This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2720761, IEEE Transactions on Smart Grid

Distributed Optimal Dispatch of Distributed Energy Resources over Lossy Communication Networks

Junfeng Wu, Tao Yang, *Member, IEEE*, Di Wu, *Member, IEEE*, Karanjit Kalsi, *Member, IEEE*, and Karl H. Johansson, *Fellow, IEEE*

***Abstract*—Driven by smart grid technologies, a great effort has been made in developing distributed energy resources (DERs) in recent years for improving reliability and efficiency of distribution systems. Emerging DERs require effective and efficient control and coordination in order to harvest their potential benefits. In this paper, we consider optimal DER coordination problem, where the goal is to minimize the total generation cost while meeting total demand and satisfying individual generator output limit. This paper develops a distributed algorithm for solving the optimal DER coordination problem over lossy communication networks with packet-dropping communication links. Under the assumption that the underlying communication network is strongly connected with a positive probability and the packet drops are independent and identically distributed (i.i.d.), we show that the proposed algorithm is able to solve the optimal DER coordination problem even in the presence of packet drops. Numerical simulation results are used to validate and illustrate the proposed algorithm.**

***Index Terms*—Distributed algorithms; Optimal DER coordination; Packet drops; Power systems; Smart grids.**

I. Introduction

IN the past decades, power systems have been undergoing a transition from a system with conventional generation power plants and inflexible loads to a system with a large numbers of distributed generators, energy storages, and flexible loads, often referred to as distributed energy resources (DERs) [1]. These resources are small and highly flexible compared with conventional generators, and can be aggregated to provide power necessary to meet the regular demand. As the electricity grid continues to modernize, DER can help facilitate the transition to a smarter grid.

In order to achieve an effective deployment among DERs, one needs to properly design the coordination and control among them. One approach is through a completely centralized control strategy, where a single control center accesses the entire network's information and provides control signals to the entire system. However, centralized approaches have a few drawbacks, such as a single point failure, high communication requirement and computation burden, and limited flexibility [2], [3].

To overcome these limitations, recently, by using the results in the fields of distributed control and multi-agent systems [4], [5], various distributed strategies have been proposed for solving the DER coordination problem [6]–[13]. In these distributed algorithms, each agent (generator) maintains a local estimate of an optimal incremental cost, which is the consensus variable, and updates it by exchanging information with only a few neighboring agents. Based on the consensus theory [4], [5], if the communication network is connected, all these local estimates converge to an optimal increment cost. The distributed algorithms for the DER coordination problem are progressing with generalization of communication networks, from fixed undirected networks to fixed directed networks and time-varying networks. For undirected fixed communication networks, the authors of [6] proposed a leader-follower consensus-based algorithm where the leader collects the mismatch between demand and generation. The authors of [7] develop a leaderless algorithm, where in addition to the consensus part, an innovation term is introduced to ensure the balance between system generation and demand. For directed fixed communication networks, the authors of [8] proposed a distributed algorithm based on the ratio consensus algorithm, and a consensus-based algorithm where agents collectively learn the system imbalance was developed in [9]. To further alleviate the communication burden, a distributed algorithm based on the consensus and bisection method was proposed in [11] and a minimum-time consensus-based algorithm was developed in [12]. In all aforementioned references, the communication network is assumed to be perfect with reliable communication links. However, varying communication links and communication time delays are ubiquitous in communication networks. Therefore, recent studies have been devoted to developing distributed algorithms for the DER coordination problem over communication networks which may be subject to varying communication links and/or communication time delays. The authors of [14] proposed a distributed algorithm based on nonnegative-surplus [15] to solve the DER coordination problem over time-varying directed communication networks but without time delays. To handle the case where networks are subject to both time-varying topologies and communication time delays, the authors of [16] developed a distributed algorithm based on the push-sum and gradient method [17].

In this paper, we consider the case where the communi-

This material is based upon work supported by the Laboratory Directed Research and Development (LDRD) Program through the Control of Complex Systems Initiative at the Pacific Northwest National Laboratory. The work of J. Wu and K.H. Johansson has been supported in part by the Knut and Alice Wallenberg Foundation, the Swedish Research Council and the National Natural Science Foundation of China under Grant No. 61120106011. (Corresponding author: T. Yang; Tel: +1 940-891-6876.)

