This article has been accepted for publication in IEEE Transactions on Smart Grid. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSG.2023.3241653

Consensus-based Energy Management of Microgrid with Random Packet Drops

Hongyi Li, *Student Member, IEEE,* Hongxun Hui, *Member, IEEE,* Hongcai Zhang, *Member, IEEE*

***Abstract*—Decentralized energy management using consensus-based algorithm is a vibrant research field since it can promote the local accommodation of the renewable energy generation without raising privacy and scalability issues. Most of the existing methods assume that the communication links are reliable, which might not be true in real-world implementations. This paper focuses on tackling the problem of random packet drops. We first formulate the models of the energy management problem in the microgrid and the information packet drop in the communication network. Based on the models, we conclude that losing the information about incremental electricity cost estimation is tolerable, while losing the information about the power mismatch estimation is not. We propose a novel consensus algorithm that tracks and exchanges the accumulated value of the power mismatch estimation so that the information loss can be recovered. An equivalent form of the proposed method is established by augmenting the communication links with virtual buffering nodes. Based on the augmented communication topology, we theoretically prove the convergence of the proposed algorithm and the optimality of the solution. Several case studies are provided to validate the effectiveness of the proposed algorithm.**

***Index Terms*—Consensus-based algorithm, microgrid energy management, random packet drops, renewable energy.**

I. Introduction

THE vision of carbon neutrality motivates the proliferation of distributed renewable energy sources (RESs) [1]. Distributed RESs are small-sized onsite generators whose power generation can be locally consumed by the loads, so that the transmission cost is reduced [2]. However, the local accommodation of distributed RES generation is challenging, since the RES generation is intermittent, uncertain and often non-dispatchable [3]. Microgrids, which provide interconnections among distributed RESs, dispatchable diesel generators and flexible loads, are regarded as a promising solution to promote the local accommodation of distributed RES generation [4].

To accommodate the distributed RES generation locally, the dispatchable diesel generators and the flexible loads need to be optimally coordinated, which is termed as the energy management problem of the microgrid [5]. Some researchers propose centralized methods to solve the microgrid energy management problem [6]–[8]. Although these centralized methods can promote the local accommodation of RES generation, there exist some critical issues, such as privacy protection, single point of failure and low scalability [9]. With the fast development of information and communication technologies, the entities in the microgrid can process data locally and communicate with each other [10]. Thus, these entities can form a multi-agent system, among which the energy management problem can be solved in a decentralized manner [11].

The consensus-based algorithm is widely adopted in the decentralized coordination of a multi-agent system, where the agents exchange information that are non-private, i.e., the estimations of incremental electricity cost and global power mismatch. Rahbari-Asr *et al.* [12] propose a consensus-based algorithm for energy management of the smart grid, which adopts an undirected communication topology. Yang *et al.* [13] propose a consensus algorithm for a more general directed communication topology. Yan *et al.* [14], Zhao *et al.* [15] and Yang *et al.* [16] further study consensus-based energy management with different objectives over different communication topologies. The above-mentioned algorithms assume that each agent knows its out-degree, i.e., the number of agents that receive its information, which might not be true in real-world implementations [17]. One typical scenario is that the communication links are non-ideal, i.e., the information packet drops randomly [18], as illustrated in Fig. 1. Since random packet drops cannot be detected *a priori*, it becomes difficult for the sending agents to accurately estimate the out-degrees, implying that the optimality and feasibility of the obtained solution might be lost.

To tackle this issue, Mao *et al.* [19] propose a similarity-based method to estimate the lost information. However, since the lost information cannot be fully recovered through approximation, the optimality of the solution cannot be guaranteed. Zhang *et al.* [20] and Wang *et al.* [21] propose gossip updating rules for the information exchanged among the agents, i.e., the estimation of power mismatch and the estimation of incremental electricity cost, respectively. Duan *et al.* [22] propose a corrective method based on the symmetry of the undirected graphs. Although these methods guarantee the convergence to global optimum, their application might be limited as they require the communication topologies to be undirected. For a more general scenario where the communication topology is directed, Hadjicostis *et al.* [23] propose the method of running sums to solve the average consensus problem with random packet drops. Wu *et al.* [24] introduce the method of running sums to energy management, and propose a ratio-consensus-based algorithm over a directed graph. However, this algorithm requires a diminishing step size to guarantee the convergence,

This paper is funded in part by the National Natural Science Foundation of China under Grant 52007200 and in part by the Science and Technology Development Fund, Macau SAR (File no. SKL-IOTSC(UM)-2021-2023, and File no. 0011/2021/AGJ). (Corresponding author: *Hongcai Zhang*.)

H. Li, H. Hui, and H. Zhang are with the State Key Laboratory of Internet of Things for Smart City and Department of Electrical and Computer Engineering, University of Macau, Macao, 999078 China (email: hczhang@um.edu.mo).

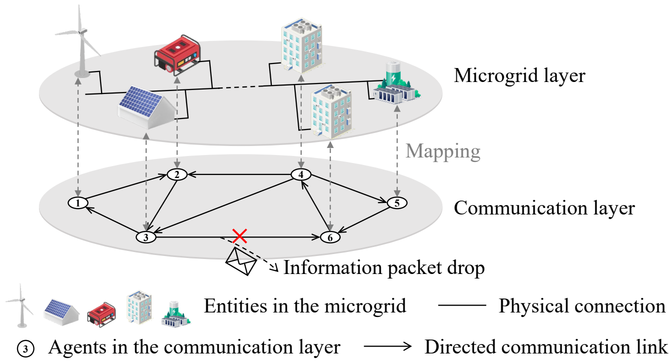


Fig. 1. In a microgrid with directed communication network, random packet drop might occur when the communication links are non-ideal. For example, in this figure, agent 6 is supposed to receive an information packet from agent 3 at every iteration. However, random packet drops might happen, which deteriorate the optimality of the solution obtained by the consensus-based energy management algorithm.

which might lead to slow convergence.

In this paper, we propose a consensus-based energy management algorithm to handle the random packet drops over directed communication topologies. We first theoretically the impact of random packet drops on the conventional method. We conclude that it is tolerable to lose the information about the estimation of incremental electricity cost, while the estimation of power mismatch is critical to the feasibility of the solution. Therefore, we propose to exchange the accumulated value of the power mismatch estimation among agents, instead of exchanging the power mismatch estimation itself. In this way, the information loss can be compensated when the failed communication link resumes. Our proposed method can obtain the global optimal solution of the energy management problem in the presence of random packet drops, over both directed and undirected communication topologies. The effectiveness of the proposed method is validated both theoretically and numerically. Multiple scenarios are simulated to show the effectiveness of the proposed method in a comprehensive way. Beyond tackling the problem of random packet drops, the proposed method converges at a faster rate, compared with existing methods, which is illustrated through comprehensive experiments. The contributions of this paper are threefold:

1) We build the models of the energy management in a microgrid and the random packet drops in the communication network. Based on the models, we theoretically analyze the convergence of the conventional method and the optimality of the obtained solution, under the impact of random packet drops.
2) We propose a novel consensus-based algorithm that allows the agents to exchange the accumulated values of their power mismatch estimation. In this way, the vulnerability to random packet drops is eliminated and the feasibility of the solution is guaranteed, when solving the microgrid energy management problem over both directed and undirected communication topologies.
3) We develop an equivalent form of the proposed method by augmenting the communication topology with virtual buffering nodes. With the equivalent form, we theoretically prove that the proposed algorithm converges to the global optimum in both mean square sense and almost sure sense, using the matrix perturbation theory.

The remainder of this paper is organized as follows. Section II includes the background knowledge and the theoretical analysis of the conventional method. Section III introduces the design of the proposed algorithm and presents the proofs of convergence and optimality. In Section IV, the performance of the proposed algorithm is validated through multiple experiments. Section V concludes the paper.

II. Preliminaries and Conventional Method

In this section, we introduce the optimization model of the energy management problem and the formulation of the conventional consensus-based algorithm over directed communication topologies. Through an example, we show that the conventional method is vulnerable to packet drops. We also theoretically analyze the impact of random packet drops on the convergence and the optimality, which paves the way to the design of our proposed method.

A. Terms, Notations and Assumptions

Before presenting the mathematical models, we first introduce some necessary terms, notations and assumptions. In this paper, the energy management problem is modeled as the optimal coordination of the RESs, the diesel generators and the flexible loads, which are all termed as *agents* throughout the following discussion. The set of all agents is defined as $\mathcal{I} = \mathcal{I}_{\mathrm{R}} \cup \mathcal{I}_{\mathrm{G}} \cup \mathcal{I}_{\mathrm{L}}$, where $\mathcal{I}_{\mathrm{R}}$, $\mathcal{I}_{\mathrm{G}}$ and $\mathcal{I}_{\mathrm{L}}$ are the sets of the RESs, the diesel generators and the flexible loads, respectively. The number of all agents is $N = |\mathcal{I}_{\mathrm{R}}| + |\mathcal{I}_{\mathrm{G}}| + |\mathcal{I}_{\mathrm{L}}|$, where $|\mathcal{I}_{\mathrm{R}}|$, $|\mathcal{I}_{\mathrm{G}}|$ and $|\mathcal{I}_{\mathrm{L}}|$ are the cardinalities of $\mathcal{I}_{\mathrm{R}}$, $\mathcal{I}_{\mathrm{G}}$ and $\mathcal{I}_{\mathrm{L}}$, respectively. Regarding each agent as a node in the graph, we depict the communication topology as a directed graph $\mathcal{G} = \{\mathcal{I}, \mathcal{E}\}$, where $\mathcal{I}$ and $\mathcal{E}$ are the set of nodes and edges, respectively. In the presence of random packet drops, the directed graph $\mathcal{G}$ becomes time-varying, i.e., $\mathcal{G}(k) = \{\mathcal{I}, \mathcal{E}(k)\}$, where k is the iteration index and $\mathcal{E}(k) \subseteq \mathcal{E}$. An edge $(i,j) \in \mathcal{E}(k)$ implies that agent i successfully receives information from agent j at iteration k. The set of in-neighbors of agent i is defined as $\mathcal{N}_i^+(k) = \{j | (i,j) \in \mathcal{E}(k), \ i \neq j\}$, while the set of out-neighbors of agent i is defined as $\mathcal{N}_i^-(k) = \{j | (j,i) \in \mathcal{E}(k), \ i \neq j\}$. The in-degree and out-degree of the node is defined as the cardinalities of the in-neighbor set and the out-neighbor set, respectively. We use a binary variable $\gamma_{ij}(k)$ to indicate packet drop. We let $\gamma_{ij}(k) = 1$ if agent i successfully receives the information from agent j at iteration k, otherwise $\gamma_{ij}(k) = 0$. We use $\mathbf{0}_{\mathrm{N}}$ and $\mathbf{1}_{\mathrm{N}}$ to denote an N dimensional vector of all zeros/ones. Symbol $\mathbf{0}_{\mathrm{M}\times\mathrm{N}}$ denotes an $\mathrm{M}\times\mathrm{N}$ matrix with all elements equal to zero. The identity matrix of size N, which is an $\mathrm{N}\times\mathrm{N}$ square matrix with ones on the main diagonal and zeros elsewhere, is denoted by $\mathbf{I}_{\mathrm{N}}$. For a vector $\boldsymbol{a}$, we denote its transpose by $\boldsymbol{a}^{\mathrm{T}}$. Throughout this paper, we make the following assumptions:

[A1] The directed graph $\mathcal{G}$ is fixed and strongly connected. We also assume that $\mathcal{G}$ is a simple graph, i.e., self-loops and multiple edges are not considered.

[A2] Each agent holds a unique identifier (ID), which is used to label the information sent by the agent[1]. Each agent knows its out-degree $\left|\mathcal{N}_i^-\right|$ under ideal communication topology, where $\mathcal{N}_i^- = \{j | (j,i) \in \mathcal{E},\ i \neq j\}$.

[A3] At every iteration, each agent packs its information in one packet and sends the packet to its out-neighbors only once, i.e., a packet drop results in the loss of all information and there is no retransmission.

[A4] The occurrences of packet drops are identically distributed and independent among different communication links and different iterations. The probability that the packet from agent j to agent i drops is denoted by p_{ij}, i.e., $\Pr\left(\gamma_{ij}(k) = 0\right) = p_{ij}$. For $(i,j) \in \mathcal{E}$, $0 \leq p_{ij} < 1$.

[A5] The packet drop is only perceived by the receiving agent. For example, if $(i,j) \in \mathcal{E}$ while $(i,j) \notin \mathcal{E}\left(k\right)$, agent i perceives the packet drop while agent j does not.

[A6] Each agent sends identical information to all its out-neighbors, i.e., the message is *broadcasted*. However, this assumption might not hold in some circumstances, which will be discussed later.

