_2, \cdots, Ed_q\}, \quad \forall t \in T^{sch} \\ \text{s.t. } T_t^{\text{CD}} &< t_{\max,y,t} \end{aligned} \quad (25)$$

$$T^{\text{td}}_{\text{Ed}\to\text{CD},t} = \frac{data_{y,t}}{R_{p,t}}, \quad \forall \text{y} \in \text{Y}, \forall p \in P, \ \forall t \in T^{sch} \quad (26)$$

$$T^{\text{cd}}_{\text{CD},t} = \frac{w_{y,t}}{f_{p,t}}, \quad \forall \text{y} \in \text{Y}, \forall p \in P, \ \forall t \in T^{sch} \quad (27)$$

In this case, the transmission rate is constant regardless of the bandwidth value because information is exchanged between the ENs and the cloud center via wired transmission.

D. EDGE NODE DEPLOYMENT MODEL

Since, the different affiliations between BNs and ENs will affect the offload scheme, this section will discuss this issue. The transmission delays and queuing delays obtained through Part B and Part C are used to determine the deployment of the EN. Equation (28) represents the sum of wired transmission delay $T^{\text{td}}_{\text{BN}\to\text{LE}}$ and wireless transmission delay $T^{\text{td}}_{\text{Ed}\to\text{CD}}$ between the BN, EN, and cloud center. As seen in Equation (28), x_{yqp} is a binary variable representing the diversity of inter-level affiliation problems between cloud edge collaboration frameworks, especially between ENs and BNs.

$$\min \sum_{y=1}^{Y} \sum_{q=1}^{Q} \sum_{p=1}^{P} x_{yqp} \Big(\sum_{t=1}^{T^{sch}/T} (T^{\text{td}}_{\text{BN}\to\text{LE},t} + T^{\text{td}}_{\text{Ed}\to\text{CD},t})/t \Big), \quad x_{yqp} \in \{0,1\}, \quad \forall t \in T^{sch} \quad (28)$$

where if x_{yqp} is equal to 1, it means that BN y sends information to the cloud center p through EN q.

The constraints are used to avoid one BN sending information to multiple ENs or one EN sending information to multiple cloud centers by Equation (28) to (33). Equation (29) and Equation (30) indicate that there is only one link for BNs to receive scheduling tasks and send messages from the cloud center, respectively.

$$\sum_{q \in \Delta^+_{(p)}} \sum_{y \in Y} x_{yqp} = 1, \quad \forall p \in P, \ \forall q \in Q \quad (29)$$

$$\sum_{p \in \Delta^-_{(P)}} \sum_{y \in Y} x_{yqp} = 1, \quad \forall p \in P, \ \forall q \in Q \quad (30)$$

where $\Delta^+_{(\cdot)}$ denotes the set of links emanating from node, $\Delta^-_{(\cdot)}$ denotes the set of links from back to node. Equation (31) represents the same link between the cloud center p and the EN q for information upload and download.

$$\sum_{p \in \Delta^-_{(q)}} x_{yqp} - \sum_{p \in \Delta^+_{(q)}} x_{ypq} = 0, \quad \forall y \in Y, \ \forall q \in Q \quad (31)$$

Equation (32) represents the same link between the EN q and the BN y for information upload and download.

$$\sum_{p\in\Delta^{-}_{(y)}} x_{yqp} - \sum_{q\in\Delta^{+}_{(y)}} x_{qyp} = 0, \quad \forall y \in Y,\ \forall q \in Q \tag{32}$$

Equation (33) Indicates that ENs upload and download information from the same cloud center, where $\Delta^{+}_{(\cdot)} \cup \Delta^{-}_{(\cdot)} = \Delta(\cdot)$. In other words, the EN q is belonging to a cloud center p.

$$\sum_{p\in\Delta(q\in Q)} x_{0qp} = 1, \quad \forall y \in Y \tag{33}$$

Equation (34) Indicates that the BN y will only belong to one EN q.

$$\sum_{q\in\Delta(y\in Y)} x_{yq0} = 1, \quad \forall y \in Y \tag{34}$$

III. CASE STUDY

We use MATLAB 2021b for simulation, the hardware used is Legion R7000P, AMD Ryzen 5 5600H with Radeon Graphics. We assume that the user's power usage habits as well as the uploading and downloading of the data from each node within the Cloud Edge Collaboration framework are successful.

In this paper, 6:00 p.m. to 8:00 p.m. is selected as the scheduling time period, and the related parameter settings as well as the reasons for the selection are shown in the Appendix A.

A. FLEXIBILITY RESOURCES OF HEMS

Fig. 4 illustrates the dispatchable resources (TL, IL, and TCL) generated by all nodes that are received by the cloud node during the scheduling period. The dispatchable demand gap of the power distribution network occurs from 6 pm to 8 pm. The start time of household electrical devices (TCL and IL) is before 6 pm (i.e., the scheduling start time), but their operating time is within the scheduling period, which leads to a large amount of data in the early scheduling stage (0-10 minutes) despite the increase in dispatchable resources. The reason why the dispatchable resources of TL devices are relatively lower than TCL and IL in the early scheduling stage is that these devices cannot be scheduled during operation. For TCL devices, the external temperature values are shown in Table 4. Taking node 38 as an example, the air conditioner's shutdown time during the scheduling period is shown in Fig. 5, where the red colour indicates that the air conditioner is not in operation. The shutdown time represents the reduced operating time of the HEMS user's air conditioner under the premise of not affecting the user experience. Fig. 6 shows the dispatchable resources of all nodes' EV users, with the charging power of EV being 52 kWh, and the charging time from 0% to 100% battery capacity is 180 minutes. Fig. 7 shows the dispatchable resources of all nodes' washing machine (WM) users, with the operating power of 0.39 kWh and operating time of 60 minutes. Single user scheduling results will be reflected in Appendix B.

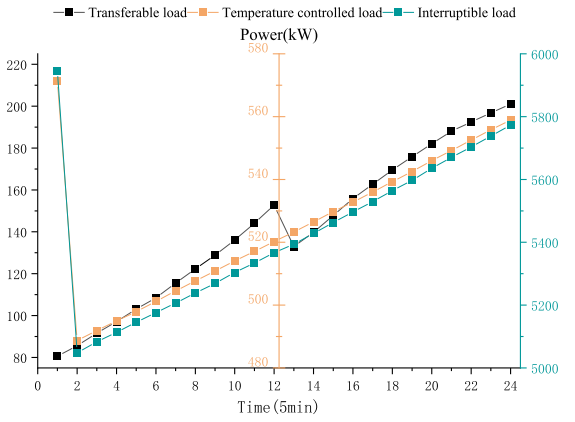

FIGURE 4. The maximum dispatchable household load within all nodes.

TABLE 4. Ambient temperature.

Sampling interval	1	2	3	4	5	6	7	8
Temperature (℃)	31.5	31.4	32.3	32.2	32.6	33.1	33.0	33.5
Sampling interval	9	10	11	12	13	14	15	16
Temperature (℃)	33.6	34.1	34.5	34.7	35.6	36.1	35.7	35.1
Sampling interval	17	18	19	20	21	22	23	24
Temperature (℃)	35.4	34.8	34.2	33.6	33.5	33.5	33.4	33.3

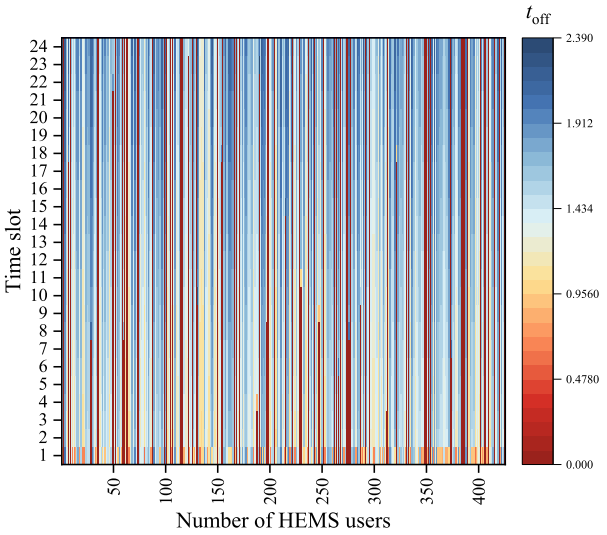

FIGURE 5. Air conditioning downtime for HEMS users within node 38.