J. Wu and K. H. Johansson are with the ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, Stockholm 10044, Sweden. Email: {junfengw, kallej}@kth.se.

T. Yang is with the Department of Electrical Engineering, University of North Texas, Denton, TX 75203. Email: Tao.Yang@unt.edu.

D. Wu and K. Kalsi are with the Pacific Northwest National Laboratory, Richland, WA 99352. Email: {di.wu, karanjit.kalsi}@pnnl.gov.

cation networks may be subject to unreliable communication links, which is common in communication networks. Here the reliability of communication links is treated as packet drops. Although time-varying communication networks may be used to model packet drops, a more realistic modeling approach is based on the probability framework, i.e., the communication link fails with a certain probability. In such a probability setting, the previously developed DER coordination algorithms in [14], [16] for time-varying communication networks are not able to handle packet drops. The main contribution of this paper is to propose a robustified extension of the distributed algorithm proposed in our earlier work [16] and show that this robustified distributed algorithm is able to solve the DER coordination problem over communication networks even in the presence of packet drops.

While the motivation for this work is driven by power system applications, the proposed framework is also useful in addressing similar problems that arise in other networked cyber-physical systems where the cyber communication network is subject to packet-dropping links. In this regard, our work is closely related to the literature of distributed optimization [17]–[21]. In [17], a distributed algorithm based on the push-sum and (sub)gradient method was developed to solve the optimization problem over directed time-varying communication networks. In [18], the authors proposed a distributed algorithm to solve the optimization problem over communication networks with packet drops, where packet drops are modeled by time-varying graphs. In [19], [20], the authors developed an extension of the ratio consensus algorithm in which messages are encoded as running sums and show that the extended algorithm is able to solve the average consensus problem in the presence of packet drops, i.e., average consensus is achieved almost surely. In contrast to [19], [20], we aim to go beyond finding a feasible solution (i.e., we include some optimization criteria in the problem formulation) and try to solve the distributed optimization problem even in the presence of unreliable communication links with packet drops. To do so, we propose a distributed algorithm by integrating our previously proposed algorithm in [16] based on the push-sum and gradient method [17] without packet-dropping links with the robustified strategy proposed in [19], [20]. Notice that our work can also be viewed as a robustified extension of the distributed algorithm developed in [17]. A similar distributed optimization in the presence of packet drops has also been considered in [21], where a distributed algorithm based on the Newton-Raphson consensus approach [22] and the robustified strategies in [19], [20] were developed. Compared with [22] where the second derivative of the local cost (objective) function was used, our proposed algorithm only uses the gradient (first derivative) of the local cost function and thus enjoys less computation burden.

The remainder of the paper is organized as follows: In Section II, we introduce some preliminaries on graph theory, the problem formulation for DER coordination, and our previously proposed algorithm for solving the DER coordination problem over networks without packet drops. Section III presents an example that motivates our study. In Section IV, a distributed algorithm based on our previous algorithm and a robustified strategy is proposed to solve the DER coordination problem over unreliable communication networks with packet-dropping communication links. Case studies are presented in Section V to illustrate and validate the proposed algorithm. Finally, concluding remarks are offered in Section VI.

II. Preliminary

This section first presents some background on graph theory [23], which is needed to describe the communication network among DERs. In addition, we formulate the DER coordination problem and briefly summarize our previously developed distributed algorithm for communication networks with reliable links [16].

A. Communication Network

In this paper, we assign each bus in the power system an agent (node). Information exchanges among the agents occur over a communication network, described by a directed graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ denotes the index set of the agents with N being the number of agents and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes the set of communication links between some pairs of the agents. In particular, $(j, i) \in \mathcal{E}$ if there exists a directed communication link from agent i to agent j. For notational convenience, we assume that $(j, j) \notin \mathcal{E}$ for all $j \in \mathcal{V}$ although each agent has an access to its own information. A directed path from node i_1 to node i_k is a sequence of nodes $i_1, \ldots, i_k$ such that $(i_{j+1}, i_j) \in \mathcal{E}$ for $j = 1, \ldots, k-1$. If there exists a directed path from node i to node j, then node j is said to be reachable from node i. A directed graph $\mathcal{G}$ is said to be strongly connected if every node is reachable from every other node. Let $\mathcal{N}_j^{\text{in}}$ and $\mathcal{N}_j^{\text{out}}$ denote the in- and out-neighbors of node j, respectively, i.e.,

$$\mathcal{N}_j^{\text{in}} = \{i \in \mathcal{V} \mid (j, i) \in \mathcal{E}\},$$
$$\mathcal{N}_j^{\text{out}} = \{\ell \in \mathcal{V} \mid (\ell, j) \in \mathcal{E}\},$$

and d_j^{out} denotes the out-degree of node j, i.e., $d_j^{\text{out}} = |\mathcal{N}_j^{\text{out}}|$. To support information prorogation from one agent of the network to another, we will make the following assumption on graph connectivity.