B. Formulation of Microgrid Energy Management Problem

Assuming that the RES generation has no cost, the energy management problem of a microgrid can be formulated as[2]:

$$\min_{P_i, i \in I} \ \sum_{i \in \mathcal{I}_{\rm G}} C_i(P_i) - \sum_{i \in \mathcal{I}_{\rm L}} U_i(P_i) \tag{1a}$$

$$\text{s.t.} \ \sum_{i \in \mathcal{I}_{\rm R}} P_i + \sum_{i \in \mathcal{I}_{\rm G}} P_i - \sum_{i \in \mathcal{I}_{\rm L}} P_i = 0, \tag{1b}$$

$$P_i^{\min} \leq P_i \leq P_i^{\max},\ \forall i \in \mathcal{I}, \tag{1c}$$

where P_i is the power generation/consumption of agent i; $C_i(\cdot)$ is the generation cost function in \$; $U_i(\cdot)$ is the utility function of the flexible loads that describes the satisfaction of electricity consumption in \$; $P_i^{\min}$ and $P_i^{\max}$ are the lower and upper bounds of the power generation/consumption, respectively. Objective (1a) is designed to maximize the social welfare of the microgrid. Constraint (1b) ensures that the power supply and demand is balanced. Constraint (1c) ensures that the power generation/consumption lies within the feasible range. With the assumptions that the generation cost function is convex [12], [13] and the utility function is concave [26], problem (1) is a convex optimization problem with affine constraints. To ensure that there exists at least one feasible solution to problem (1). we assume that the following relationship holds:

$$\sum_{i \in \mathcal{I}_{\rm R} \cup \mathcal{I}_{\rm G}} P_i^{\min} \leq \sum_{i \in \mathcal{I}_{\rm L}} P_i^{\min} < \sum_{i \in \mathcal{I}_{\rm L}} P_i^{\max} \leq \sum_{i \in \mathcal{I}_{\rm R} \cup \mathcal{I}_{\rm G}} P_i^{\max}. \tag{2}$$

C. Consensus Algorithm with Ideal Communication Links

The consensus-based algorithm solves problem (1) in an iterative and interactive manner, which is introduced as follows. Since problem (1) is convex, strong duality holds. Thus, the classical primal-dual decomposition method [27] can be applied and the partial Lagrangian dual of the objective is:

$$\begin{aligned} \mathcal{L}\left(\boldsymbol{P}, \lambda\right) = & \sum_{i \in \mathcal{I}_{\rm G}} C_i(P_i) - \sum_{i \in \mathcal{I}_{\rm L}} U_i(P_i) \\ & + \lambda \left(\sum_{i \in \mathcal{I}_{\rm L}} P_i - \sum_{i \in \mathcal{I}_{\rm G}} P_i - \sum_{i \in \mathcal{I}_{\rm R}} P_i \right), \end{aligned} \tag{3}$$

where λ is the Lagrangian multiplier. In the context of energy management, λ is also interpreted as the incremental cost of electricity. We refer to the update of λ as the solving of the master problem, which is defined as:

$$\lambda\left(k+1\right) = \lambda\left(k\right) + \eta \zeta\left(k\right), \tag{4}$$

$$\zeta\left(k\right) = \left(\sum_{i \in \mathcal{I}_{\rm L}} P_i\left(k\right) - \sum_{i \in \mathcal{I}_{\rm G}} P_i\left(k\right) - \sum_{i \in \mathcal{I}_{\rm R}} P_i\left(k\right) \right), \tag{5}$$

where the step size η is a small positive constant; $\zeta\left(k\right)$ depicts the global power mismatch, which is also the gradient of the Lagrangian dual with respect to λ.

Since the Lagrangian dual (3) is separable and the power limit constraints (1c) are local constraints of individual agents, given $\lambda(k+1)$, each agent can solve its sub-problem locally, which is defined as:

$$P_i\left(k+1\right) = \begin{cases} \begin{array}{l} \underset{P_i}{\text{argmin}} \ -\lambda\left(k+1\right) P_i \\ \quad \text{s.t. } 0 \leq P_i \leq P_i^{\max} \end{array}, & \forall i \in \mathcal{I}_{\rm R}, \\ \begin{array}{l} \underset{P_i}{\text{argmin}} \ C_i(P_i) - \lambda\left(k+1\right) P_i \\ \quad \text{s.t. } P_i^{\min} \leq P_i \leq P_i^{\max} \end{array}, & \forall i \in \mathcal{I}_{\rm G}, \\ \begin{array}{l} \underset{P_i}{\text{argmin}} \ \lambda\left(k+1\right) P_i - U_i(P_i) \\ \quad \text{s.t. } P_i^{\min} \leq P_i \leq P_i^{\max} \end{array}, & \forall i \in \mathcal{I}_{\rm L}. \end{cases} \tag{6}$$

By solving the master problem and the sub-problems iteratively, the optimal solution of problem (1) can be obtained. However, the calculation of equations (4) and (5) requires global information, which is inaccessible to individual agents if problem (1) is solved in a decentralized way. Thus, two local consensus variables, namely λ_i and ζ_i, are introduced to denote the local estimations of the global information. Without considering random packet drops in the communication network, the conventional consensus-based algorithm to solve problem (1) over a directed graph is outlined as follows [16]:

[1]Each agent holds a unique ID is practical in real-world implementations. For example, the IDs can be identical to the media access control (MAC) address of the device, which is widely used in most communication technologies that follow IEEE 802 standards, e.g., Ethernet, Wi-Fi and Bluetooth. Since the MAC address of a device is unique within a network segment, it can be used as the ID of the agent.

[2]The power transmission loss in the microgrid is not considered in the problem formulation (1), since it is not the main focus of this paper. The modeling of the transmission loss in a microgrid can be found in [25] and [15]. The authors in [15] propose a convex relaxation of the microgrid energy management problem with power transmission loss. Based on this relaxed formulation, our proposed method can be applied to solve the energy management problem considering power transmission loss and random packet drops simultaneously.

This article has been accepted for publication in IEEE Transactions on Smart Grid. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSG.2023.3241653

1) Step 1 - Initialization: In the consensus-based algorithm, each agent maintains the following variables: the power generation/consumption $P_i(k)$; the price estimation $\lambda_i(k)$; the power mismatch estimation $\zeta_i(k)$. Setting the iteration index k to 0, $P_i(0)$ can be initialized to any admissible value within $\left[P_i^{\min}, P_i^{\max}\right]$, and the rest of the variables are initialized as:

$$\lambda_i(0) = \begin{cases} 0, & \forall i \in \mathcal{I}_{\mathrm{R}}, \\ C_i'(P_i(0)), & \forall i \in \mathcal{I}_{\mathrm{G}}, \\ U_i'(P_i(0)), & \forall i \in \mathcal{I}_{\mathrm{L}}, \end{cases} \tag{7}$$

$$\zeta_i(0) = \begin{cases} -P_i(0), & \forall i \in \mathcal{I}_{\mathrm{R}} \cup \mathcal{I}_{\mathrm{G}}, \\ P_i(0), & \forall i \in \mathcal{I}_{\mathrm{L}}, \end{cases} \tag{8}$$

where $C_i'(\cdot)$ and $U_i'(\cdot)$ are the first-order derivatives of $C_i(\cdot)$ and $U_i(\cdot)$, respectively.

2) Step 2 - Information exchange: At the beginning of iteration k, each agent sends the current value of two consensus variables, namely $\lambda_i(k)$ and $\zeta_i(k)$, to its out-neighbors.

3) Step 3 - Local update: Based on the received $\lambda_j(k)$ and $\zeta_j(k)$, each agent performs the following updates locally:

a) Update price estimation:

$$\lambda_i(k+1) = \lambda_i(k) + \sum_{j \in \mathcal{N}_i^+} a_{ij}\left(\lambda_j(k) - \lambda_i(k)\right) + \eta \zeta_i(k), \tag{9}$$

where the weight a_{ij} is defined as:

$$a_{ij} = \begin{cases} \frac{1}{|\mathcal{N}_i^+|+1}, & \forall j \in \mathcal{N}_i^+, \\ 0, & \text{otherwise}. \end{cases} \tag{10}$$

b) Optimal response: Each agent optimally responds to $\lambda_i(k+1)$ by solving the sub-problem (6) locally.

c) Update power mismatch estimation:

$$\zeta_i(k+1) = b_{ii}\zeta_i(k) + \sum_{j \in \mathcal{N}_i^+} b_{ij}\zeta_j(k) + \hat{P}_i(k), \tag{11}$$

where the weight b_{ij} is defined as:

$$b_{ij} = \begin{cases} \frac{1}{|\mathcal{N}_j^-|+1}, & \forall i \in \mathcal{N}_j^- \cup \{j\}, \\ 0, & \text{otherwise}, \end{cases} \tag{12}$$

and the change of power $\hat{P}_i(k)$ is defined as:

$$\hat{P}_i(k) = \begin{cases} P_i(k) - P_i(k+1), \forall i \in \mathcal{I}_{\mathrm{R}} \cup \mathcal{I}_{\mathrm{G}}, \\ P_i(k+1) - P_i(k), \forall i \in \mathcal{I}_{\mathrm{L}}. \end{cases} \tag{13}$$

4) Step 4 - Termination check: After local updating of the variables, each agent checks the following terminating criteria:

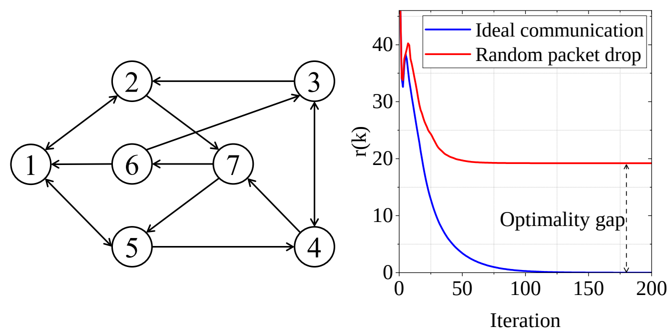
$$\begin{cases} \|\lambda_i(k+1) - \lambda_i(k)\| \le \epsilon_\lambda, \\ \|\zeta_i(k+1) - \zeta_i(k)\| \le \epsilon_\zeta, \end{cases} \tag{14}$$

where ϵ_λ and ϵ_ζ are thresholds for λ_i and ζ_i, respectively. If the conditions in (14) are satisfied simultaneously, agent i reaches a stationary point and recognizes it as the optimal solution; otherwise, set $k = k+1$ and return to Step 2.

In the following discussion, we refer to the conventional consensus-based algorithm, namely equations (6) - (14), as ***Algorithm 1***.

TABLE I
PARAMETERS OF AGENTS IN THE MICROGRID

Agent	Type	q ($/kW^2)	l ($/kW)	$P^{\max}$ (kW)	$P^{\min}$ (kW)	P^* (kW)
1	RES	-	-	46	0	46.0000
2	Diesel generator	0.00040	0	93	20	62.3419
3	Diesel generator	0.00045	0	82	30	55.4150
4	Diesel generator	0.00056	0	56	15	44.5299
5	Flexible load	-0.0072	0.8352	58	45	54.5366
6	Flexible load	-0.0084	1.3776	82	60	79.0313
7	Flexible load	-0.0076	1.1856	78	58	74.7188



(a) Communication topology $\mathcal{G}_1$. (b) Performance of Algorithm 1.

Fig. 2. The communication topology $\mathcal{G}_1$ used in the motivating example and the results of Algorithm 1 in different scenarios. With ideal communication, Algorithm 1 converges to the optimum while there exist an optimality gap in the presence of random packet drops.

D. Impact of Random Packet Drops on Algorithm 1

We first discuss the convergence of Algorithm 1 in the presence of random packet drops. The first two terms of equation (9) calculate the weighted average of the local estimation and the received information. According to the definition of a_{ij}, for agents $j \in \mathcal{N}_i^+$, the weights a_{ij} are equal and sum up to 1, thus the variables $\lambda_i, i \in \mathcal{I}$, will gradually converge to a consensus value. Considering random packet drops, the weights a_{ij} can be calculated on an iteration-by-iteration basis, based on the in-degree observed by agent i at every iteration. In this way, the weights $a_{ij}(k)$ remain equal and sum up to 1, i.e., the first two terms are always guiding the variables $\lambda_i, i \in \mathcal{I}$ to approach the same value.

The last term of equation (9) is an innovation term, where $\zeta_i(k)$ can be interpreted as the global net demand estimated by agent i at iteration k. We explain the rationale behind this heuristic term with microeconomics. If there is a net demand, agent i expects the cost of electricity to increase, since more incentives are needed to encourage more generation, and vice versa. According to the sub-problems (6), as the price estimation increases, the power generation of an agent will monotonically increase before its upper bound is reached, while the power consumption of an agent will monotonically decrease until the lower bound is reached. According to equation (13), $\hat{P}_i(k)$ will be negative, leading power mismatch estimation ζ_i to approach zero when it is updated with equation (11).

Therefore, with random packet drops, Algorithm 1 iterates to a stationary point where the price estimations reach a consensus value and the power mismatch estimations reach zero, i.e., $\lambda_i = \hat{\lambda}, \zeta_i = 0, \forall i \in \mathcal{I}$. Since the terminating criteria (14) are satisfied, Algorithm 1 is regarded as converged. However, the algorithm converges to a non-optimal, or even infeasible solution of the energy management problem (1).

With a motivating example, we demonstrate that Algorithm 1 converges to the global optimum of problem (1) with ideal communication, while it fails when there exist random packet drops. The parameters of the microgrid, termed as MG_1, are outlined in Table I. Without losing generality, we adopt quadratic generation cost functions and utility functions, where q and l are the coefficients of the quadratic term and the linear term, respectively. The last column of Table I, P^*, presents the optimal solution obtained in a centralized way using GUROBI [28]. The directed communication topology $\mathcal{G}_1$ is adopted, as shown in Fig. 2a. We set the probabilities of packet drops of all communication links to 5%, i.e., $p_{ij} = 0.05, \forall (i,j) \in \mathcal{E}$. To describe the gap between the obtained solution and the global optimum, we define the residual as the Euclidean distance to the optimum:

$$r(k) = \sqrt{\sum_{i \in \mathcal{I}} \left(P_i\left(k\right) - {P_i}^*\right)^2}. \tag{15}$$

The results of Algorithm 1 with both ideal and non-ideal communication links are shown in Fig. 2b. Without packet drops, Algorithm 1 successfully converges to the global optimum. However, in the presence of packet drops, although the algorithm still converges, the obtained solution is non-optimal, or even infeasible, as we can observe an optimality gap. Recall that assumption **[A3]** implies that when a packet drop occurs on link (i,j) at iteration k, both the information about $\lambda_j(k)$ and $\zeta_j(k)$ are lost. However, the impacts of the loss of $\lambda_j(k)$ and $\zeta_j(k)$ on the feasibility and optimality of the solution are different, which are discussed in the following paragraphs.