B. BANDWIDTH ALLOCATION SCHEME (MULTIPLE TIME SCALES)

Appendix C shows the bandwidth prediction results for different scheduling intervals. The evaluation method for

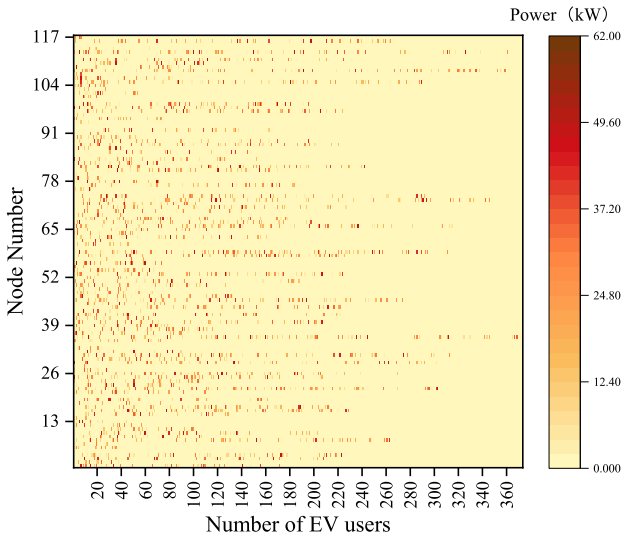

FIGURE 6. EV user dispatchable resources within all node.

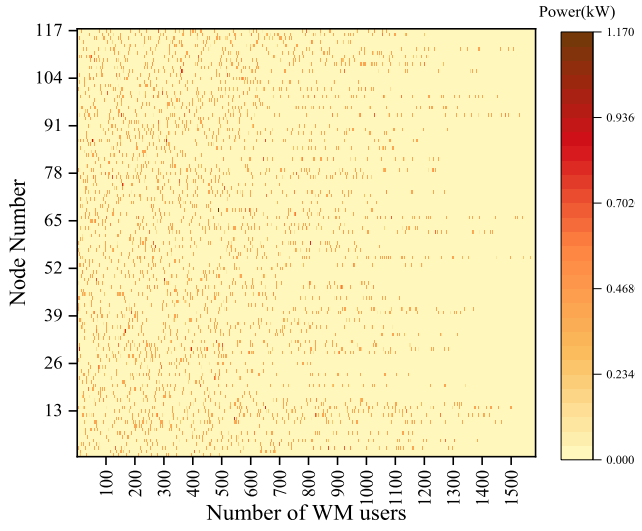

FIGURE 7. WM user dispatchable resources within all node.

bandwidth utilization is consistent with [29]. Fig. 8 shows a graph of the predicted value of bandwidth versus the actual value of bandwidth at node 117 in the second scheduling time slot. According to the adopted algorithm, the trend of the curve of the predicted value of bandwidth of each node in the second scheduling time slot is consistent with the trend of the actual value, and most of the nodes' predicted values are accurate with the actual value, the error rate is within the interval of 6%∼10%. The bandwidth allocation scheme based on predicted values within the allowed bandwidth range during scheduling is illustrated in Fig. 9. Fig. 10 describes the allocation scheme under fixed bandwidth, which results in three possible scenarios. The first scenario is when the bandwidth utilization is relatively high. In Fig. 12, the highest bandwidth utilization rate is 1.845, which is caused by a large amount of dispatchable data generated at the node during dispatchable scheduling. The fixed bandwidth allocation

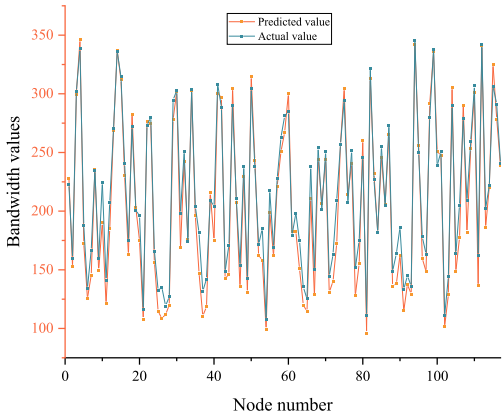

FIGURE 8. Bandwidth error during second dispatch.

scheme cannot meet the actual demand, which exacerbates transmission delay. The second scenario is when the bandwidth utilization rate is relatively low. The lowest bandwidth utilization rate in Fig.12 is 0.325, indicating that the node has been allocated too much bandwidth, which does not reduce transmission delay as the node's data volume remains constant, but only wastes bandwidth resources. The third scenario is when the bandwidth is fully utilized (defined as a bandwidth utilization rate within the range of 0.9 to 1.0). Fig. 11 depicts the bandwidth utilization rate of each node after implementing the bandwidth allocation scheme. It can be seen that the bandwidth utilization rate range decreased from 0.325-1.845 to 0.9-1.102, indicating that the nodes made full use of bandwidth resources during scheduling. As shown in Fig. 13, taking the second scheduling moment as an example, after the dynamic allocation of bandwidth, the surplus bandwidth resources of the nodes with low bandwidth utilization are allocated to the nodes with high bandwidth utilization, which reduces the transmission delay (the maximum transmission delay is reduced from 1.0s to 0.2s), especially for the nodes with surplus bandwidth resources, whose bandwidth allocation values can still satisfy the data transmission without increasing transmission delay requirements of the task. This strategy ensures that the transmission delay of the scheduling task information of each node is minimized while increasing the bandwidth utilization.

C. SOLUTIONS FOR OFFLOADING

This section selects a reasonable offloading strategy to mitigate the queuing delay based on the transmission delay of each node during the scheduling period obtained in the previous section. The Cloud Edge Collaboration Framework edge nodes and their governing node schemes are shown in Appendix D. To alleviate queuing delays caused by insufficient computing resources at ENs, we adopt an offloading strategy based on cloud-edge collaboration framework, and the partial results of the offloading are shown in Fig.14, while the overall offloading results can be found in Appendix B1. Specifically, we first determine the BNs that need to be offloaded, and define them as the offload nodes. Specifically,

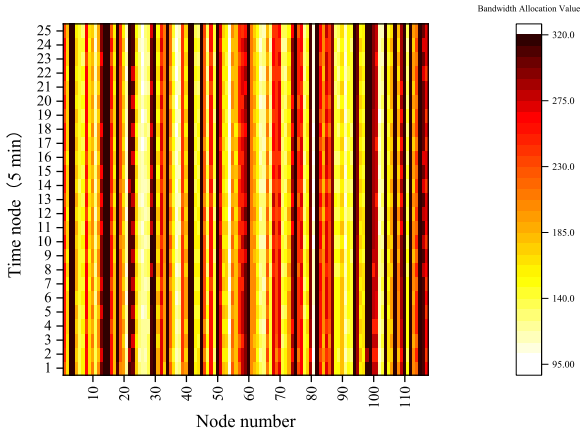

FIGURE 9. Bandwidth allocation scheme under the allowed range of bandwidth.