Assumption 1. *The graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ is strongly connected. Each agent j knows its own out-degree d_j^{out}.*

B. Distributed Dispatch over Networks with Reliable Communication Links

The goal of the DER coordination problem is to minimize the total generation cost while meeting total demand and satisfying individual generator output limits, formulated as:

$$\min_{x_i} \quad \sum_{i=1}^{N} C_i(x_i) \tag{1a}$$

$$\text{subject to} \quad \sum_{i=1}^{N} x_i = D, \tag{1b}$$

$$x_i \in X_i := [x_i^{\min}, x_i^{\max}], \; i = 1, \ldots, N, \tag{1c}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2720761, IEEE Transactions on Smart Grid

where x_i is the power generation of agent i, $C_i(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is the cost function of agent i, where $\mathbb{R}^+$ is the set of non-negative real numbers, $x_i^{\min}$ and $x_i^{\max}$ are respectively the lower and upper bounds of the power generation of agent i, and D is the total demand satisfying $\sum_{i=1}^N x_i^{\min} \leq D \leq \sum_{i=1}^N x_i^{\max}$ in order to ensure the feasibility of problem (1).

Compared to most studies where cost functions are assumed to be quadratic, this paper considers general convex cost functions that satisfy the following assumption.

Assumption 2. *For each $i \in \{1, \ldots, N\}$, the cost function $C_i(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is strictly convex and continuously differentiable.*

Since i) each cost function $C_i(\cdot)$ is convex, ii) the constraint (1b) is affine, and iii) the set $X_1 \times \cdots \times X_N$ is a polyhedral set, if we dualize problem (1) with respect to the constraint (1b), there is zero duality gap. Moreover, the dual optimal set is nonempty [24]. Consequently, solutions of the DER coordination problem can be obtained by solving its dual problem.

For convenience, let $\mathbf{x} := [x_1, \ldots, x_N]^\top \in \mathbb{R}_+^N$. Then, define the Lagrangian function

$$\mathcal{L}(\mathbf{x}, \lambda) = \sum_{i=1}^N C_i(x_i) - \lambda \left(\sum_{i=1}^N x_i - D \right).$$

The corresponding Lagrange dual problem is

$$\max_{\lambda \in \mathbb{R}_+} \sum_{i=1}^N \Psi_i(\lambda) + \lambda D, \tag{2}$$

where

$$\Psi_i(\lambda) = \min_{x_i \in X_i} C_i(x_i) - \lambda x_i. \tag{3}$$

Under Assumption 2, for any given $\lambda \in \mathbb{R}_+$, the right-hand side of (3) has a unique minimizer given by

$$x_i(\lambda) = \mathrm{proj}_{X_i}\left(\nabla C_i^{-1}(\lambda)\right), \tag{4}$$

where ∇C_i^{-1} denotes the inverse function of ∇C_i, which exists over $[\nabla C_i(x_i^{\min}), \nabla C_i(x_i^{\max})]$ since ∇C_i is continuous and strictly increasing due to Assumption 2, and $\mathrm{proj}_{X_i}\left(\nabla C_i^{-1}(\lambda)\right)$ denotes the projection of $\nabla C_i^{-1}(\lambda)$ to the set X_i, defined as

$$\mathrm{proj}_{X_i}\left(\nabla C_i^{-1}(\lambda)\right) = \min\{\max\{\nabla C_i^{-1}(\lambda), x_i^{\min}\}, x_i^{\max}\}.$$

Furthermore, there is at least one optimal solution to the dual problem (2), and the unique optimal solution of the primal DER coordination problem is given by

$$x_i^* = x_i(\lambda^*), \quad \forall i = 1, 2, \ldots, N, \tag{5}$$

where λ^* is any dual optimal solution.