As previously discussed, by calculating a_{ij} on an iteration-by-iteration basis, the weights $a_{ij}(k)$ remain equal and sum up to 1. Thus, the first two terms of equation (9) always guide the variables $\lambda_i, i \in \mathcal{I}$ to approach the consensus value, and the lost of $\lambda_j(k)$ has trivial impact. In the previous discussions, we also show λ_i is heuristically tuned by the estimation of global power mismatch, ζ_i. Thus, it is important for agent i to correctly estimate the global power mismatch. However, the lost of $\zeta_j(k)$ could result in agent i having an inaccurate estimation of the global power mismatch.

Due to the lack of global information, $\zeta_i(k)$ can only track the global power mismatch in a collective way, i.e.,

$$\sum_{i \in \mathcal{I}} \zeta_i\left(k\right) = \zeta\left(k\right). \tag{16}$$

To prove this, we sum up both sides of equation (11) and get

$$\sum_{i \in \mathcal{I}} \zeta_i\left(k+1\right) = \sum_{i \in \mathcal{I}} \zeta_i\left(k\right) + \zeta\left(k+1\right) - \zeta\left(k\right), \tag{17}$$

which holds for every k given that matrix $\mathcal{B} = [b_{ij}]$ is column-stochastic, according to definition (12). According to equations (5) and (8), equation (16) also holds for $k = 0$. Thus, based on equation (17), we conclude that equation (16) holds for every k. As previously discussed, whether λ_i increases or decreases, according to equation (6), the reaction of agent i leads its power mismatch estimation to approach zero. Therefore, we conclude that $\zeta_i(k) \to 0, \forall i \in \mathcal{I}$ as $k \to \infty$. As $\zeta_i(k) \to 0$, equation (16) implies that $\zeta(k) \to 0$ as $k \to \infty$, i.e., the equality constraint (1b) is satisfied. Thus, one of the necessary conditions for the feasibility of the solution is that equation (16) holds.

However, since the weights b_{ij} are set by the sending agents, as suggested in assumption **[A5]**, the weight matrix $\mathcal{B}$ might not be column-stochastic in the presence of random packet drops. Therefore, equation (16) might not hold, implying that the obtained solution might not satisfy the equality constraint (1b). Thereby, we can conclude that although both the information about $\lambda_j(k)$ and $\zeta_j(k)$ are lost, only $\zeta_j(k)$ needs to be recovered. In the next section, we propose an algorithm that can recover the lost information about $\zeta_j(k)$ when the failed communication link resumes, and show that the convergence to the global optimum is guaranteed.

III. Proposed Consensus-based Energy Management Algorithm

Based on the previous discussion, we find that the estimations of global power mismatch, i.e., $\zeta_i, \forall i \in \mathcal{I}$, are critical to the feasibility of the solution. In this section, we propose a method that allows the receiving agents to recover the lost information with auxiliary variables. Before numerical demonstration of the effectiveness, we theoretically prove that the proposed algorithm converges to the optimal solution of problem (1) in the presence of random packet drops.

A. Design of the Proposed Algorithm

Since the packet drops are inevitable, our proposed method recovers and adds the lost information when the packet is successfully delivered again. Taking communication link (i,j) as an example, this can be achieved by introducing two auxiliary variables κ_j and κ_{ij}. Variable κ_j is maintained by the sending agent j, which is defined as the accumulated value of $b_{ij}\zeta_j$ that agent j sends from the first iteration:

$$\kappa_j\left(k\right) = \sum_{m=0}^{k} b_{ij}\zeta_j\left(m\right). \tag{18}$$

Variable κ_{ij} is maintained by the receiving agent i, which is defined as the last κ_j received from agent j:

$$\kappa_{ij}\left(k+1\right) = \begin{cases} \kappa_j\left(k\right), & \gamma_{ij}\left(k\right) = 1, \\ \kappa_{ij}\left(k\right), & \text{otherwise.} \end{cases} \tag{19}$$

At every iteration, agent j sends κ_j to agent i, instead of $b_{ij}\zeta_j$. The rationale behind this is explained as follows. Assuming that the communication at iteration $k-1$ is successful, agent i will let $\kappa_{ij}(k) = \kappa_j(k-1)$. If agent i receives $\kappa_j(k)$ from agent j at iteration k, agent i will have $b_{ij}\zeta_j(k) = \kappa_j(k) - \kappa_{ij}(k)$. Thereby, agent i can use $\kappa_j(k) - \kappa_{ij}(k)$ to replace $b_{ij}\zeta_j(k)$ in equation (11). Then, agent i updates $\kappa_{ij}(k+1) = \kappa_j(k)$, according to equation (21). If from iteration k, the information packets from agent j drop for $a \geq 1$ iterations, i.e., for $m = k, k+1, \ldots, k+a-1$, agent i does not receive $\kappa_j(m)$, $\kappa_{ij}(m)$ remains unchanged until iteration $k+a$. At iteration $k+a$, when the information packet from agent j is successfully delivered, agent i receives $\kappa_j(k+a)$. By calculating the difference between $\kappa_j(k+a)$ and $\kappa_{ij}(k+a-1) = \kappa_j(k-1)$, agent i recovers $\sum_{m=0}^{a} b_{ij}\zeta_j(k+m)$, which is the sum of the information from agent j that agent i is supposed to use to update ζ_i from iteration k to iteration $k+a$. Thus, by using $\kappa_j(k) - \kappa_{ij}(k)$ to replace $b_{ij}\zeta_j(k)$ in equation (11), agent i avoids the lost of information from agent j.

According to Assumption [A6], each agent i sends identical information to its out-neighbors $j \in \mathcal{N}_i^-$. Therefore, with $\left|\mathcal{N}_i^+\right|$ in-neighbors, it should maintain $\left|\mathcal{N}_i^+\right| + 1$ auxiliary variables, which are:

$$\kappa_i\left(k\right) = \sum_{m=0}^{k} \frac{1}{\left|\mathcal{N}_i^-\right| + 1}\zeta_i\left(m\right), \tag{20}$$

$$\kappa_{ij}\left(k+1\right) = \begin{cases} \kappa_j\left(k\right), & \gamma_{ij}\left(k\right) = 1, \\ \kappa_{ij}\left(k\right), & \text{otherwise.} \end{cases}, \forall j \in \mathcal{N}_i^+. \tag{21}$$

The steps of the proposed algorithm are outlined as follows:

1) Step 1 - Initialization: In addition to equations (7) and (8), agent i initializes a group of auxiliary variables as:

$$\kappa_i\left(0\right) = \frac{1}{\left|\mathcal{N}_i^-\right| + 1}\zeta_i\left(0\right), \tag{22}$$

$$\kappa_{ij}\left(0\right) = 0, \forall j \in \mathcal{N}_i^+. \tag{23}$$

2) Step 2 - Information exchange: At the beginning of iteration k, each agent sends the current value of two variables to its out-neighbors, namely $\lambda_i(k)$ and $\kappa_i(k)$. As assumed in **[A4]**, each agent i generates a binary variable $\gamma_{ij}(k)$ for every agent $j \in \mathcal{N}_i^+(k)$, which indicate packet drops.

3) Step 3 - Local update: Based on the received $\lambda_j\left(k\right)$ and $\kappa_j(k)$, each agent performs the following updates locally:

a) Update price estimation:

$$\begin{aligned}\lambda_i\left(k+1\right) = &\lambda_i\left(k\right) + \sum_{j\in\mathcal{N}_i^+(k)} a_{ij}\left(k\right)\left(\lambda_j\left(k\right) - \lambda_i\left(k\right)\right) \\ &+ \eta\zeta_i\left(k\right),\end{aligned} \tag{24}$$

where the weight $a_{ij}(k)$ is calculated as:

$$a_{ij}(k) = \begin{cases} \frac{1}{\left|\mathcal{N}_i^+(k)\right|+1}, & \forall j \in \mathcal{N}_i^+(k), \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

b) Optimal response: Each agent optimally responds to $\lambda_i\left(k+1\right)$ by solving the sub-problem (6) locally.

c) Update power mismatch estimation: Each agent executes the following equation instead of equation (11):

$$\begin{aligned}\zeta_i\left(k+1\right) = &\frac{1}{\left|\mathcal{N}_i^-\right| + 1}\zeta_i\left(k\right) + \sum_{j\in\mathcal{N}_i^+(k)}\left(\kappa_j\left(k\right) - \kappa_{ij}\left(k\right)\right) \\ &+ \hat{P}_i\left(k\right).\end{aligned} \tag{26}$$

d) Update auxiliary variables: For an agent $i \in \mathcal{I}$, it updates the auxiliary variables $\kappa_i(k+1)$ and $\kappa_{ij}(k+1), \forall j \in \mathcal{N}_i^+$ based on equations (20) and (21).

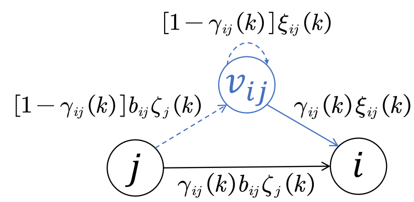


Fig. 3. The non-ideal communication link is augmented by a virtual buffering node, which is shown in blue. The virtual node buffers the lost information when packet drop occurs and release it to the receiving agent when the communication link resumes.

4) Step 4 - Termination check: After the local update of the variables, each agent checks the terminating criteria (14). If the terminating criteria are not satisfied simultaneously, set $k = k+1$ and return to Step 2.

Remark 1: We formulate the proposed algorithm following assumption **[A6]**, i.e., agent i sends identical messages to all its out-neighbors. Thus, agent i only needs to maintain one auxiliary variables κ_i to track the accumulated value of $b_{ji}\zeta_i$ it sends, for all $j \in \mathcal{N}_i^-$. However, assumption **[A6]** might not hold, i.e., agent i can send information to its out-neighbors with different weights. In this scenario, agent i needs to maintain a group of auxiliary variables, κ_{ji}, each for a chosen weight, to track the information it sends. In the following discussion, we use b_{ij} to denote the weights of ζ_j, without emphasizing the setting of weights.

Remark 2: Without packet drops, i.e., $\gamma_{ij}(k) \equiv 1, \forall (i,j) \in \mathcal{E}$, our proposed algorithm reduces to Algorithm 1. This can be concluded from the observation that given $\gamma_{ij}(k) \equiv 1$, there hold $\kappa_j(k) - \kappa_{ij}(k) \equiv b_{ij}\zeta_j(k)$ and $\kappa_{ij}(k+1) \equiv \kappa_j(k)$. Thus, our proposed algorithm can be regarded as a robustified extension of Algorithm 1 that can handle random packet drops.

Remark 3: The increase of system-overheads due to introducing auxiliary variables is insignificant. We analyze the system-overheads from the perspectives of communication burden and computational burden. Since the number of variables to be exchanged along a communication link remains the same and the proposed method does not require a retransmission, the communication burden is not increased. The computational burden is increased due to the introduction of auxiliary variables. The number of auxiliary variables maintained by agent i is limited by the number of its communication neighbors. The additional operations, namely equations (18) and (19), only involve linear combination, which can be easily handled by modern microprocessors. Furthermore, although the values of κ_i and κ_{ij} grow as the algorithm iterates, they are upper-bounded, since the values of $\zeta_i, i \in \mathcal{I}$ converges to zero. Thus, the increase of computational burden is insignificant.

B. Theoretical Proofs of Convergence and Optimality

The accumulating feature of the auxiliary variables makes it difficult to study the convergence of the proposed algorithm directly. Thus, we develop an equivalent form of the proposed method, based on which the convergence of the algorithm and the optimality of the solution are theoretically proved.

Let us take link (i,j) as an example. The proposed method is equivalent to adding a virtual buffering node to link (i,j), which is denoted in blue in Fig 3. The virtual buffering node stores the lost information when the packet from agent j drops, and releases the buffered information to agent i when link (i, j) works again. Denoting the buffered information by ξ_{ij}, given $\xi_{ij}(0) = 0$, equations (20), (21) and (26) can be equivalently written as:

$$\zeta_i(k+1) = \left(1 - \sum_{h\in\mathcal{N}_i^-} b_{hi}\right)\zeta_i(k) + \hat{P}_i(k) + \sum_{j\in\mathcal{N}_i^+(k)} \gamma_{ij}(k)\left(b_{ij}\zeta_j(k) + \xi_{ij}(k)\right), \quad (27)$$

$$\xi_{ij}(k+1) = \left(1 - \gamma_{ij}(k)\right)\left(\xi_{ij}(k) + b_{ij}\zeta_j(k)\right). \quad (28)$$

Remark 4: To keep the mathematical formulation simple and clear, we include a variable ξ_{ij} for any two nodes $i, j \in \mathcal{I}$. If $(i, j) \notin \mathcal{E}$, by letting $\gamma_{ij}(k) \equiv 0$ and $b_{ij} \equiv 0$, ξ_{ij} always equals to 0 and has no impact on the other variables.

Since the convergence of a linear homogeneous system can be easily analyzed by studying its eigenvalues, we first derive the linear homogenous form of equations (24), (27) and (28).