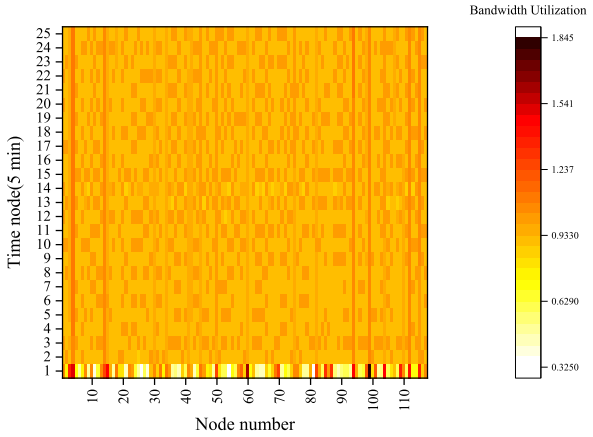

FIGURE 12. Bandwidth utilization of fixed bandwidth.

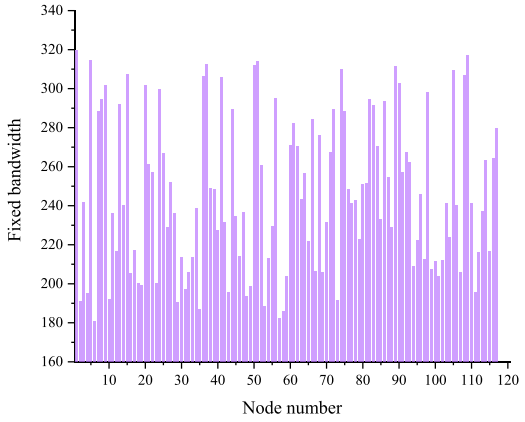

FIGURE 10. Statistics of the fixed bandwidth.

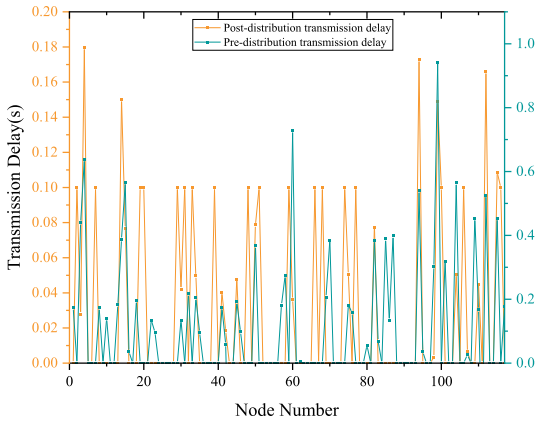

FIGURE 13. Transmission delay in the second scheduling interval.

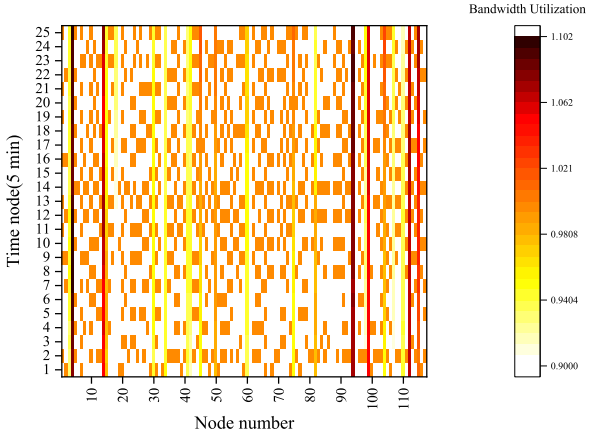

FIGURE 11. Bandwidth utilization of the allocation scheme.

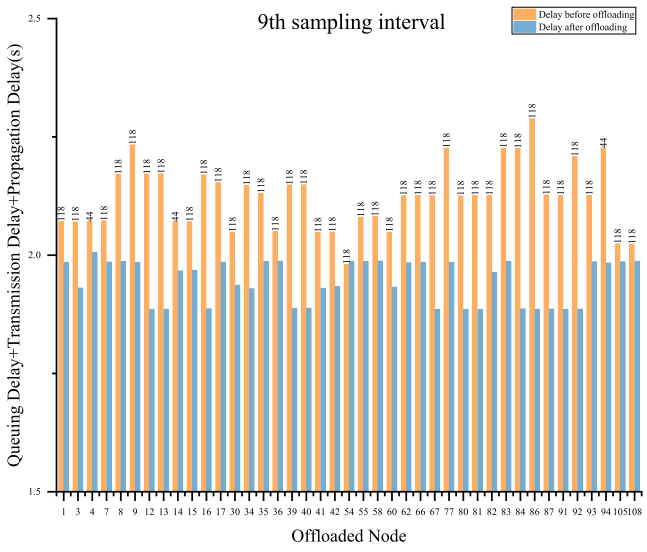

FIGURE 14. Comparison of delay before and after offloading.

as the computing resources of the cloud center (128 GB) are much higher than those of the ENs (10 GB) [40], we prioritize the allocation of the offloaded BNs to the cloud center. When the computing resources of the cloud center are exhausted, the remaining offloaded BNs are allocated to the ENs with the minimum propagation delay and sufficient computing resources. As shown in Fig.14, most of the offloaded BNs are allocated to the cloud center, with node 118 designated as the cloud center. Due to relatively sufficient computing resources at ENs 44 and 46 during the scheduling

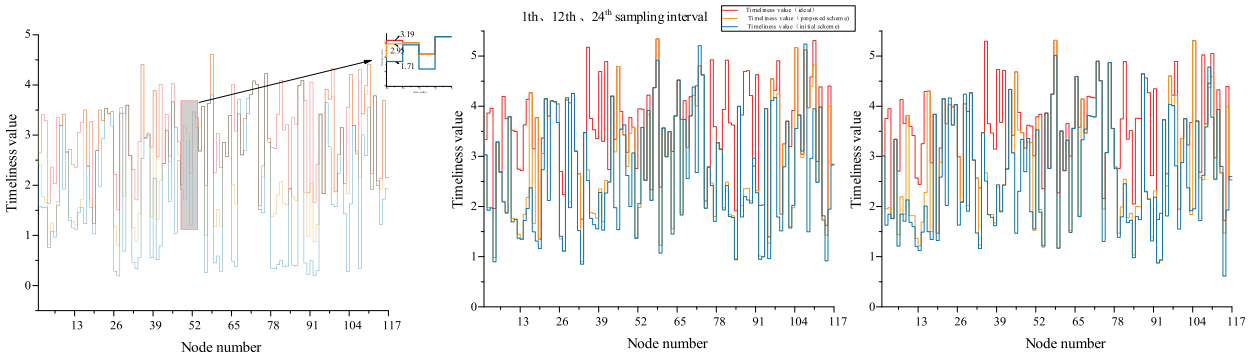

FIGURE 15. Comparing the timeliness value of dispatchable resources in multiple nodes across multiple time periods and scenarios.

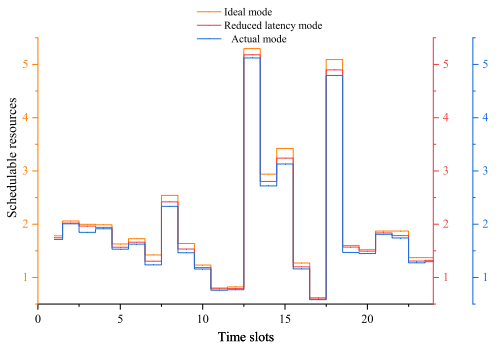

FIGURE 16. Comparison chart of schedulable resources.

process, the remaining offloaded BNs are allocated to these two ENs. Fig.14 shows the comparison of delay before and after offloading. The orange bars represent the delay before offloading, and the blue bars represent the delay after offloading. Among other things, after offloading, node 84 reduces the time to complete an upload from 2.25s to 1.83s for schedulable resources within the node. The reduction in total delay is particularly significant for the BNs offloaded to the cloud center. The propagation distance between nodes will be detailed in Appendix E2. The overall time delay is greatly improved by the offloading strategy based on cloud-side collaboration as shown in Appendix E2, which will facilitate the subsequent timeliness analysis of schedulable resources.