For any given $\lambda \in \mathbb{R}_+$, because of the uniqueness of $x_i(\lambda)$, the dual function $\sum_{i=1}^N \Psi_i(\lambda) + \lambda D$ is differentiable at λ and its gradient is given by $-(\sum_{i=1}^N x_i(\lambda) - D)$ [25]. We can then apply the gradient method to solve the dual problem (2):

$$\lambda(t+1) = \lambda(t) - \gamma(t) \left(\sum_{i=1}^N x_i(\lambda(t)) - D \right), \tag{6}$$

where $\lambda(0) \in \mathbb{R}$ can be arbitrarily assigned and $\gamma(t)$ is the step-size at time instant (step) t.

When designing a distributed algorithm based on (6), the main challenge is how to obtain the global quantity $\sum_{i=1}^N x_i(\lambda(t)) - D$ in a distributed manner. To do so, we note that the dual problem (2) can be converted into

$$\max_{\lambda \in \mathbb{R}_+} \sum_{i=1}^N \Phi_i(\lambda), \tag{7}$$

where

$$\Phi_i(\lambda) = \min_{x_i \in X_i} C_i(x_i) - \lambda(x_i - D_i), \tag{8}$$

and D_i is a virtual local demand at each bus such that $\sum_{i=1}^N D_i = D$. Note that there is no physical meaning to D_i's. The purpose of introducing these parameters is for designing a distributed algorithm by applying the gradient method based on the dual problem (7). The gradient of $\Phi_i(\lambda)$ is

$$\nabla \Phi_i(\lambda) = -\left(x_i(\lambda) - D_i\right). \tag{9}$$

In our previous work [16], we have proposed a distributed algorithm based on the push-sum and gradient method [17] for solving the DER coordination problem over strongly connected networks with reliable communication links. In the proposed algorithm, each agent j maintains scalar variables $v_j(t)$, $w_j(t)$, $y_j(t)$, $\lambda_j(t)$, $x_j(t)$, where $x_j(t)$ and $\lambda_j(t)$ are the estimates of the optimal generation (primal optimal solution) and the optimal incremental cost (dual optimal solution), respectively. At each time step t, each agent $j \in \mathcal{V}$ updates its variables through information exchanges with its neighbors according to

$$w_j(t+1) = \sum_{i \in \mathcal{N}_j^{\text{in}}(t) \cup \{j\}} \frac{v_i(t)}{d_i^{\text{out}} + 1}, \tag{10a}$$

$$y_j(t+1) = \sum_{i \in \mathcal{N}_j^{\text{in}}(t) \cup \{j\}} \frac{y_i(t)}{d_i^{\text{out}} + 1}, \tag{10b}$$

$$\lambda_j(t+1) = \frac{w_j(t+1)}{y_j(t+1)}, \tag{10c}$$

$$x_j(t+1) = \mathrm{proj}_{X_j}\left(\nabla C_j^{-1}(\lambda)\right), \tag{10d}$$

$$v_j(t+1) = w_j(t+1) - \gamma(t+1)(x_j(t+1) - D_j). \tag{10e}$$

The step size $\gamma(t+1) > 0$ satisfies the following assumption.

Assumption 3. *The sequence $(\gamma(t))_{t \in \mathbb{N}}$ satisfies the following conditions:*

$$\sum_{t=1}^\infty \gamma(t) = \infty, \quad \sum_{t=1}^\infty \gamma^2(t) < \infty, \text{ and}$$
$$0 < \gamma(t) \leq \gamma(s) \quad \text{for all } t > s \geq 0. \tag{11}$$

The algorithm (10) is initialized at time instant $t = 0$, with an arbitrary value to $v_j(0)$ at agent j and $y_j(0) = 1$ for all $j \in \mathcal{V}$.

Remark 1. *We compare the distributed algorithm (10) with existing ones in the literature. According to (9), $-(x_j(t+1) - D_j)$ in (10e) is the gradient of the function $\Phi_j(\lambda)$ at $\lambda = \lambda_j(t+1)$. Without (10d) and the gradient term in (10e), the algorithm is reduced to a particular version of push-sum*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2720761, IEEE Transactions on Smart Grid

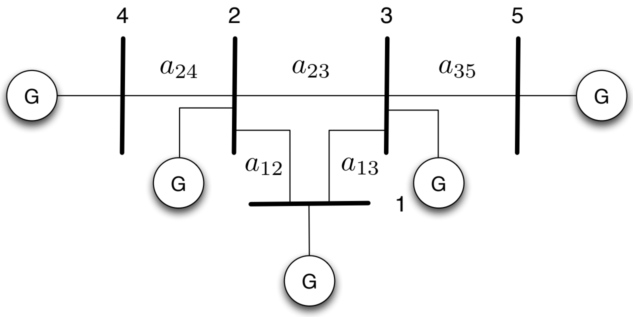


Fig. 1. IEEE five-bus power system.