Theorem 1. *With a small enough step size η, equations (24), (27) and (28) can be reformulated into the following linear homogeneous system:*

$$\begin{bmatrix} \boldsymbol{\lambda}(k+1) \\ \boldsymbol{\zeta}(k+1) \\ \boldsymbol{\xi}(k+1) \end{bmatrix} = \boldsymbol{W}(\eta, k) \begin{bmatrix} \boldsymbol{\lambda}(k) \\ \boldsymbol{\zeta}(k) \\ \boldsymbol{\xi}(k) \end{bmatrix}, \quad (29)$$

$$\boldsymbol{W}(\eta, k) = \begin{bmatrix} \mathcal{A}(k) & \eta\mathbf{I}_N & \mathbf{0}_{N^2\times N} \\ \boldsymbol{K}(k)\left(\mathcal{A}(k) - \mathbf{I}_N\right) & \tilde{\boldsymbol{W}}_1(\eta, k) & \tilde{\boldsymbol{W}}_2(k) \\ \mathbf{0}_{N\times N^2} & \tilde{\boldsymbol{W}}_3(k) & \tilde{\boldsymbol{W}}_4(k) \end{bmatrix}, \quad (30)$$

where $\boldsymbol{\lambda}(k)$, $\boldsymbol{\zeta}(k)$ and $\boldsymbol{\xi}(k)$ are column vectors consisting of $\lambda_i(k)$, $\zeta_i(k)$ and $\xi_{ij}(k), \forall i, j \in \mathcal{I}$, respectively; $\mathcal{A} = [a_{ij}]$ is a weight matrix; $\mathbf{I}_N$ is an $N \times N$ identity matrix; $\mathbf{0}_{M\times N}$ is an $M \times N$ zero matrix; the definitions of $\boldsymbol{K}(k)$, $\tilde{\boldsymbol{W}}_1(k)$, $\tilde{\boldsymbol{W}}_2(k)$, $\tilde{\boldsymbol{W}}_3(k)$ and $\tilde{\boldsymbol{W}}_4(k)$ are provided in Appendix A.

Proof. Please see Appendix A. ■

Before studying the convergence of the linear homogeneous system (29), we first intuitively identify its stationary point. From equation (24) we infer that at the stationary point, for every $i \in \mathcal{I}$, there hold $\lambda_i = \lambda^*$ and $\zeta_i = 0$. According to (28), we infer that $\xi_{ij} = 0, \forall i, j \in \mathcal{I}$ given that $\zeta_i = 0, \forall i \in \mathcal{I}$. Thus, $\boldsymbol{\lambda}(k)$, $\boldsymbol{\zeta}(k)$ and $\boldsymbol{\xi}(k)$ converge to $\lambda^*\mathbf{1}_{\mathrm{N}}$, $\mathbf{0}_{\mathrm{N}}$ and $\mathbf{0}_{\mathrm{N}^2}$, respectively. In the presence of random packet drops, the communication topology varies across iterations. Thus, we study the convergence of the linear homogeneous system (29) with two probabilistic descriptions, which are defined as:

Definition 1 (Convergence in *Mean Square* Sense [29])**.** *System (29) converges in mean square sense if given $k \to \infty$, there hold $E\left[\left\|\boldsymbol{\lambda}(k) - \lambda^*\mathbf{1}_N\right\|_2^2\right] \to 0$, $E\left[\left\|\boldsymbol{\zeta}(k) - \mathbf{0}_N\right\|_2^2\right] \to 0$ and $E\left[\left\|\boldsymbol{\xi}(k) - \mathbf{0}_{N^2}\right\|_2^2\right] \to 0$.*

Definition 2 (Convergence in *Almost Sure* Sense [29])**.** *System (29) converges in almost sure sense if given $k \to \infty$, $[\boldsymbol{\lambda}(k); \boldsymbol{\zeta}(k); \boldsymbol{\xi}(k)] \to [\lambda^*\mathbf{1}_N; \mathbf{0}_N; \mathbf{0}_{N^2}]$ with probability 1.*

Remark 5: Definition 1 requires that the expectation of the mean square error between the stationary point and the point that system (29) converges to is zero. Definition 2 depicts that system (29) will eventually approach the stationary point, although the communication topology is time-varying.

The convergence of the proposed algorithm and the optimality of the solution are given by the following theorems.

Theorem 2. *Given a small enough step size η, system (29) converges in both mean square sense and almost sure sense.*

Proof. Please see Appendix B. ■

Theorem 3. *If problem (1) has at least one feasible solution, the proposed algorithm obtains the global optimal solution of problem (1) as it converges.*

Proof. Please see Appendix C. ■

Theorem 2 and Theorem 3 imply that in the presence of random packet drops, given long enough iterations, the proposed algorithm always converges to the optimum of problem (1).

IV. Case Study

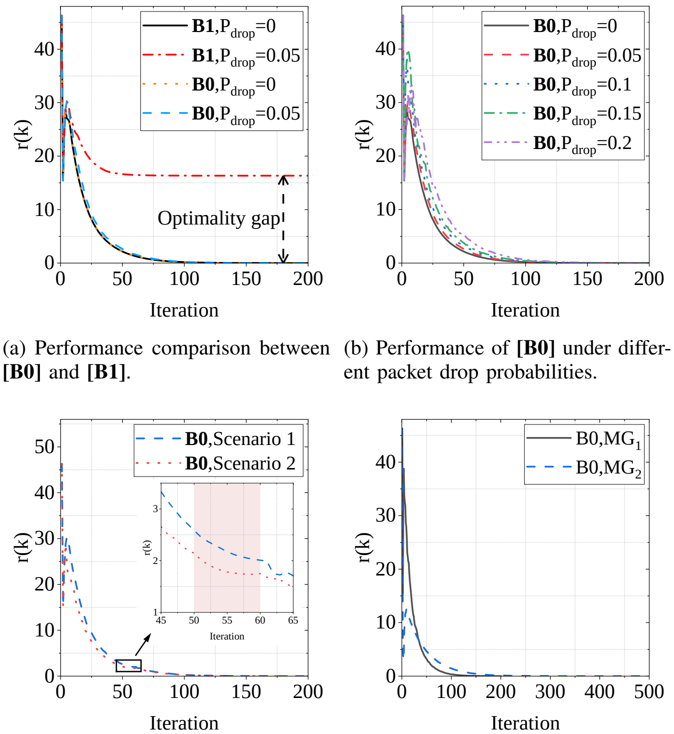
In this section, the effectiveness of the proposed method is tested with multiple experiments. We first validate that the proposed method can tackle the problem of random packet drops. We further compare the performance of the proposed method with the existing methods in the literature. The impacts of changing the structure of the communication topology and increasing the size of the system are also investigated. The involved methods are listed as follows:

- **[B0]** The method proposed in this paper to handle random packet drops over directed communication topologies.
- **[B1]** The conventional consensus-based algorithm, which is presented in [16] and is vulnerable to random packet drops.
- **[B2]** The corrective-based algorithm proposed in [22], which requires the communication topology to be undirected.
- **[B3]** The ratio-consensus-based algorithm with running sums proposed in [24], which works on directed communication topologies.

Without losing generality, we assume that the probability of packet drop is a constant, i.e., $p_{ij} = P_{drop}$. In each case, with a given P_{drop}, a sequence of communication topologies, $\{\mathcal{G}(k)\}$, is generated and adopted by all the algorithms.

A. Case 1: Effectiveness of the Proposed Method in Handling Random Packet Drops

In case 1, we verify the effectiveness of the proposed method in different scenarios of random packet drops. We first test **[B0]** with the same microgrid MG_1 and communication topology $\mathcal{G}_1$, as shown in Table I and Fig. 2a, respectively. The probability of packet drop is assumed to be 5%, i.e., $P_{drop} = 0.05$. The results of **[B0]** and **[B1]** are shown in Fig. 4a. When $P_{drop} = 0$, both **[B0]** and **[B1]** converge to the global optimum, as shown by the solid black line and the dotted orange line. The traces of residual are the same, since our proposed method **[B0]** reduces to **[B1]** when there



(a) Performance comparison between **[B0]** and **[B1]**.

(b) Performance of **[B0]** under different packet drop probabilities.

(c) Performance of **[B0]** in two extreme scenarios.

(d) Performance of **[B0]** in MG_1 and MG_2.

Fig. 4. The performance of the proposed method **[B0]** is compared with the conventional method **[B1]** with different P_{drop}. When $P_{drop} = 0$, the performances of **[B0]** and **[B1]** are the same. When random packet drops exist, **[B0]** can still obtain the optimum, while**[B1]** cannot. Our proposed method **[B0]** is effective with different P_{drop} and in different microgrids.

is no packet drop. With random packet drops, **[B1]** fails to converge to the optimal solution, as there exists an optimality gap between the dash-dotted red line and zero. Our proposed method successfully handles the random packet drops and converges to the global optimum, as shown by the dashed blue line. Since only the converged state of the algorithm is adopted and applied to the microgrid, the oscillation of the trajectories does not pose threats to the system. It is also observed that with a 5% probability of packet drop, the proposed algorithm **[B0]** converges at a similar rate as in the ideal scenario.

The presence of random packet drop might affect the convergence rate of the consensus-based algorithm. Thus, we test the proposed method **[B0]** under different P_{drop}. The results are presented in Fig. 4b. In general, as P_{drop} increases, the number of iterations needed for **[B0]** to converge also increases, while the convergence to the global optimum is always guaranteed. Thus, we conclude that our proposed method is effective under different packet drop probabilities. In the following experiments, the random packet drop probability is set to 0.1, if not specified.

To further show the effectiveness of the proposed method **[B0]**, we simulate two extreme scenarios:

1) Scenario 1: The information packets drop continuously on one communication link. To simulate this, beyond random packet drops, we manually block the information packet from agent 2 to agent 7 for 10 successive iterations, starting from the 50th iteration.
2) Scenario 2: The communication topology is not strongly connected at some iterations due to packet drop. To simulate this, beyond random packet drops, we manually block the information packet from agent 7 to agent 6 for 10 successive iterations, starting from the 50th iteration.

TABLE II
PARAMETERS OF AGENTS IN MG_2

Agent	Type	q ($/kW^2)	l ($/kW)	$P^{\max}$ (kW)	$P^{\min}$ (kW)	P^* (kW)
1	RES	-	-	42	0	42.0000
2	RES	-	-	37	0	37.0000
3	RES	-	-	46	0	46.0000
4	Diesel generator	0.00056	0	56	38	40.2256
5	Flexible load	-0.0072	0.8352	58	45	54.8713
6	Flexible load	-0.0084	1.0416	62	52	59.3183
7	Flexible load	-0.0076	0.8208	54	50	51.0360

TABLE III
PARAMETERS OF AGENTS IN MG_3

Agent	q ($/kW^2)	$P^{\max}$ (kW)	$P^{\min}$ (kW)	D_i (kW)	P^* (kW)
1	0.000547	88.0791	33.2450	63.5784	62.0347
2	0.000592	67.6062	38.6613	51.0647	57.3521
3	0.000560	80.9790	42.5112	63.5165	60.6086
4	0.000538	81.0776	49.3689	65.7591	63.1033
5	0.000506	67.6160	45.9689	58.2633	67.1744
6	0.000559	68.9019	44.8891	67.7741	60.8071
7	0.000557	87.3433	47.2230	62.1199	60.9959

The results are presented in Fig. 4c. From the 50th iteration, the convergence of **[B0]** is slightly slowed down, as the trajectories are flattened in the light red zone in Fig. 4c. This is because the connectivity of the communication topology is weakened when we manually block one communication link. After the 60th iteration, the trajectories become steeper as the blocked communication links resume. In both scenarios, **[B0]** converges to the global optimum solution, indicating its effectiveness in extreme scenarios.

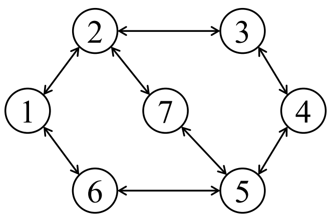
Considering that RES is taking an increasing share of the generation in microgrids, we introduce a microgrid MG_2, whose power generation is dominated by the RESs, to test the performance of the proposed method **[B0]**. The parameters of MG_2 are listed in Table II and the results are shown in Fig. 4d. With more RES in the microgrid, **[B0]** still converges to the global optimum, which indicates that our proposed method can be applied to microgrids with different generation portfolios.

B. Case 2: Performance Comparison with Existing Methods

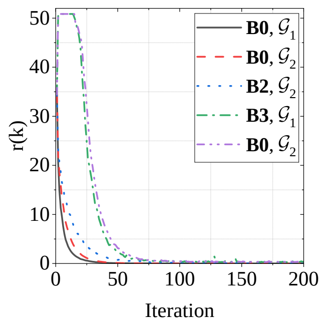
In Case 2, we compare the proposed method **[B0]** with two benchmarks, **[B2]** and **[B3]**. Since **[B2]** and **[B3]** are not designed for the cases with flexible loads, we adopt a new microgrid MG_3 with only diesel generators and list the parameters in Table III. Following the assumptions made in [22] and [24], we assign a virtual load D_i to each agent and modify the initialization of ζ_i to:

$$\zeta_i(0) = \begin{cases} -P_i(0) + D_i & , \forall i \in \mathcal{I}_{\mathrm{G}}, \\ P_i(0) + D_i & , \forall i \in \mathcal{I}_{\mathrm{L}}. \end{cases} \tag{31}$$

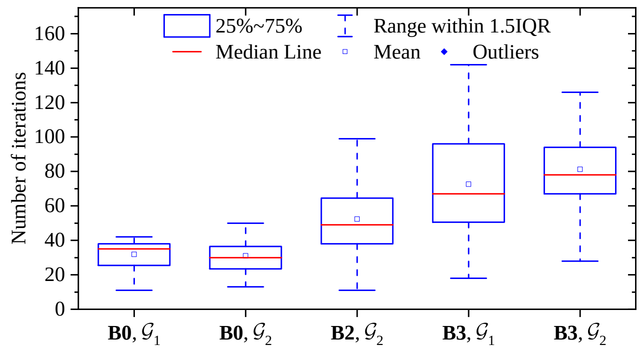
This article has been accepted for publication in IEEE Transactions on Smart Grid. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSG.2023.3241653



(a) Communication topology $\mathcal{G}_2$.