D. THE TIMELINESS VALUE OF DISPATCHABLE RESOURCES

Fig. 15 shows a comparison of the timeliness value of dispatchable resources in different nodes during the period of dispatchable resource scheduling in three different time periods. We have set up three different scenarios, namely the ideal state, the proposed method, and the initial method (fixed bandwidth, without offloading strategy). As the setting of wind and solar power stations varies among nodes, and the dispatchable energy in each node can directly alleviate the shortage of dispatchable resources, the value of dispatchable energy in such nodes is set to 0.9-1. For nodes with only one type of new energy power station, the value of dispatchable energy is set to 0.8-0.9, while for nodes without any new energy power station, it is set to 0.7-0.8. The dispatchable resource timeliness value is shown in the following Equation (35) [38]:

$$Value_i(t) = \varpi_i \log[T^{\mathrm{LE}}(t) + T^{\mathrm{OE}}(t) + T^{\mathrm{CD}}(t)], \ \forall i \in Y \cup Q \tag{35}$$

Compared with the initial approach, the proposed method has increased the overall timeliness value of dispatchable resources by 19.2%, 17.2%, and 18.6% at the 1st, 12th, and 24th stages respectively. In the 1st stage, 52 nodes are ENs, and there is no transmission delay or propagation delay in the data transmission process, so these nodes do not need to consider bandwidth and computing resource issues. However, for nodes 49, 50, and 51, the proposed method significantly improves the timeliness value compared to the initial approach. For example, in the initial approach, the timeliness value of dispatchable resources within node 49 was only 1.71, while after implementing the proposed method, the timeliness value increased to 2.95, representing a 72.5% improvement. The overall timeliness value results can be found in Appendix F. Improvement in the timeliness of schedulable resources means that users can respond faster to the aggregator's scheduling commands, increasing the amount of actual schedulable resources for a single user, which means more schedulable resources for the aggregator and more revenue for the user, as shown in Figure 16.

The above analysis shows that the proposed framework can effectively improve the timeliness of dispatchable resources in each node, but it still has some limitations. For example, this paper only considers the lack of new energy during the scheduling period, which reduces the power demand of users, but fails to consider the situation when the new energy output is too much, which increases the power demand of users; for the overall scheduling, the relevant scheduling model should be constructed according to the size of the time delay; in terms of the bandwidth prediction, only the ARIMA model is considered, which fails to compare the prediction

with the existing algorithms such as the machine learning algorithm and so on, which makes the prediction results more convincing.

IV. CONCLUSION

To improve the timeliness of dispatchable resources in home energy management systems, this paper proposes a strategy. This strategy reduces the delay in the transmission of schedulable resource information to the scheduling center, which will help to improve the timeliness of schedulable resources within each node as well as the amount of schedulable resources for each home user. In dispatchable resource management, there are two main timeliness issues: transmission delay and queuing delay. To address these issues, two solutions are proposed in this paper:

First, for transmission delay, this paper uses a differential integral moving average autoregressive model based on historical bandwidth demand data to identify the optimal solution for bandwidth allocation. This approach optimizes bandwidth utilization efficiency, thereby improving information transmission speed. We stabilized the bandwidth utilization rate of each node during scheduling within the range of 0.9 to 1.102, avoiding high transmission delay caused by high bandwidth utilization rates and excessive idle bandwidth caused by low bandwidth utilization. And the transmission delay of the node is greatly alleviated the highest transmission delay from 1s to 0.18s, which will greatly improve the timeliness of data transmission and make effective support for the determination of the subsequent offloading strategy.

Second, for queuing delay, this paper adopts a cloud-edge collaborative framework and an offloading strategy. Specifically, this paper offloads computing tasks from ENs to cloud nodes and ENs with assigned computing resources to improve resource utilization and minimize queuing delay. In this part, the transmission delay between each part determined in the above part is used as a basis and combined with the counting of the remaining computing resources of each edge node for decision making, as seen in the algorithmic analysis, its overall delay rushes down from 2.25s to 1.83s. the decrease of the overall time delay implies that the amount of resources that can actually be dispatched by the scheduling center will be improved.

To verify the effectiveness of the proposed strategy, experiments were conducted that considered factors such as the number and value of dispatchable resources, transmission delay, and queuing delay of nodes. The experimental results show that the proposed strategy can significantly improve the timeliness of dispatchable resources in home energy management systems, with a 19.2%, 17.2%, and 18.6% improvement in timeliness value observed in three different time periods before, during, and after the experiment. And three simulation models are analyzed to analyze the actual schedulable resources of a single household, and the adoption of the scheme proposed in this paper will effectively improve the schedulable resources of a single user.

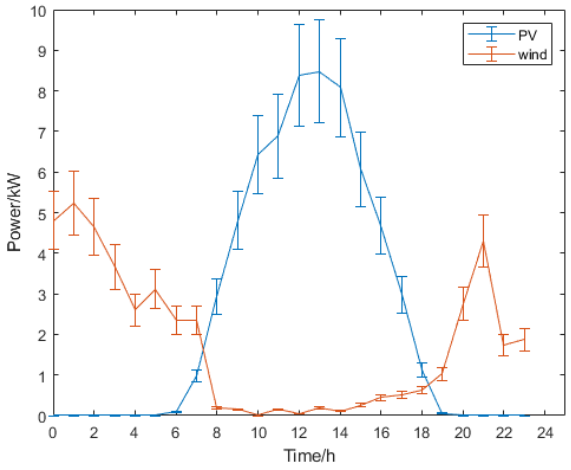

FIGURE 17. Day-ahead output of 6kW PV and 10kW wind power.

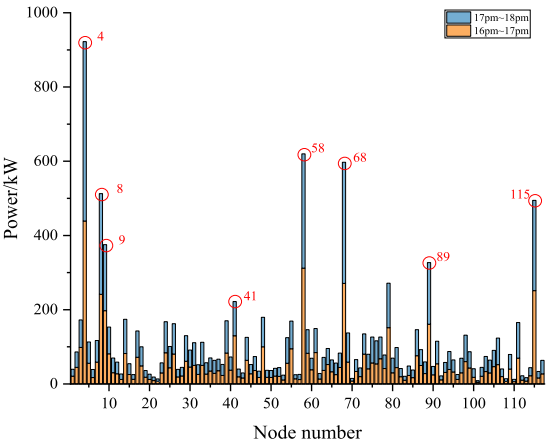

FIGURE 18. Total load of 117 nodes at 16pm~18pm.

TABLE 5. The locations of nodes for the installation of PV and wind power plants.

Node of PV power plants	Node of Wind power plants
24,62,65,66,69,70,72,73,74,7 6,77,80,82,85,87,89,90,91,92 ,99,100,103,104,105,107,110 ,111,112,116	34,36,40,42,46,49,54,55,56,59,61,77,78,7 9,80,85,86,87,88,89,90,91,92,95,96,97,98 ,99,100,101,102,103,104,105,106,107,10 8,109,110,111,112

The determination of the actual value of dispatchable resources will help the scheduling center to be more reasonable in generating the scheduling plan, avoiding the actual response value to be much lower than the pre-set value of the scheduling plan, and the actual benefit of the user will be enhanced.