TABLE I
GENERATOR PARAMETERS

Bus	a_i (kW^{2}h)	b_i ($/kWh)	c_i ($/h)	Range (kW)
1	0.00024	0.0267	0.38	[30,60]
2	0.00052	0.0152	0.65	[20,60]
3	0.00042	0.0185	0.4	[50,200]
4	0.00052	0.0152	0.65	[20,60]
5	0.00031	0.0297	0.3	[20,140]

algorithm [26], or ratio consensus algorithm [27], [28] for computing the average of initial values in directed graphs. In this case, all $\lambda_j(t+1)$ converge to the average of the initial values. The inclusion of the gradient term in the update of $v_j(t+1)$ is to ensure that all $\lambda_j(t+1)$ converge to an optimal incremental cost λ^*.

We are now ready to recall our previous result which states that the proposed distributed algorithm (10) solves the optimal dispatch problem for distributed energy resources over reliable communication networks.

Lemma 1 ([16] Theorem 1)**.** *Under Assumptions 1, 2 and 3, the distributed algorithm* (10) *solves the optimal DER coordination problem, i.e., $\lambda_i(t) \to \lambda^*$, and $x_i(t) \to x_i^*$ as $t \to \infty$ for all $i \in \mathcal{V}$.*

The proof of Lemma 1 was motivated by [17] and was carried out in two steps. The first step shows that $\lambda_i(t+1)$ tracks the average $\bar{v}(t) = \frac{1}{N}\sum_{i=1}^{N} v_i(t)$ for $t \geq 0$ increasingly well as time goes on. The second step shows that the average $\bar{v}(t)$ converges to an optimal incremental cost λ^*.

III. Motivating Example and Problem Statement

In this section, we first present a motivating example for this study. We consider the IEEE 5-bus system shown in Fig. 1, where each bus is connected with a generator whose cost function is quadratic, i.e., $C_i(x_i) = a_i x_i^2 + b_i x_i + c_i$. The parameters of the generators including the parameters of the quadratic cost functions are given in Table I. The communication network is not necessarily the same as the physical topology and is modeled by a fixed directed graph depicted in Fig. 2.

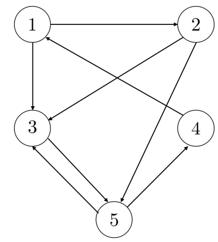
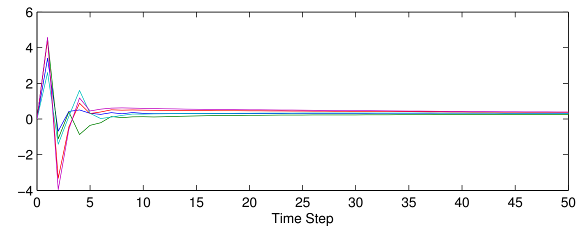
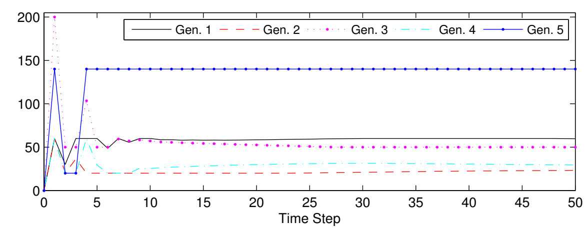


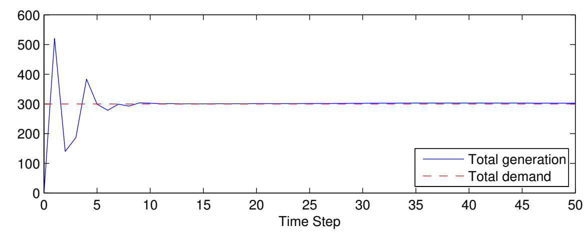
Fig. 2. Directed communication network.