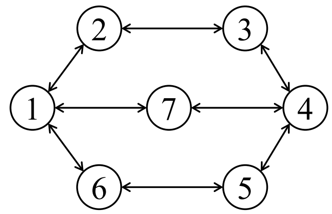


(b) Performance comparison.

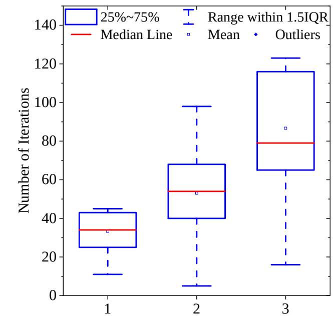


(c) Number of iterations needed for each benchmarks to converge over different communication topologies.

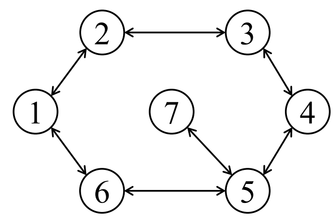
Fig. 5. The undirected communication topology $\mathcal{G}_2$ used in Case 2 and the performance of the three algorithms, which are all able to handle random packet drops. The proposed algorithm **[B0]** achieves faster convergence to the global optimum than the other benchmarks over both undirected or directed communication topology. Given the same convergence criterion and 100 randomly generated initial states, **[B0]** takes fewer iterations to converge, compared to the other benchmarks.



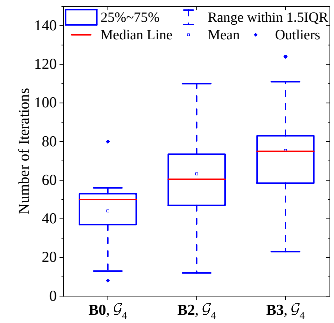
(a) Communication topology $\mathcal{G}_3$.



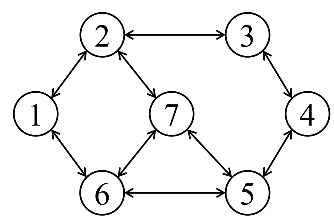
(b) Performance comparison.



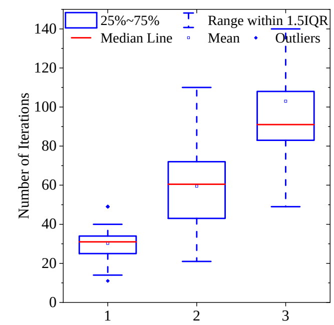
(c) Communication topology $\mathcal{G}_4$.



(d) Performance comparison.



(e) Communication topology $\mathcal{G}_5$.



(f) Performance comparison.

Fig. 6. The undirected communication topologies used to investigate the impact of changing the structure of the communication topology and the performance of **[B0]**, **[B2]** and **[B3]** over these topologies. The changing of communication topology structure does not affect the convergence of the proposed method, and the proposed method takes fewer iterations to converge, compared to the existing methods. In general, the number of iterations needed decreases, as the number of communication links in the communication topology increases.

Since **[B2]** can only be applied over undirected communication topologies, besides the directed graph $\mathcal{G}_1$ in Fig. 2a, we introduce an undirected graph $\mathcal{G}_2$, which is shown in Fig. 5a. The performance of **[B0]**, **[B2]** and **[B3]** are tested on both $\mathcal{G}_1$ and $\mathcal{G}_2$, with $P_{drop} = 0.1$. The results are presented in Fig. 5b. Over the undirected communication topology $\mathcal{G}_2$, all the algorithms can obtain the optimal solution in the presence of random packet drops. However, only **[B0]** and **[B3]** can converge to the global optimum over a directed communication topology $\mathcal{G}_1$.

In Fig. 5b, our proposed algorithm **[B0]** converges at a faster rate than **[B2]** and **[B3]** over both the undirected and directed communication topologies. To further test the convergence performance of **[B0]**, **[B2]** and **[B3]**, we further compare the number of iterations needed for each method to converge, with given initial states. The following terminating criterion is used:

$$\left| \sum_{i \in \mathcal{I}} P_i(k) - \sum_{i \in \mathcal{I}} D_i \right| \leq \epsilon_P, \tag{32}$$

where ϵ_P is the terminating threshold.

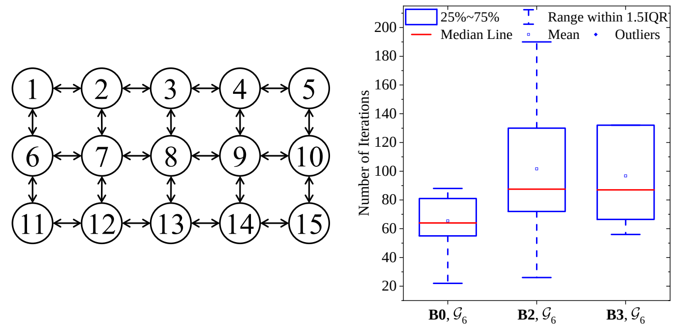
Based on the parameters of MG_3, we randomly generate 100 initial states, from which **[B0]**, **[B2]** and **[B3]** start to iterate. The results are presented through box plots in Fig. 5c. Over both $\mathcal{G}_1$ and $\mathcal{G}_2$, our proposed method always converges within 50 iterations, with an average of 31.84 iterations and 30.97 iterations, respectively. With the same initial states, **[B2]** takes an average of 52.35 iterations to converge over the undirected communication topology $\mathcal{G}_2$. Over $\mathcal{G}_1$ and $\mathcal{G}_2$, the average numbers of iterations needed by **[B3]** to converge are 72.59 iterations and 81.28 iterations, respectively. Compared with **[B2]** and **[B3]**, our proposed method **[B0]** takes fewer iterations to converge, implying a greater potential for real-world implementations.

C. Case 3: Impact of the Communication Topology

The structure of the communication topology will have an impact on the number of iterations needed for the consensus-

TABLE IV
PARAMETERS OF AGENTS IN MG_4

Agent	q ($\$/kW^2$)	P^{max} (kW)	P^{min} (kW)	D_i (kW)	P^* (kW)
1	0.00041	64.4126	41.7450	65.2739	62.4056
2	0.00047	65.9377	49.6478	73.4669	54.7055
3	0.00043	61.3586	45.6827	49.7420	59.7226
4	0.00042	79.2455	45.7484	58.3376	61.1313
5	0.00046	71.8018	40.7513	50.3982	55.6684
6	0.00042	60.4568	43.4763	44.1542	60.4568
7	0.00042	76.2020	31.5642	41.5480	60.6567
8	0.00040	70.4250	34.6561	61.8493	63.7516
9	0.00047	77.2366	48.2115	52.5515	54.9844
10	0.00044	66.3277	49.1088	68.8069	57.7933
11	0.00049	63.6817	41.2425	53.4780	52.2083
12	0.00050	62.7240	46.4255	75.8791	51.6288
13	0.00047	67.4327	34.4361	61.9511	55.2180
14	0.00049	61.8293	31.7159	54.3683	52.9330
15	0.00045	73.1841	34.2260	48.1706	56.7112



(a) Communication topology $\mathcal{G}_6$. (b) Performance comparison.

Fig. 7. The communication topology $\mathcal{G}_6$ with 15 agents and the performance of **[B0]**, **[B2]** and **[B3]** over $\mathcal{G}_6$. With 100 randomly generated initial states, the number of iterations needed for **[B0]** to converge to the global optimum is fewer than **[B2]** and **[B3]**, implying that our proposed method outperforms the benchmarks in a larger system.

based algorithm to converge. To study the impact of communication topology change, we introduce three new topologies $\mathcal{G}_3$, $\mathcal{G}_4$ and $\mathcal{G}_5$, which are adapted from $\mathcal{G}_2$. As shown in Fig. 6a, Fig 6c and Fig 6e, the difference between the three communication topologies and $\mathcal{G}_2$ are outlined as follows:

1) $\mathcal{G}_3$: The number of communication links are the same as $\mathcal{G}_2$, while the agents that communicate with agent 7 are different.
2) $\mathcal{G}_4$: The communication link between agent 2 and agent 7 is removed, compared to $\mathcal{G}_2$.
3) $\mathcal{G}_5$: One communication link is added between agent 6 and agent 7, compared to $\mathcal{G}_2$.

We test **[B0]**, **[B2]** and **[B3]** over $\mathcal{G}_3$, $\mathcal{G}_4$ and $\mathcal{G}_5$ with the same 100 initial states used over $\mathcal{G}_2$. The results are presented in Fig. 6b, Fig 6d and Fig. 6f. All three methods can converge to the global optimum over these communication topologies, under the impact of random packet drop. Similar to the case over $\mathcal{G}_2$, compared to **[B2]** and **[B3]**, our proposed method **[B0]** takes fewer iterations to converge over $\mathcal{G}_3$, $\mathcal{G}_4$ and $\mathcal{G}_5$. In general, the convergence of the proposed method is not affected when the structure of the communication topology is changed. With the same number of communication links, the number of iteration that **[B0]** takes over $\mathcal{G}_3$ falls in a similar range as over $\mathcal{G}_2$. However, with less communication links, **[B0]** takes more iterations to converge over $\mathcal{G}_4$ than over $\mathcal{G}_2$. In contrast, the number of iterations needed over $\mathcal{G}_5$ is the least, since $\mathcal{G}_5$ has more communication links than other communication topologies.

D. Case 4: Performance Comparison in a Larger Microgrid

To demonstrate that our proposed method **[B0]** still outperforms the benchmarks as the size of the microgrid increases, we introduce a microgrid system with 15 agents, termed as MG_4, whose parameters are shown in Table IV. The communication topology, termed as $\mathcal{G}_6$, is shown in Fig. 7a. Similarly, we randomly generate 100 initial points for MG_4 and test the performance of **[B0]**, **[B2]** and **[B3]**. The results are presented in Fig. 7b. As the size of the microgrid increases, all three methods need more iterations to converge to the global optimum. In MG_4, our proposed method always converges within 88 iterations, with an average of 65.39 iterations. However, **[B2]** and **[B3]** take an average of 101.65 iterations and 96.76 iterations to converge, respectively. Thus, **[B0]** outperforms **[B2]** and **[B3]** in MG_4. According to [30], we can assume that at each iteration, the information exchange and the local computation both take 1 second. Therefrom, a maximum of 88 iterations to converge implies that the optimal solution is obtained by the proposed method within 3 minutes. Thus, our proposed method is capable to be applied in real implementations.

V. CONCLUSION

To promote the local accommodation of RES generation, we study the energy management problem of a microgrid and its decentralized solving with the consensus-based algorithm. Although the conventional consensus algorithm converges to the global optimum with ideal communication, it might converge to an infeasible solution when random packet drop exist. Motivated by this observation, we analyze the impact of the random packet drops resulting from non-ideal communication links and identify the critical information that needs to be recovered. Then, we propose a novel algorithm with auxiliary variables to compensate for the information loss brought by random packet drops. The convergence of the proposed algorithm and the optimality of the obtained solution are theoretically proved, by augmenting the original communication topology with virtual buffering nodes. Through multiple numerical experiments, we validate that the proposed method can obtain the global optimal solution in the presence of random packet drops. We also demonstrate that the proposed algorithm converges at a faster rate than the benchmarks. The impact of the structure of the communication topology and the scale of the microgrid system is also investigated, which shows the great potential of the proposed method for real-world implementations.

APPENDIX A
PROOF OF THEOREM 1

Firstly, we reformulate $\hat{P}_i(k)$ as a function of $\lambda_i(k+1)$ and $\lambda_i(k)$ with the following lemma.

Lemma 1. *Given that the step size η is small enough, the local optimal response to $\lambda_i(k+1)$, namely equation (6), can be linearized around the given point $P_i(k)$ [21].*

The first order optimality conditions for $i \in \mathcal{I}_{\mathrm{G}}$ in (6) are:

$$\lambda_i(k) = C_i^{'}(P_i(k)). \tag{33}$$

Lemma 1 implies that by lifting the power limit constraint (1c) and taking the first order Taylor expansion around a given point $P_i(k)$, we can get:

$$\lambda_i(k) + \Delta\lambda_i(k) = C_i^{'}(P_i(k)) + C_i^{''}(P_i(k))\Delta P_i(k) + R_1, \tag{34}$$

where R_1 is the remainder.

Assuming that $C_i(\cdot)$ is three times differentiable at point $P_i(k)$, the remainder can be expressed as:

$$R_1 = \frac{C_i^{'''}(\tilde{P}_i)}{2}\Delta P_i^2(k), \tag{35}$$

where $C_i^{'''}(\cdot)$ is the third order derivative of the generation cost function; $\tilde{P}_i$ is some value between $P_i(k)$ and $P_i(k)+\Delta P_i(k)$.