However, this paper only considers the impact of time delay on the timeliness value of each node, but ignores the scheduling strategy of time delay on the schedulable resources of each node in the demand response process, which is what we will study subsequently. In addition, in the cloud-edge collaboration framework, edge computing can process the data uploaded by each BN to reduce bad data

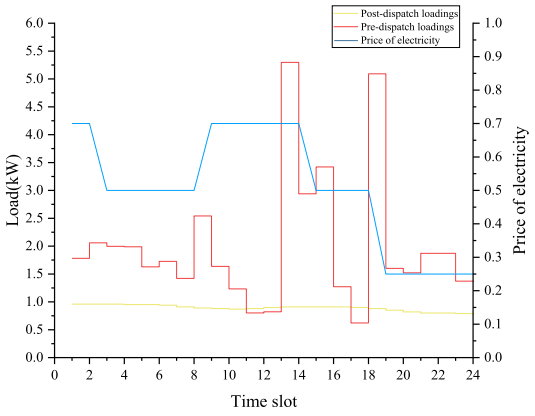

FIGURE 19. Single user scheduling results.

and unsuccessfully sent data, etc. to relieve the computational pressure on the cloud center, but this is not consistent with what is discussed in this paper, so it is not reflected in the paper. These are the directions of our subsequent research, and we will improve these studies to better enhance the real-time response of schedulable resources and the integrity of the scheduling and communication process.

APPENDIX

A. SCENARIO DESCRIPTION

The IEEE 118 node system diagram has been employed in this article, alongside node 1 representing the cloud node. A time range of 2 hours from 6 pm to 8 pm is used to test the performance of the proposed model. Fig. 17 depicts the electricity production statistics of individual photovoltaic and wind power plants, with a 15% margin of error. Domestic photovoltaic and wind power stations have capacities of 0.6 MW and 0.8 MW, respectively, and these values are converted to kW for ease of analysis. Table 5 shows the node locations within the station area where photovoltaic and wind power stations are installed. Fig. 18 depicts the 117 nodes' load conditions between 6 and 8 p.m. It can be deduced that during the two hours after households return home around 6 p.m., such as nodes 4, 8, 9, 41, 58, and 68, etc., the output of new energy remains constant. The overall load will far exceed the new energy power station's output limit, resulting in a significant power shortfall.

B. SINGLE USER SCHEDULING RESULTS

Figure 19 shows the load profile of a single user before and after participating in the dispatch, and it can be seen that the load profile decreases significantly and is regionally smooth. The customer's comfort level for temperature is 0.60417, the comfort level for electricity consumption is 0.34259, and the total price of electricity during the scheduling period is 19.942.

C. SCHEDULING INTERVAL SELECTION

Results of the proposed method with diverse time scales of bandwidth allocation scheme are presented in Figs. 20-21.

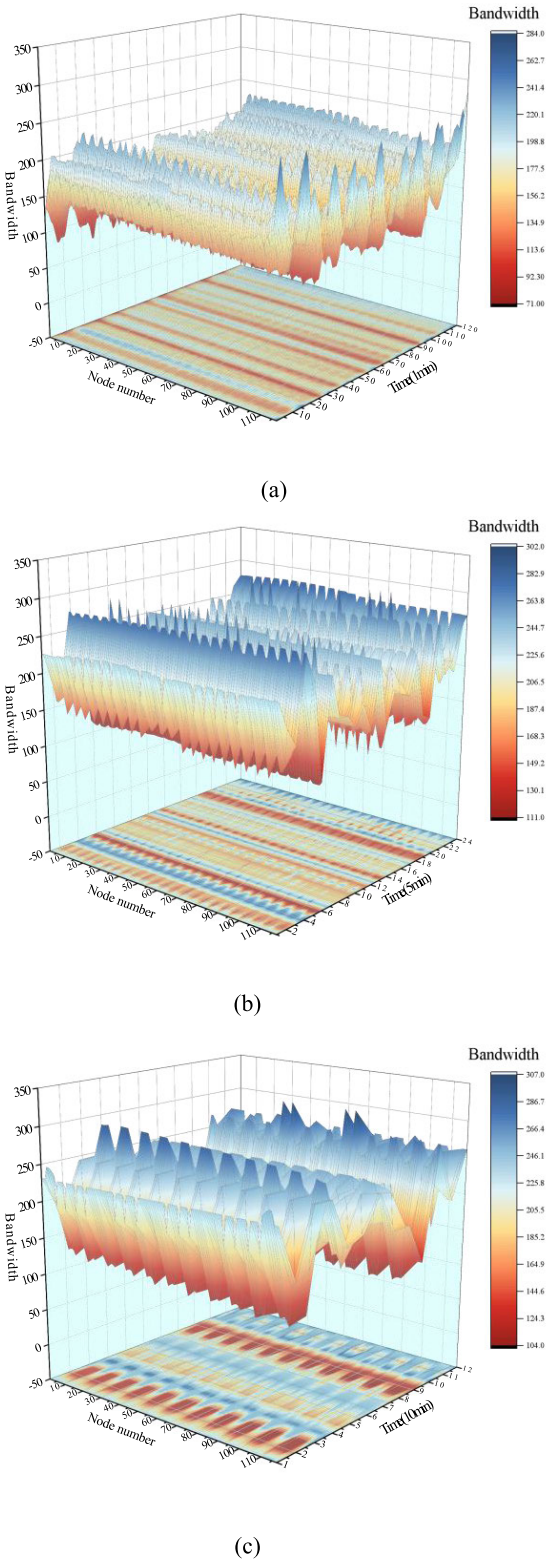

FIGURE 20. Predicted bandwidth of each node under diverse sampling interval.

The Fig (a), (b), and (c) in Fig. 20 represent the bandwidth allocation results for different nodes during a 2-hour

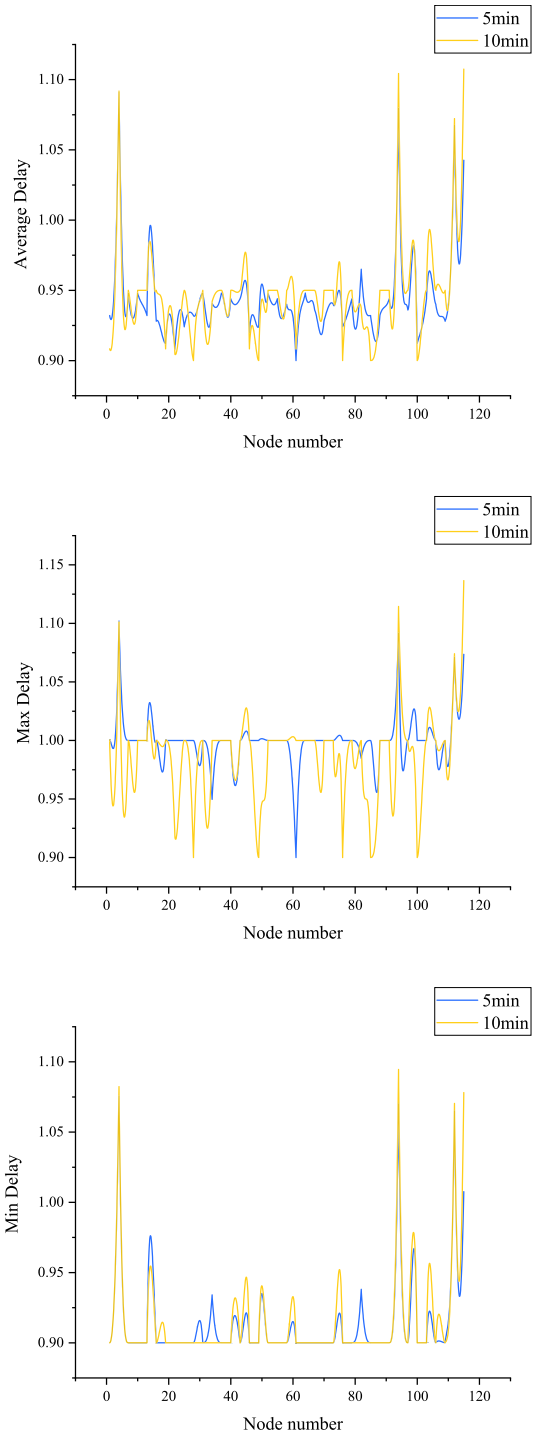

FIGURE 21. 5 minute and 10 minute sampling periods: Average, Max and Min delay of each node during the simulation.