For the quadratic generation cost functions that are widely used (e.g., in [12]), $C_i^{'''}(\tilde{P}_i) = 0$ always holds, implying that $R_1 = 0$. Thus, the following equations hold:

$$\begin{aligned}\lambda_i(k) + \Delta\lambda_i(k) &= C_i^{'}(P_i(k)) + C_i^{''}(P_i(k))\Delta P_i(k) \\ \Rightarrow \Delta P_i(k) &= \left[C_i^{''}(P_i(k))\right]^{-1}\Delta\lambda_i(k).\end{aligned} \tag{36}$$

If third order derivatives of the quadratic cost functions and utility functions are non-zero, equation (36 becomes an approximation. There exists an upper bound M such that:

$$\left|C_i^{'''}(\tilde{P}_i)\right| \le M, \tag{37}$$

where M is a positive value. Thus, the upper bound of the remainder can be written as:

$$|R_1| \le \frac{M}{2}\Delta P_i^2(k). \tag{38}$$

To ensure the accuracy of the approximation, a small enough step size η could be selected to ensure that $\Delta P_i^2(k) = [P_i(k+1) - P_i(k)]^2$ is small enough, such that the absolute value of the remainder $|R_1|$ is within the acceptable range.

Considering the power limits, for $i \in \mathcal{I}_{\mathrm{G}}$, the change of power at iteration k can be expressed as:

$$\begin{aligned}\hat{P}_i(k) = P_i(k) - P_i(k+1) = K_i(k)\left(\lambda_i(k+1) - \lambda_i(k)\right), \\ -\left[C_i^{''}(P_i(k))\right]^{-1} < K_i(k) < 0,\end{aligned} \tag{39}$$

where $K_i(k)$ is the coefficient that relates $\hat{P}_i(k)$ to $\lambda_i(k+1) - \lambda_i(k)$.

Similarly, for $i \in \mathcal{I}_{\mathrm{L}}$, we can express the change of power at iteration k as:

$$\begin{aligned}\hat{P}_i(k) = P_i(k+1) - P_i(k) = K_i(k)\left(\lambda_i(k+1) - \lambda_i(k)\right), \\ \left[U_i^{''}(P_i(k))\right]^{-1} < K_i(k) < 0.\end{aligned} \tag{40}$$

Assuming $K_i(k) = 0, \forall i \in \mathcal{I}_{\mathrm{R}}$, we can conclude that:

$$\hat{P}_i(k) = K_i(k)\left(\lambda_i(k+1) - \lambda_i(k)\right), \forall i \in \mathcal{I}. \tag{41}$$

With equation (41), we define a diagonal matrix $\boldsymbol{K}(k) = \mathrm{diag}\{K_i(k)\}$ and reformulate the update rules (24), (27) and (28) into the linear homogeneous system (29). The matrices $\tilde{\boldsymbol{W}}_1(k)$, $\tilde{\boldsymbol{W}}_2(k)$, $\tilde{\boldsymbol{W}}_3(k)$ and $\tilde{\boldsymbol{W}}_4(k)$ are defined as:

$$\tilde{\boldsymbol{W}}_1(\eta, k) = \eta\boldsymbol{K}(k) + \boldsymbol{\Gamma}(k) \circ \mathcal{B}, \tag{42}$$

where $\boldsymbol{\Gamma}(k) = [\gamma_{ij}(k)]$ is a matrix indicating the packet drops at iteration k; $\circ$ denotes the Hadamard product of two matrices;

$$\tilde{\boldsymbol{W}}_2(k) = \begin{bmatrix} \boldsymbol{\gamma}_1(k) & \mathbf{0}_{1\times\mathrm{N}} & \cdots & \mathbf{0}_{1\times\mathrm{N}} \\ \mathbf{0}_{1\times\mathrm{N}} & \boldsymbol{\gamma}_2(k) & \cdots & \mathbf{0}_{1\times\mathrm{N}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times\mathrm{N}} & \mathbf{0}_{1\times\mathrm{N}} & \cdots & \boldsymbol{\gamma}_{\mathrm{N}}(k) \end{bmatrix}, \tag{43}$$

where $\boldsymbol{\gamma}_i(k)$ denotes the ith row of $\boldsymbol{\Gamma}(k)$;

$$\tilde{\boldsymbol{W}}_3(k) = \begin{bmatrix} \mathrm{diag}\{(1 - \boldsymbol{\gamma}_1(k)) \odot \boldsymbol{b}_1\} \\ \vdots \\ \mathrm{diag}\{(1 - \boldsymbol{\gamma}_{\mathrm{N}}(k)) \odot \boldsymbol{b}_{\mathrm{N}}\} \end{bmatrix}, \tag{44}$$

where $\mathrm{diag}\{(1-\boldsymbol{\gamma}_i(k))\odot\boldsymbol{b}_i\}$ denotes a $N\times N$ diagonal matrix with elements of $(1 - \boldsymbol{\gamma}_i(k)) \odot \boldsymbol{b}_i$ as its diagonal entries; $\odot$ denotes the component-wise multiplication of two vectors; $\boldsymbol{b}_i$ is the ith row of $\mathcal{B}$;

$$\tilde{\boldsymbol{W}}_4(k) = \begin{bmatrix} 1 - \gamma_{11}(k) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 - \gamma_{\mathrm{NN}}(k) \end{bmatrix}, \tag{45}$$

where the diagonal entries are $1 - \gamma_{ij}(k),\ i, j = 1, \cdots N.$ ■

APPENDIX B
PROOF OF THEOREM 2

We prove the convergence of the proposed algorithm with the matrix perturbation theory. Given that η is small enough, we regard $\boldsymbol{W}(\eta, k)$ as $\boldsymbol{W}(0, k)$ perturbed by $\eta\boldsymbol{\Delta}$, i.e.,

$$\boldsymbol{W}(\eta, k) = \boldsymbol{W}(0, k) + \eta\boldsymbol{\Delta}, \tag{46}$$

where $\boldsymbol{\Delta}(k)$ is defined as:

$$\boldsymbol{\Delta}(k) = \begin{bmatrix} \mathbf{0}_{\mathrm{N}\times\mathrm{N}} & \mathbf{I}_{\mathrm{N}} & \mathbf{0}_{\mathrm{N}\times\mathrm{N}^2} \\ \mathbf{0}_{\mathrm{N}\times\mathrm{N}} & \boldsymbol{K}(k) & \mathbf{0}_{\mathrm{N}\times\mathrm{N}^2} \\ \mathbf{0}_{\mathrm{N}^2\times\mathrm{N}} & \mathbf{0}_{\mathrm{N}^2\times\mathrm{N}} & \mathbf{0}_{\mathrm{N}^2\times\mathrm{N}^2} \end{bmatrix}. \tag{47}$$

To study the eigenvalues of $\boldsymbol{W}(\eta, k)$, we first study the eigenvalues of $\boldsymbol{W}(0, k)$ and then calculate the eigenvalue derivatives with respect to η. The matrix $\boldsymbol{W}(0, k)$ is a lower triangular block matrix whose diagonal blocks are $\mathcal{A}(k)$ and $\tilde{\boldsymbol{W}}(0, k) = \left[\tilde{\boldsymbol{W}}_1(0, k), \tilde{\boldsymbol{W}}_2(k); \tilde{\boldsymbol{W}}_3(k), \tilde{\boldsymbol{W}}_4(k)\right]$.

According to definition (10), the elements of $\mathcal{A}(k)$ are non-negative and the sum of each row of $\mathcal{A}(k)$ equals to 1. Thus, $\mathcal{A}(k)$ is a row-stochastic matrix for every iteration k. According to definitions (42) and (44), for the first N columns of $\tilde{\boldsymbol{W}}(0, k)$, the components of the jth column are non-negative and sum up to $\sum_{i=1}^{N} \gamma_{ij}b_{ij} + \sum_{i=1}^{N}(1 - \gamma_{ij})b_{ij} = \sum_{i=1}^{N} b_{ij}$. Since definition (12) implies that $\sum_{i=1}^{N} b_{ij} = 1$, the column sums of the first N columns of $\tilde{\boldsymbol{W}}(0, k)$ equal to 1. According to definitions (43) and (45), for the rest of the columns of $\tilde{\boldsymbol{W}}(0, k)$, there are two non-zero components γ_{ij} and $(1 - \gamma_{ij})$, which are positive and sum up to 1. Thus, the elements of $\tilde{\boldsymbol{W}}(0,k)$ are non-negative and the sum of each column of $\tilde{\boldsymbol{W}}(0,k)$ equals to 1, implying that $\tilde{\boldsymbol{W}}(0,k)$ is column-stochastic for every iteration k.

As row-stochastic or column-stochastic matrices, both $\mathcal{A}(k)$ and $\tilde{\boldsymbol{W}}(0,k)$ have one eigenvalue with modulus 1 and the rest of the eigenvalues lying in the open unit disk. Since the eigenvalues of a lower triangular block matrix are the eigenvalues of the diagonal blocks, $\boldsymbol{W}(0,k)$ has two eigenvalues with modulus 1 and the rest of the eigenvalues lying in the open unit disk, i.e.,

$$1=|\sigma_1|=|\sigma_2|>|\sigma_3|\ldots\geq|\sigma_{2\mathrm{N}+\mathrm{N}^2}|. \quad (48)$$

We denote the right and left eigenvectors corresponding to eigenvalue σ_i as $\boldsymbol{\nu}_i$ and $\boldsymbol{\mu}_i$, which satisfy:

$$\begin{cases}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{W}(0,k)=\sigma_i\boldsymbol{\mu}_i^{\mathrm{T}}, \boldsymbol{W}(0,k)\boldsymbol{\nu}_i=\sigma_i\boldsymbol{\nu}_i\\ \boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\nu}_i=1, \boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\nu}_j=0\end{cases}. \quad (49)$$

For $|\sigma_1|=|\sigma_2|=1$, the corresponding right and left eigenvectors that satisfy equation (49) are:

$$\boldsymbol{\mu}_1=\begin{bmatrix}-\boldsymbol{K}(k)\mathbf{1}_{\mathrm{N}}\\ \mathbf{1}_{\mathrm{N}+\mathrm{N}^2}\end{bmatrix}, \boldsymbol{\mu}_2=\begin{bmatrix}\frac{1}{\mathbf{1}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{x}_1}\boldsymbol{x}_1\\ \mathbf{0}_{\mathrm{N}+\mathrm{N}^2}\end{bmatrix}, \quad (50)$$

$$\boldsymbol{\nu}_1=\begin{bmatrix}\mathbf{0}_{\mathrm{N}}\\ \frac{1}{\mathbf{1}_{\mathrm{N}+\mathrm{N}^2}^{\mathrm{T}}\boldsymbol{y}_1}\boldsymbol{y}_1\end{bmatrix}, \boldsymbol{\nu}_2=\begin{bmatrix}\mathbf{1}_{\mathrm{N}}\\ \frac{\hat{K}(k)}{\mathbf{1}_{\mathrm{N}+\mathrm{N}^2}^{\mathrm{T}}\boldsymbol{y}_1}\boldsymbol{y}_1\end{bmatrix}, \quad (51)$$

where $\mathbf{1}_{\mathrm{N}}$ is an N dimensional column vector of ones; $\mathbf{0}_{\mathrm{N}}$ is an N-dimensional column vector of zeros; $\boldsymbol{a}^{\mathrm{T}}$ denotes the transpose of $\boldsymbol{a}$; $\boldsymbol{x}_1$ is the left eigenvector of $\mathcal{A}(k)$ corresponding to eigenvalue 1; $\boldsymbol{y}_1$ is the right eigenvector of $\tilde{\boldsymbol{W}}(0,k)$ corresponding to eigenvalue 1; $\hat{K}(k)$ is the sum of all $K_i(k)$.