scheduling period with sampling periods of 1 minute, 5 minutes, and 10 minutes, respectively, where the bandwidth demand of each node is derived from the data volume of the hems dispatchable resources. During the 1-minute sampling interval, some predicted bandwidth values fall below the lower limit of the allowable bandwidth range, resulting in actual bandwidth allocation significantly lower than the true

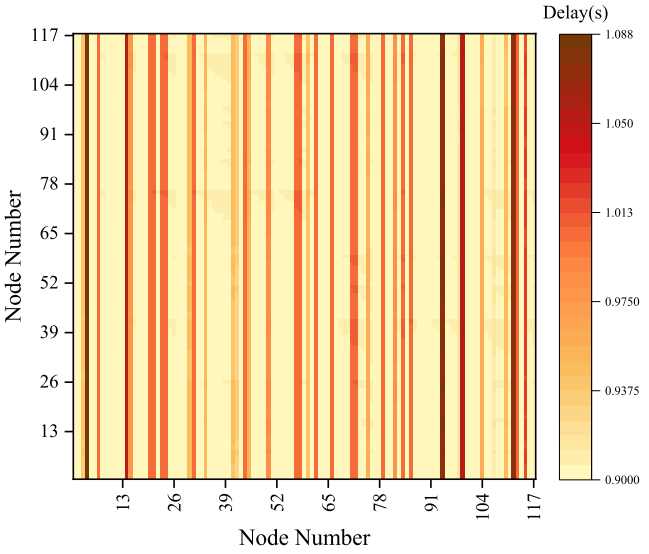

FIGURE 22. Inter-node transmission delay and propagation delay.

values of the nodes, with a large peak-to-valley difference in bandwidth allocation. This is due to the fact that the 1-minute sampling interval is much shorter than the device's operating time and the randomness of user-generated power consumption data. When household devices stop running or there are no dispatchable devices (FL) at time t, there will be a sudden decrease in data volume, which increases the computation time of bandwidth prediction. Fig. 21 displays the average, maximum, and minimum delays for each node during the simulation process. When a 5-minute sampling interval is used, most nodes have average, maximum, and minimum delays smaller than those observed with a 10-minute sampling interval. Therefore, this paper will adopt a 5-minute sampling interval.

D. LAYOUT OF EDGE NODES

According to the bandwidth allocation scheme for each node determined in Fig. 22, the delay of data transmission between nodes is composed of two parts: one is the transmission delay caused by the bandwidth, and the other is the propagation delay between the two nodes. In the IEEE 118-node diagram, a single node can exchange data with multiple nodes. In this paper, it is assumed that the data transmission rate between nodes is the same as the medium. Under this premise, the factors that cause data transmission delay between nodes are the bandwidth of the data sending node and the distance between nodes. As shown in Fig. 22, the minimum delay for data transmission between nodes is 0.9 seconds, and the maximum delay is 1.088 seconds. In order to avoid choosing nodes with high delay or sending to multiple nodes during data transmission, suitable ENs need to be selected to be responsible for data reception. The selection of ENs is shown in Table 6. Node 1 is designated as the cloud center, so the ENs and their responsible nodes are counted starting from node 2 in Table 6. Under this scheme, while keeping the lowest

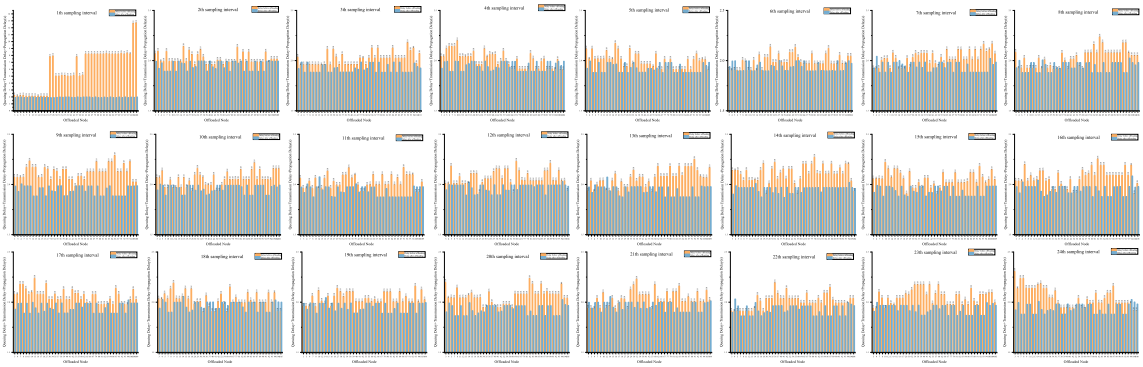

FIGURE 23. 24 moment BN offload solution.

TABLE 6. Edge node and business node affiliation & delay.

Node number	Edge node	Transmission delay & Propagation Delay (s)
2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20	3	0.9001、0.9000、0.9004、0.9007、0.9010、0.9012、0.9014、0.9021、0.9002、0.9006、0.9011、0.9013、0.9015、0.9021、0.9025、0.9026、0.9010、0.9010、0.9012
21、22、23、24	21	1、1.0002、1.0002、1.0004
25、26、27	26	1.0005、1、1.0003
28、29、53、54、55、56、57、58、59、60、	60	1.0029、1.0027、1.0024、1.0021、1.0017、1.0014、1.0013、1.0009、1.0005、1
30、31、32、33、34、35、36、37、38、39、40、41、42、43、61、62	32	1.0001、1.0001、1、1.0001、1.0004、1.0007、1.0010、1.0013、1.0016、1.0017、1.0019、1.0022、1.0024、1.0026
44、45	45	1.0002、1
46、47、48	47	1.0001、1、1.0003
49、50、51、52	52	1.0007、1.0006、1.0003、1
63、64、65、66、67、68、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95	64	1.0005、1.0007、0.9001、0.9000、0.9004、0.9008、0.9011、0.9003、0.9007、0.9009、0.9011、0.9012、0.9016、0.9020、0.9024、0.9010、0.9016、0.9020、0.9002、0.9007、0.9009、0.9012、0.9016、0.9019、0.9021
69、70、71、72	72	1.0007、1.0003、1.0001、1
73、74、75、76、77	75	1.0005、1.0003、1、1.0004、1.0008
96、97、98、99	98	1.0005、1.0002、1、1.0003
100、114、115、116	115	1.0007、1.0003、1、1.0005
101、102	102	1.0003、1
103	103	1
104、105、106、107、108、109、110、111	108	0.9259、0.9258、0.9254、0.9249、0.9247、0.9252、0.9253、0.9254
112、113	113	1.0675、1.0673
117、118	118	0.9007、0.9000

delay unchanged, the highest delay decreases to 1.0675, and most of the nodes have a delay of less than 1 second.

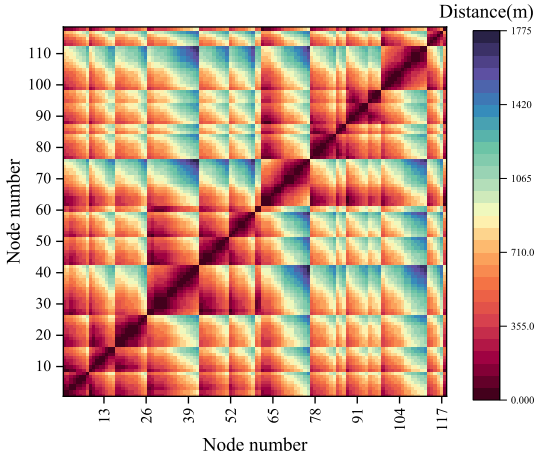

FIGURE 24. 24 moment BN offload solution.

The evaluation method for transmission rate is consistent with [30].

E. OVERALL OFFLOAD RESULTS

In this section the offloading of all BNs and the distance between each node will be shown.

1) OFFLOAD SCHEME DURING THE DISPATCH PERIOD

Fig.23 illustrates the offloading of each BN for 24 scheduling moments. It is worth noting that the delay experienced before the initial offloading moment differs considerably from the delay encountered after the offloading. This discrepancy arises due to the presence of a substantial number of dispatchable resources during the first moment, leading to the generation of a large volume of data. Consequently, the original bandwidth and computational resources become inadequate. However, with the adoption of the strategy proposed in this paper, a significant reduction in delay is achieved. Furthermore, as the dispatching period progresses, the delay tends to stabilize, resulting in a smoother operation.

2) DISTANCE BETWEEN NODES

Fig.24 displays the distances between 118 nodes, where a distance of 0 indicates the node itself. The interconnected

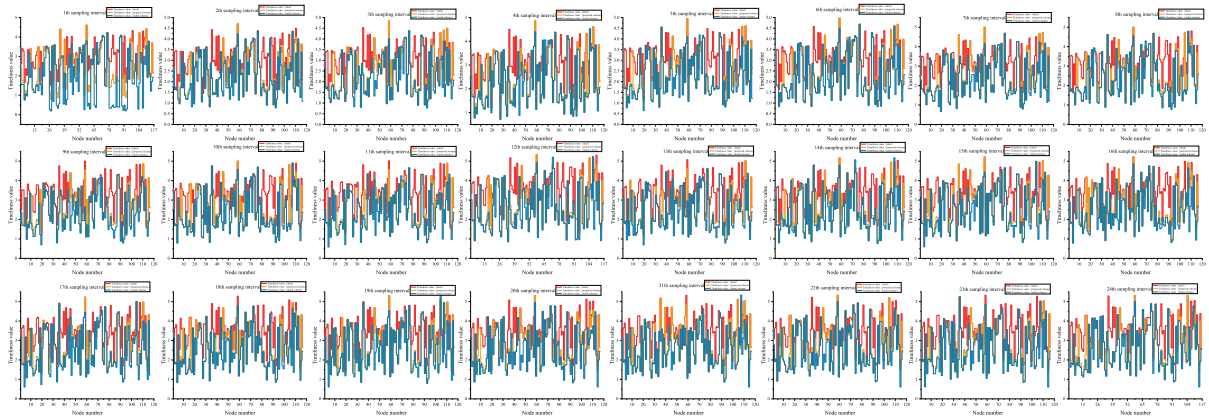

FIGURE 25. 24 moment' each node timeliness value.

nature of the nodes within the IEEE 118 node graph results in varying distances between them, leading to diverse propagation delays. These propagation delays is insignificant but still play a crucial role in determining the affiliation confirmation between ENs and BNs.

F. OVERALL TIMELINESS VALUE

This section will show the timeliness value of dispatchable resources at each node for 24 moments.

REFERENCES

[1] G. Lin, Y. Yang, F. Pan, S. Zhang, F. Wang, and S. Fan, "An optimal energy-saving strategy for home energy management systems with bounded customer rationality," *Future Internet*, vol. 11, no. 4, p. 88, Apr. 2019.

[2] F. Shen, Q. Wu, S. Huang, X. Chen, H. Liu, and Y. Xu, "Two-tier demand response with flexible demand swap and transactive control for real-time congestion management in distribution networks," *Int. J. Electr. Power Energy Syst.*, vol. 114, Jan. 2020, Art. no. 105399.

[3] M. Elkazaz, M. Sumner, E. Naghiyev, S. Pholboon, R. Davies, and D. Thomas, "A hierarchical two-stage energy management for a home microgrid using model predictive and real-time controllers," *Appl. Energy*, vol. 269, Jul. 2020, Art. no. 115118.

[4] A. S. Gazafroudi, M. Shafie-khah, E. Heydarian-Forushani, A. Hajizadeh, A. Heidari, J. M. Corchado, and J. P. S. Catalão, "Two-stage stochastic model for the price-based domestic energy management problem," *Int. J. Electr. Power Energy Syst.*, vol. 112, pp. 404–416, Nov. 2019.

[5] N. Koltsaklis, I. Panapakidis, G. Christoforidis, and J. Knápek, "Smart home energy management processes support through machine learning algorithms," *Energy Rep.*, vol. 8, no. 3, pp. 1–6, Jun. 2022.

[6] H. Shareef, M. S. Ahmed, A. Mohamed, and E. Al Hassan, "Review on home energy management system considering demand responses, smart technologies, and intelligent controllers," *IEEE Access*, vol. 6, pp. 24498–24509, 2018.

[7] J. Leitão, P. Gil, B. Ribeiro, and A. Cardoso, "A survey on home energy management," *IEEE Access*, vol. 8, pp. 5699–5722, 2020.

[8] N. Ghadimi, S. Nojavan, O. Abedinia, and A. B. Dehkordi, "Deterministic-based energy management of DC microgrids," in *Risk-Based Energy Management*, S. Nojavan, M. Shafieezadeh, and N. Ghadimi, Eds. New York, NY, USA: Academic, 2020, pp. 11–30.

[9] R. Sharifi, S. H. Fathi, and V. Vahidinasab, "A review on demand-side tools in electricity market," *Renew. Sustain. Energy Rev.*, vol. 72, pp. 565–572, May 2017.

[10] B. Jeddi, Y. Mishra, and G. Ledwich, "Distributed load scheduling in residential neighborhoods for coordinated operation of multiple home energy management systems," *Appl. Energy*, vol. 300, Oct. 2021, Art. no. 117353.

[11] F. Luo, W. Kong, G. Ranzi, and Z. Y. Dong, "Optimal home energy management system with demand charge tariff and appliance operational dependencies," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 4–14, Jan. 2020.

[12] O. Kebotogetse, R. Samikannu, and A. Yahya, "Review of key management techniques for advanced metering infrastructure," *Int. J. Distrib. Sensor Netw.*, vol. 17, no. 8, Aug. 2021, Art. no. 155014772110415.

[13] I. Gomes, K. Bot, M. G. Ruano, and A. Ruano, "Recent techniques used in home energy management systems: A review," *Energies*, vol. 15, no. 8, p. 2866, 2022.

[14] X. Wei, M. A. Amin, Y. Xu, T. Jing, Z. Yi, X. Wang, Y. Xie, D. Li, S. Wang, and Y. Zhai, "Two-stage cooperative intelligent home energy management system for optimal scheduling," *IEEE Trans. Ind. Appl.*, vol. 58, no. 4, pp. 5423–5437, Jul. 2022.