According to [31], given $\eta=0$, the eigenvalue derivatives with respect to η are:

$$\frac{\partial\sigma_i}{\partial\eta}=\boldsymbol{\mu}_i^{\mathrm{T}}\left.\frac{d\boldsymbol{W}(\eta,k)}{d\eta}\right|_{\eta=0}\boldsymbol{\nu}_i=\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Delta}(k)\boldsymbol{\nu}_i. \quad (52)$$

Thus, the derivatives of $|\sigma_1|=|\sigma_2|=1$ are:

$$\frac{\partial\sigma_1}{\partial\eta}=\begin{bmatrix}\frac{1}{\mathbf{1}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{x}_1}\boldsymbol{x}_1^{\mathrm{T}} & \mathbf{1}_{\mathrm{N}^2}^{\mathrm{T}}\end{bmatrix}\boldsymbol{\Delta}(k)\begin{bmatrix}\mathbf{0}_{\mathrm{N}}\\ \frac{1}{\mathbf{1}_{\mathrm{N}}^{\mathrm{T}}y_1}y_1\end{bmatrix}=0, \quad (53)$$

$$\begin{aligned}\frac{\partial\sigma_2}{\partial\eta}&=\begin{bmatrix}-\mathbf{1}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{K}(k) & \mathbf{0}_{\mathrm{N}+\mathrm{N}^2}^{\mathrm{T}}\end{bmatrix}\boldsymbol{\Delta}(k)\begin{bmatrix}\mathbf{1}_{\mathrm{N}}\\ \frac{\hat{K}(k)}{\mathbf{1}_{\mathrm{N}+\mathrm{N}^2}^{\mathrm{T}}\boldsymbol{y}_1}\boldsymbol{y}_1\end{bmatrix}\\ &=\frac{\hat{K}(k)}{\mathbf{1}_{\mathrm{N}+\mathrm{N}^2}^{\mathrm{T}}\hat{\boldsymbol{x}}_1\cdot\mathbf{1}_{\mathrm{N}+\mathrm{N}^2}^{\mathrm{T}}\boldsymbol{y}_1}\hat{\boldsymbol{x}}_1^{\mathrm{T}}\boldsymbol{y}_1,\end{aligned} \quad (54)$$

where $\hat{\boldsymbol{x}}_1$ is an $N+N^2$ dimensional vector with $\boldsymbol{x}_1$ as the first N elements. Since $\boldsymbol{y}_1$ is the right eigenvector of $\tilde{\boldsymbol{W}}(0,k)$ corresponding to eigenvalue 1, the components of $\boldsymbol{y}_1$ are non-negative. Thus, $\mathbf{1}_{\mathrm{N}+\mathrm{N}^2}^{\mathrm{T}}\hat{\boldsymbol{x}}_1\cdot\mathbf{1}_{\mathrm{N}+\mathrm{N}^2}^{\mathrm{T}}\boldsymbol{y}_1$ and $\hat{\boldsymbol{x}}_1^{\mathrm{T}}\boldsymbol{y}_1$ are positive real scalars. According to equations (39), (40) and (41), $\hat{K}(k)<0$ holds for every iteration k. Thus, from equation (54), $\frac{\partial\sigma_2}{\partial\eta}<0$. Thereby, there exists a small positive constant $\epsilon_1>0$ such that $|\sigma_2|<|\sigma_1|=1$ if $\eta<\epsilon_1$. Since the eigenvalues are continuous functions of η, there exists another small positive constant $\epsilon_2>0$ such that the rest of the eigenvalues remain in the open unit disk if $\eta<\epsilon_2$. Thus, letting $\eta<\min(\epsilon_1,\epsilon_2)$, $\boldsymbol{W}(\eta,k)$ has only one eigenvalue with modulus 1 and the others lie in the open unit disk, i.e.,

$$1=|\sigma_1|>|\sigma_2|\ldots\geq|\sigma_{2\mathrm{N}+\mathrm{N}^2}|. \quad (55)$$

In the following discussion, we assume that $\eta<\min(\epsilon_1,\epsilon_2)$ and omit η[3].

According to (24), (27) and (28), at a stationary point, λ_i, ζ_i and ξ_{ij} should converge to λ^*, 0 and 0, respectively. Inspired by [21], we define the consensus error as:

$$\boldsymbol{e}(k)=[\boldsymbol{\lambda}(k);\boldsymbol{\zeta}(k);\boldsymbol{\xi}(k)]-[\lambda^*\mathbf{1}_{\mathrm{N}};\mathbf{0}_{\mathrm{N}};\mathbf{0}_{\mathrm{N}^2}]. \quad (56)$$

Based on the row-stochastic nature of $\mathcal{A}(k)$, we have:

$$\begin{aligned}\boldsymbol{e}(k+1)&=\begin{bmatrix}\boldsymbol{\lambda}(k+1)\\ \boldsymbol{\zeta}(k+1)\\ \boldsymbol{\xi}(k+1)\end{bmatrix}-\begin{bmatrix}\lambda^*\mathbf{1}_{\mathrm{N}}\\ \mathbf{0}_{\mathrm{N}}\\ \mathbf{0}_{\mathrm{N}^2}\end{bmatrix}\\ &=\boldsymbol{W}(k)\begin{bmatrix}\boldsymbol{\lambda}(k)\\ \boldsymbol{\zeta}(k)\\ \boldsymbol{\xi}(k)\end{bmatrix}-\begin{bmatrix}\lambda^*\mathbf{1}_{\mathrm{N}}\\ \mathbf{0}_{\mathrm{N}}\\ \mathbf{0}_{\mathrm{N}^2}\end{bmatrix}=\boldsymbol{W}(k)\boldsymbol{e}(k),\end{aligned} \quad (57)$$

which implies that the evolution of $\boldsymbol{e}(k)$ can be described by the same linear homogeneous system as (29).

Defining a new vector $\hat{\boldsymbol{e}}(k)=\boldsymbol{e}(k)\otimes\boldsymbol{e}(k)$, where $\otimes$ is the Kronecker product, we have:

$$\begin{aligned}\hat{\boldsymbol{e}}(k+1)&=(\boldsymbol{W}(k)\boldsymbol{e}(k))\otimes(\boldsymbol{W}(k)\boldsymbol{e}(k))\\ &=(\boldsymbol{W}(k)\otimes\boldsymbol{W}(k))\hat{\boldsymbol{e}}(k).\end{aligned} \quad (58)$$

Thus, denoting the expectation as $E[\cdot]$, we can get:

$$E[\hat{\boldsymbol{e}}(k+1)|\hat{\boldsymbol{e}}(k)]=E[\boldsymbol{W}(k)\otimes\boldsymbol{W}(k)]\hat{\boldsymbol{e}}(k). \quad (59)$$

As assumed in **[A4]**, the sequences $\{\boldsymbol{W}(k)\}$, and thereby $\{\boldsymbol{W}(k)\otimes\boldsymbol{W}(k)\}$, are independent and identically distributed. Thus, we omit the time index k hereafter and derive the following equation from equation (59):

$$E[\hat{\boldsymbol{e}}(k+1)|\hat{\boldsymbol{e}}(k)]=E[\boldsymbol{W}\otimes\boldsymbol{W}]^k\hat{\boldsymbol{e}}(0). \quad (60)$$

Since the eigenvalues of $\boldsymbol{W}\otimes\boldsymbol{W}$ are $\sigma_i\sigma_j, i,j=1,\ldots,2\mathrm{N}+\mathrm{N}^2$ [35], defining $\varsigma_m=\sigma_i\sigma_j$, according to (55), we have:

$$1=|\varsigma_1|>|\varsigma_2|\geq|\varsigma_3|\geq\ldots\geq|\varsigma_{(2\mathrm{N}+\mathrm{N}^2)^2}|. \quad (61)$$

The right and left eigenvectors of $\boldsymbol{W}\otimes\boldsymbol{W}$ corresponding to ς_1 are $\boldsymbol{r}_1=\boldsymbol{\rho}_1\otimes\boldsymbol{\rho}_1$, $\boldsymbol{l}_1^{\mathrm{T}}=\boldsymbol{\omega}_1^{\mathrm{T}}\otimes\boldsymbol{\omega}_1^{\mathrm{T}}$, where $\boldsymbol{\rho}_1$ and $\boldsymbol{\omega}_1$ are the right and left eigenvectors of $\boldsymbol{W}$, respectively [36]. Considering the requirements in (49), we select $\boldsymbol{\rho}_1$ and $\boldsymbol{\omega}_1$ as:

$$\boldsymbol{\rho}_1=\begin{bmatrix}\mathbf{1}_{\mathrm{N}}\\ \mathbf{0}_{\mathrm{N}+\mathrm{N}^2}\end{bmatrix}, \boldsymbol{\omega}_1^{\mathrm{T}}=\frac{1}{\hat{K}}\begin{bmatrix}\mathbf{1}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{K} & -\mathbf{1}_{\mathrm{N}+\mathrm{N}^2}^{\mathrm{T}}\end{bmatrix}. \quad (62)$$

There exists a non-singular matrix $\boldsymbol{S}$ such that

$$\boldsymbol{S}^{-1}E[\boldsymbol{W}\otimes\boldsymbol{W}]\boldsymbol{S}=\begin{bmatrix}1 & \mathbf{0}_{\mathrm{N}^2+2\mathrm{N}-1}^{\mathrm{T}}\\ \mathbf{0}_{\mathrm{N}^2+2\mathrm{N}-1} & \boldsymbol{J}\end{bmatrix}, \quad (63)$$

where $\boldsymbol{J}$ is the Jordan block that corresponds to the eigenvalues inside the open unit disk. Given that $k\to\infty$, we have:

$$E[\boldsymbol{W}\otimes\boldsymbol{W}]^k=\boldsymbol{S}\begin{bmatrix}1 & \mathbf{0}_{\mathrm{N}^2+2\mathrm{N}-1}^{\mathrm{T}}\\ \mathbf{0}_{\mathrm{N}^2+2\mathrm{N}-1} & \boldsymbol{J}^k\end{bmatrix}\boldsymbol{S}^{-1}\to\boldsymbol{r}_1\boldsymbol{l}_1^{\mathrm{T}}. \quad (64)$$

[3]The step size η can be time-varying, e.g, η can be based on the states of the agents [32], [33] or be vanishing [34]. To keep the mathematical formulation clear and easy to understand, we adopt constant step size in this paper.

Since $\boldsymbol{r}_1\boldsymbol{l}_1^{\mathrm{T}} = (\boldsymbol{\rho}_1 \otimes \boldsymbol{\rho}_1)\left(\boldsymbol{\omega}_1^{\mathrm{T}} \otimes \boldsymbol{\omega}_1^{\mathrm{T}}\right) = \boldsymbol{\rho}_1\boldsymbol{\omega}_1^{\mathrm{T}} \otimes \boldsymbol{\rho}_1\boldsymbol{\omega}_1^{\mathrm{T}}$, we first compute $\boldsymbol{\rho}_1\boldsymbol{\omega}_1^{\mathrm{T}}$:

$$\boldsymbol{\rho}_1\boldsymbol{\omega}_1^{\mathrm{T}} = \frac{1}{\hat{K}}\begin{bmatrix} \mathbf{1}_{\mathrm{N}} \\ \mathbf{0}_{\mathrm{N+N^2}} \end{bmatrix}\begin{bmatrix} \mathbf{1}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{K} & -\mathbf{1}_{\mathrm{N+N^2}}^{\mathrm{T}} \end{bmatrix} \tag{65}$$

$$= \frac{1}{\hat{K}}\begin{bmatrix} \mathbf{1}_{\mathrm{N\times N}}\boldsymbol{K} & -\mathbf{1}_{\mathrm{N\times(N+N^2)}} \\ \mathbf{0}_{\mathrm{(N+N^2)\times N}} & \mathbf{0}_{\mathrm{(N+N^2)\times(N+N^2)}} \end{bmatrix}. \tag{66}$$

From equations (60) and (64), we can get:

$$\begin{aligned} E\left[\hat{\boldsymbol{e}}(k+1)\middle|\,\hat{\boldsymbol{e}}(k)\right] &= \boldsymbol{r}_1\boldsymbol{l}_1^{\mathrm{T}}\hat{\boldsymbol{e}}(0) \\ &= \left(\boldsymbol{\rho}_1\boldsymbol{\omega}_1^{\mathrm{T}} \otimes \boldsymbol{\rho}_1\boldsymbol{\omega}_1^{\mathrm{T}}\right)(\boldsymbol{e}(0) \otimes \boldsymbol{e}(0)) \\ &= \boldsymbol{\rho}_1\boldsymbol{\omega}_1^{\mathrm{T}}\boldsymbol{e}(0) \otimes \boldsymbol{\rho}_1\boldsymbol{\omega}_1^{\mathrm{T}}\boldsymbol{e}(0). \end{aligned} \tag{67}$$

Since $\boldsymbol{e}(0) = [\boldsymbol{\lambda}(0); \boldsymbol{\zeta}(0); \boldsymbol{\xi}(0)] - [\lambda^*\mathbf{1}_{\mathrm{N}}; \mathbf{0}_{\mathrm{N}}; \mathbf{0}_{\mathrm{N^2}}]$, we have:

$$\begin{aligned} \boldsymbol{\rho}_1\boldsymbol{\omega}_1^{\mathrm{T}}\boldsymbol{e}(0) &= \frac{1}{\hat{K}}\begin{bmatrix} \mathbf{1}_{\mathrm{N\times N}}\boldsymbol{K} & -\mathbf{1}_{\mathrm{N\times(N+N^2)}} \\ \mathbf{0}_{\mathrm{(N+N^2)\times N}} & \mathbf{0}_{\mathrm{(N+N^2)\times(N+N^2)}} \end{bmatrix}\begin{bmatrix} \boldsymbol{\lambda}(0) \\ \boldsymbol{\zeta}(0) \\ \boldsymbol{\xi}(0) \end{bmatrix} \\ &\quad - \frac{1}{\hat{K}}\begin{bmatrix} \mathbf{1}_{\mathrm{N\times N}}\boldsymbol{K} & -\mathbf{1}_{\mathrm{N\times(N+N^2)}} \\ \mathbf{0}_{\mathrm{(N+N^2)\times N}} & \mathbf{0}_{\mathrm{(N+N^2)\times(N+N^2)}} \end{bmatrix}\begin{bmatrix} \lambda^*\mathbf{1}_{\mathrm{N}} \\ \mathbf{0}_{\mathrm{N}} \\ \mathbf{0}_{\mathrm{N^2}} \end{bmatrix} \\ &= \frac{1}{\hat{K}}\begin{bmatrix} \sum_{i\in\mathcal{I}}(P_i(0) - \zeta_i(0) - \xi_i(0))\mathbf{1}_{\mathrm{N}} \\ \mathbf{0}_{\mathrm{N+N^2}} \end{bmatrix} \\ &\quad - \frac{1}{\hat{K}}\begin{bmatrix} \sum_{i\in\mathcal{I}}(P_i^* - \zeta_i^* - \xi_i^*)\mathbf{1}_{\mathrm{N}} \\ \mathbf{0}_{\mathrm{N+N^2}} \end{bmatrix} \\ &= \mathbf{0}_{\mathrm{2N+N^2}}. \end{aligned} \tag{68}$$

Thus, from equations (67) and (68), we conclude that:

$$E\left[\hat{\boldsymbol{e}}(k+1)\middle|\,\hat{\boldsymbol{e}}(k)\right] = \mathbf{0}_{\mathrm{2N+N^2}} \otimes \mathbf{0}_{\mathrm{2N+N^2}}, \tag{69}$$

which implies that as $k \to \infty$, the expectation of the consensus error converges to zero, i.e., the linear homogeneous system (29) converges in mean square sense. Since system (29) is a discrete-time Markov jump system, it converges in almost sure sense, given that it converges in mean square sense [37]. ■

Appendix C
Proof of Theorem 3

Since problem (1) is a convex optimization problem, the optimality of the solution is proved by showing that the Karush–Kuhn–Tucker (KKT) conditions are satisfied.