[15] O. Almughram, S. B. Slama, and B. Zafar, "Model for managing the integration of a vehicle-to-home unit into an intelligent home energy management system," *Sensors*, vol. 22, no. 21, p. 8142, 2022.

[16] S. Gherairi, "Design and implementation of an intelligent energy management system for smart home utilizing a multi-agent system," *Ain Shams Eng. J.*, vol. 14, no. 3, Apr. 2023, Art. no. 101897.

[17] P. Singh, S. Dhundhara, Y. P. Verma, and N. Tayal, "Optimal battery utilization for energy management and load scheduling in smart residence under demand response scheme," *Sustain. Energy, Grids Netw.*, vol. 26, Jun. 2021, Art. no. 100432.

[18] Z. Yang, M. Ni, and H. Liu, "Pricing strategy of multi-energy provider considering integrated demand response," *IEEE Access*, vol. 8, pp. 149041–149051, 2020.

[19] A. Anees, T. Dillon, S. Wallis, and Y.-P.-P. Chen, "Optimization of day-ahead and real-time prices for smart home community," *Int. J. Electr. Power Energy Syst.*, vol. 124, Jan. 2021, Art. no. 106403.

[20] J. Kim and G. Hwang, "Adaptive bandwidth allocation based on sample path prediction with Gaussian process regression," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4983–4996, Oct. 2019.

[21] X. Dai, X. Wang, and N. Liu, "Optimal scheduling of data-intensive applications in cloud-based video distribution services," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 1, pp. 73–83, Jan. 2017.

[22] E. Li, F. He, Q. Li, and H. Xiong, "Bandwidth allocation of stream-reservation traffic in TSN," *IEEE Trans. Netw. Service Manage.*, vol. 19, no. 1, pp. 741–755, Mar. 2022.

[23] J. Hui, C. Gan, L. Wu, and Z. Xu, "High bandwidth utilization DWBA algorithm for upstream channel in NG-EPON," *IEEE Access*, vol. 10, pp. 99435–99444, 2022.

[24] J. Xu, H. Wang, and L. Chen, "Bandwidth allocation for multiple federated learning services in wireless edge networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2534–2546, Apr. 2022.

[25] W. Feng and X. Zhang, "Wireless federated learning with dynamic quantization and bandwidth adaptation," *IEEE Wireless Commun. Lett.*, vol. 11, no. 11, pp. 2335–2339, Nov. 2022.

[26] J. Yang, Y. Zhou, W. Wen, J. Zhou, and Q. Zhang, "Asynchronous hierarchical federated learning based on bandwidth allocation and client scheduling," *Appl. Sci.*, vol. 13, no. 20, p. 11134, 2023.

[27] J. Cao, Z. Ma, J. Xie, X. Zhu, F. Dong, and B. Liu, "Towards tenant demand-aware bandwidth allocation strategy in cloud datacenter," *Future Generat. Comput. Syst.*, vol. 105, pp. 904–915, Apr. 2020.

[28] X. Liu, Z. Li, P. Xu, and J. Li, "Joint optimization for bandwidth utilization and delay based on particle swarm optimization," *IEEE Access*, vol. 9, pp. 92125–92133, 2021.
[29] J. Kim and G. Hwang, "Prediction based efficient online bandwidth allocation method," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2330–2334, Dec. 2019.
[30] Z. Wang, K. Han, X. Shen, W. Yuan, and F. Liu, "Achieving the performance bounds for sensing and communications in perceptive networks: Optimal bandwidth allocation," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1835–1839, Sep. 2022.
[31] S. Chen, H. Wen, J. Wu, W. Lei, W. Hou, W. Liu, A. Xu, and Y. Jiang, "Internet of Things based smart grids supported by intelligent edge computing," *IEEE Access*, vol. 7, pp. 74089–74102, 2019.
[32] S. Lu, J. Lu, K. An, X. Wang, and Q. He, "Edge computing on IoT for machine signal processing and fault diagnosis: A review," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11093–11116, Jul. 2023.
[33] Y. Chen, W. Gu, J. Xu, Y. Zhang, and G. Min, "Dynamic task offloading for digital twin-empowered mobile edge computing via deep reinforcement learning," *China Commun.*, vol. 20, no. 11, pp. 164–175, Nov. 2023.
[34] K. Sadatdiynov, L. Cui, L. Zhang, J. Z. Huang, S. Salloum, and M. S. Mahmud, "A review of optimization methods for computation offloading in edge computing networks," *Digit. Commun. Netw.*, vol. 9, no. 2, pp. 450–461, Apr. 2023.
[35] J. Huang, J. Wan, B. Lv, Q. Ye, and Y. Chen, "Joint computation offloading and resource allocation for edge-cloud collaboration in Internet of Vehicles via deep reinforcement learning," *IEEE Syst. J.*, vol. 17, no. 2, pp. 2500–2511, Jun. 2023.
[36] Z. Su, Y. Wang, T. H. Luan, N. Zhang, F. Li, T. Chen, and H. Cao, "Secure and efficient federated learning for smart grid with edge-cloud collaboration," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1333–1344, Feb. 2022.
[37] R. Wang, J. Xu, H. Gao, S. He, and J. Liu, "Cloud-edge collaboration based operation strategy in medium-low distribution network," in *Proc. Panda Forum Power Energy (PandaFPE)*, Chengdu, China, Apr. 2023, pp. 1776–1780.
[38] G. Wang, W. Xu, Y. Wang, H. Liu, J. Xiang, and Z. Tang, "Application research of edge IoT proxy technology based on 5G in power demand side management system," in *Proc. 12th Int. Conf. Comput. Eng. Netw. (CENet)*, in Lecture Notes in Electrical Engineering, vol. 961, Q. Liu, X. Liu, J. Cheng, T. Shen, and Y. Tian, Eds. Singapore: Springer, 2022, doi: 10.1007/978-981-19-6901-0_51.
[39] X. Wu, L. You, R. Wu, Q. Zhang, and K. Liang, "Management and control of load clusters for ancillary services using Internet of Electric Loads based on cloud–edge–end distributed computing," *IEEE Internet Things J.*, vol. 9, no. 19, pp. 18267–18279, Oct. 2022.
[40] Y. Zhang, X. Wang, J. He, Y. Xu, F. Zhang, and Y. Luo, "A transfer learning-based high impedance fault detection method under a cloud-edge collaboration framework," *IEEE Access*, vol. 8, pp. 165099–165110, 2020.
[41] C. Gong, F. Lin, X. Gong, and Y. Lu, "Intelligent cooperative edge computing in Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9372–9382, Oct. 2020.
[42] Y. He, B. Venkatesh, and L. Guan, "Optimal scheduling for charging and discharging of electric vehicles," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1095–1105, Sep. 2012.
[43] Q. Ou, Y. Wang, W. Song, N. Zhang, J. Zhang, and H. Liu, "Research on network performance optimization technology based on cloud-edge collaborative architecture," in *Proc. IEEE 2nd Int. Conf. Big Data, Artif. Intell. Internet Things Eng. (ICBAIE)*, Mar. 2021, pp. 274–278.

WEI ZHANG (Member, IEEE) received the Ph.D. degree in electrical engineering and its automation from Wuhan University, Wuhan, China, in 2015. He is currently a Lecturer with the Department of Electrical Engineering, University of Shanghai for Science and Technology, Shanghai, China. His current research interests include electric power demand response, the dispatch of integrated energy systems, and the electric Internet of Things.

CHENHAO HUANG is currently pursuing the B.S. degree in electrical engineering with the Department of Electrical Engineering, University of Shanghai for Science and Technology, Shanghai, China. His research interest includes the corresponding effects of communication in power systems on demand.