We prove the primal feasibility of the solution by showing that the equality constraint (1b) and the inequality constraints (1c) are satisfied. As discussed in Section II-D, one necessary condition for satisfying the equality constraint (1b) is that the global power mismatch is collectively estimated, i.e., $\sum_{i\in\mathcal{I}}\zeta_i(k) = \zeta(k)$. In our proposed method, this feature is preserved by augmenting the communication topology with virtual buffering nodes, i.e.,

$$\sum_{i\in\mathcal{I}}\zeta_i(k) + \sum_{i,j\in\mathcal{I}}\xi_{ij}(k) = \zeta(k). \tag{70}$$

We prove equation (70) by summing up both sides of equations (27) and (28) for all agents $i \in \mathcal{I}$. Since $\tilde{\boldsymbol{W}}(0,k)$ is a column-stochastic matrix, the following equation holds:

$$\begin{aligned} &\sum_{i\in\mathcal{I}}\zeta_i(k+1) + \sum_{i,j\in\mathcal{I}}\xi_{ij}(k+1) \\ &\quad = \sum_{i\in\mathcal{I}}\zeta_i(k) + \sum_{i,j\in\mathcal{I}}\xi_{ij}(k) + \zeta(k+1) - \zeta(k). \end{aligned} \tag{71}$$

Since $\sum_{i\in\mathcal{I}}\zeta_i(0) = \zeta(0)$ and $\xi_{ij}(0) = 0, \forall i,j \in \mathcal{I}$ hold, $\sum_{i\in\mathcal{I}}\zeta_i(0) + \sum_{i,j\in\mathcal{I}}\xi_{ij}(0) - \zeta(0) = 0$, i.e., equation (70) holds when $k = 0$. Substituting $k = 0$ into equation (71), one can infer that equation (70) holds when $k = 1$. Thereby, through induction, we can conclude that equation (70) holds for every k.

As proved in Section III-B, $\zeta_i(k)$ and $\xi_{ij}(k)$ both converge to 0. Thus, equation (70) implies that the global power mismatch $\zeta(k)$ converges to zero as $k \to \infty$, i.e., the equality constraint (1b) is satisfied. Since the inequality constraints (1c) are satisfied when solving the sub-problems (6), the obtained solution is primal feasible.

We prove the dual feasibility by contradiction. If $\lambda^* < 0$, equations in (6) imply that $P_i = P_i^{\max}, \forall i \in \mathcal{I}_{\mathrm{L}}$ and $P_i = P_i^{\min}, \forall i \in \mathcal{I}_{\mathrm{R}} \cup \mathcal{I}_{\mathrm{G}}$. Due to the assumption made in (2), there will exist a mismatch between power supply and demand, which is contradictory to the statement of primal feasibility.

The complementary slackness condition and the stationary condition can be written as:

$$\begin{cases} \phi_i^*\left(P_i - P_i^{\min}\right) = 0, & \forall i \in \mathcal{I} \\ \iota_i^*\left(P_i^{\max} - P_i\right) = 0, & \forall i \in \mathcal{I} \\ -\lambda^* + \phi_i^* - \iota_i^* = 0, & \forall i \in \mathcal{I}_{\mathrm{R}} \\ C_i^{'}(P_i^*) - \lambda^* + \phi_i^* - \iota_i^* = 0, & \forall i \in \mathcal{I}_{\mathrm{G}} \\ -U_i^{'}(P_i^*) + \lambda^* + \phi_i^* - \iota_i^* = 0, & \forall i \in \mathcal{I}_{\mathrm{L}} \end{cases}, \tag{72}$$

where ϕ_i^* and ι_i^* are the multipliers corresponding to the lower and upper power limits of agent i, respectively. By properly selecting ϕ_i^* and ι_i^*, conditions in (72) can be satisfied. ■

References

[1] D. M. Kammen and D. A. Sunter, "City-integrated renewable energy for urban sustainability," *Science*, vol. 352, no. 6288, pp. 922–928, 2016.

[2] H. Sun, Q. Guo, J. Qi, V. Ajjarapu, R. Bravo, J. Chow, Z. Li, R. Moghe, E. Nasr-Azadani, U. Tamrakar, G. N. Taranto, R. Tonkoski, G. Valverde, Q. Wu, and G. Yang, "Review of challenges and research opportunities for voltage control in smart grids," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2790–2801, 7 2019.

[3] Y. Li, Y. Ding, Y. Liu, T. Yang, P. Wang, J. Wang, and W. Yao, "Dense Skip Attention based Deep Learning for Day-Ahead Electricity Price Forecasting," *IEEE Transactions on Power Systems*, pp. 1–19, 2022.

[4] H. Hui, P. Siano, Y. Ding, P. Yu, Y. Song, H. Zhang, and N. Dai, "A transactive energy framework for inverter-based HVAC loads in a real-time local electricity market considering distributed energy resources," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 12, pp. 8409–8421, 12 2022.

[5] L. Olatomiwa, S. Mekhilef, M. S. Ismail, and M. Moghavvemi, "Energy management strategies in hybrid renewable energy systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 62, pp. 821–835, 2016.

[6] H. Kanchev, F. Colas, V. Lazarov, and B. Francois, "Emission reduction and economical optimization of an urban microgrid operation including dispatched PV-based active generators," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 4, pp. 1397–1405, 2014.

[7] A. Parisio, C. Wiezorek, T. Kyntäjä, J. Elo, K. Strunz, and K. H. Johansson, "Cooperative MPC-based energy management for networked microgrids," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 3066–3074, 2017.

[8] J. Arkhangelski, M. Abdou-Tankari, and G. Lefebvre, "Day-ahead optimal power flow for efficient energy management of urban microgrid," *IEEE Transactions on Industry Applications*, vol. 57, no. 2, pp. 1285–1293, 2021.

[9] Y. Chen, C. Li, D. Qi, Z. Li, Z. Wang, and J. Zhang, "Distributed event-triggered secondary control for islanded microgrids with proper trigger condition checking period," *IEEE Transactions on Smart Grid*, vol. 13, no. 2, pp. 837–848, 3 2022.

[10] H. Hui, Y. Chen, S. Yang, H. Zhang, and T. Jiang, "Coordination control of distributed generators and load resources for frequency restoration in isolated urban microgrids," *Applied Energy*, vol. 327, no. August, p. 120116, 12 2022.

[11] Y. Chen, D. Qi, H. Dong, C. Li, Z. Li, and J. Zhang, "A FDI attack-resilient distributed secondary control strategy for islanded microgrids," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 1929–1938, 5 2021.

[12] N. Rahbari-Asr, U. Ojha, Z. Zhang, and M. Y. Chow, "Incremental welfare consensus algorithm for cooperative distributed generation/demand response in smart grid," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2836–2845, 2014.

[13] S. Yang, S. Tan, and J.-X. Xu, "Consensus based approach for economic dispatch problem in a smart grid," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4416–4426, 2013.

[14] L. Yan, X. Chen, and Y. Chen, "A consensus-based privacy-preserving energy management strategy for microgrids with event-triggered scheme," *International Journal of Electrical Power and Energy Systems*, vol. 141, p. 108198, 2022.

[15] C. Zhao, J. He, P. Cheng, and J. Chen, "Consensus-based energy management in smart grid with transmission losses and directed communication," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2049–2061, 2017.

[16] D. Yang, S. Zhang, B. Zhou, and S. Bu, "Consensus-based decentralized optimization for distributed generators power allocation over time-varying digraphs in microgrids," *IEEE Systems Journal*, vol. 15, no. 1, pp. 814–825, 2021.

[17] C. Xi, V. S. Mai, R. Xin, E. H. Abed, and U. A. Khan, "Linear Convergence in Optimization over Directed Graphs with Row-Stochastic Matrices," *IEEE Transactions on Automatic Control*, vol. 63, no. 10, pp. 3558–3565, 2018.

[18] Q. Yang, G. Chen, and T. Wang, "ADMM-based distributed algorithm for economic dispatch in power systems with both packet drops and communication delays," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 3, pp. 842–852, 2020.

[19] M. Mao, C. He, L. Chang, and Y. Liu, "A distributed economic dispatch algorithm for islanding microgrid considering unreliable communication links," *2019 IEEE Energy Conversion Congress and Exposition, ECCE 2019*, no. 51577047, pp. 6278–6285, 2019.

[20] Y. Zhang, N. Rahbari-Asr, and M. Y. Chow, "A robust distributed system incremental cost estimation algorithm for smart grid economic dispatch with communications information losses," *Journal of Network and Computer Applications*, vol. 59, pp. 315–324, 2016.

[21] R. Wang, Q. Li, G. Li, and H. Liu, "A gossip-based distributed algorithm for economic dispatch in smart grids with random communication link failures," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 6, pp. 4635–4645, 2020.

[22] J. Duan and M. Y. Chow, "Robust consensus-based distributed energy management for microgrids with packet losses tolerance," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 281–290, 2020.

[23] C. N. Hadjicostis, N. H. Vaidya, and A. D. Domínguez-García, "Robust Distributed Average Consensus via Exchange of Running Sums," *IEEE Transactions on Automatic Control*, vol. 61, no. 6, pp. 1492–1507, 2016.

[24] J. Wu, T. Yang, D. Wu, K. Kalsi, and K. H. Johansson, "Distributed optimal dispatch of distributed energy resources over lossy communication networks," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 3125–3137, 2017.

[25] Y. Wu, X. Tan, L. Qian, D. H. Tsang, W. Z. Song, and L. Yu, "Optimal pricing and energy scheduling for hybrid energy trading market in future smart grid," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 6, pp. 1585–1596, 2015.

[26] C. Zhao, J. Chen, J. He, and P. Cheng, "Privacy-preserving consensus-based energy management in smart grids," *IEEE Transactions on Signal Processing*, vol. 66, no. 23, pp. 6162–6176, 2018.

[27] G. C. Calafiore and L. El Ghaoui, *Optimization Models*. Cambridge University Press, 2014.

[28] L. Gurobi Optimization, "Gurobi Optimizer Reference Manual," 2022.

[29] K. Cai and H. Ishii, "Average consensus on general strongly connected digraphs," *Automatica*, vol. 48, no. 11, pp. 2750–2761, 2012.

[30] B.-Y. Choi, S. Moon, Z.-L. Zhang, K. Papagiannaki, and C. Diot, "Analysis of point-to-point packet delay in an operational network," *Computer Networks*, vol. 51, no. 13, pp. 3812–3827, 9 2007.

[31] N. Van der Aa, H. Ter Morsche, and R. Mattheij, "Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem," *The Electronic Journal of Linear Algebra*, vol. 16, no. October 2006, pp. 300–314, 2007.

[32] P. Lin, W. Ren, and J. A. Farrell, "Distributed Continuous-Time Optimization: Nonuniform Gradient Gains, Finite-Time Convergence, and Convex Constraint Set," *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2239–2253, 2017.

[33] P. Lin, W. Ren, C. Yang, and W. Gui, "Distributed Continuous-Time and Discrete-Time Optimization with Nonuniform Unbounded Convex Constraint Sets and Nonuniform Stepsizes," *IEEE Transactions on Automatic Control*, vol. 64, no. 12, pp. 5148–5155, 2019.

[34] J. Zhang, K. You, and K. Cai, "Distributed Dual Gradient Tracking for Resource Allocation in Unbalanced Networks," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2186–2198, 2020.

[35] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. 25, no. 9, pp. 772–781, 1978.

[36] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[37] O. L. V. do Costa, R. P. Marques, and M. D. Fragoso, *Discrete-Time Markov Jump Linear Systems*, ser. Probability and Its Applications. London: Springer London, 2005.

Hongyi Li (S'21) received the B.Eng. degree from Huazhong University of Science and Technology, Wuhan, China, in 2020 and the M.Sc. degree from Imperial College London, London, UK, in 2021, both in electrical engineering. He is currently working toward the Ph.D. degree at University of Macau, Macau, China. His research interests include Internet of Things for smart energy, consensus-based energy management and blockchain-assisted optimization technology.

Hongxun Hui (S'17–M'20) received the B.E. and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2015 and 2020, respectively. He is currently a Research Assistant Professor with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao SAR, China. From 2018 to 2019, he was a visiting scholar at the Advanced Research Institute in Virginia Tech, and the CURENT Center in University of Tennessee. His research interests include power system optimization, demand-side resource control, and Internet of Things technologies for smart energy.

Hongcai Zhang (S'14–M'18) received the B.S. and Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2013 and 2018, respectively. He is currently an Assistant Professor with the State Key Laboratory of Internet of Things for Smart City and Department of Electrical and Computer Engineering, University of Macau, Macao, China. In 2018-2019, he was a postdoctoral scholar with the Energy, Controls, and Applications Lab at University of California, Berkeley, where he also worked as a visiting student researcher in 2016. His current research interests include Internet of Things for smart energy, optimal operation and optimization of power and transportation systems, and grid integration of distributed energy resources.