(a) Incremental cost ($/kWh)



(b) Generation (kW)



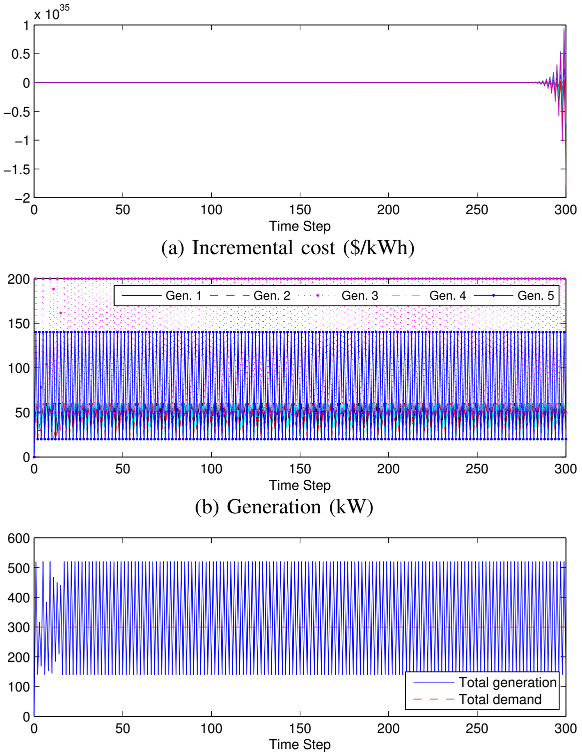
(c) Generation (kW)

Fig. 3. Results for networks with reliable communication links.

A. Perfect Communication Networks

We first consider the case where the communication network is perfect with reliable communication links. The virtual local demands at each bus are given as $D_1 = 40$ kW, $D_2 = 30$ kW, $D_3 = 100$ kW, $D_4 = 40$ kW, and $D_5 = 90$ kW. The total demand is $D = \sum_{i=1}^{5} D_i = 300$ kW, which is unknown to the agent at each bus. The simulation results of running distributed algorithm (10) with step size $\gamma(t) = \frac{0.15}{t}$ are given in Fig. 3. It is shown in Fig. 3a that all the estimates of the optimal incremental cost converge to an optimal value $\lambda^* = 0.296$ $/kWh. As shown in Fig. 3b, the power outputs of the generators also converge to their optimal values, which are $x_1^* = 56.05$ kW, $x_2^* = 26.975$ kW, $x_3^* = 50$ kW, $x_4^* = 26.975$ kW, and $x_5^* = 140$ kW, which agrees with the centralized solution. As $\lambda_i(t)$ and $x_i(t)$ converge to their optima, the total generation meets the total demand $D = 300$ kW as shown in Fig. 3c. These results are in consistence with the our previous result [16, Theorem 1], recapped in Lemma 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2720761, IEEE Transactions on Smart Grid



(a) Incremental cost ($/kWh)

(b) Generation (kW)

(c) Generation (kW)

Fig. 4. Results for networks with packet-dropping communication links

B. Unreliable Communication Networks with Packet Drops

We next consider the effect to the proposed distributed algorithm (10) when the communication networks are unreliable with packet-dropping communication links. In particular, we consider the case where each communication link $(j, i) \in \mathcal{E}$ suffers a packet drop with the same probability $q_{ji} = 0.1$. Since the packet drops are random, the iteration results at each agent vary from one simulation to another. Nevertheless, the proposed algorithm (10) always fails to converge. The simulation results of a particular run are given in Fig. 4, which shows that the algorithm fails to converge, and thus fails to solve the DER coordination problem in the presence of packet-dropping communication links.

To conclude, we find that the previously developed algorithm (10) which solves the DER coordination problem over networks with reliable links, however, fails for the case when communication links are subject to packet drops. This motivates us to propose a distributed algorithm for the DER coordination problem over unreliable networks with packet-dropping communication links.

To do so, let us first introduce a probabilistic modeling approach for packet drops. For a fixed strongly connected communication network $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, due to packet-dropping communication, the existing communication link from agent i to agent j, $(j, i) \in \mathcal{E}$ may randomly fail with some nonzero probability. Let $(\Omega, \mathcal{F})$ denote the measurable space generated by the intermittent communication over $\mathcal{E}$. We use an indicator variable $r_{ji}(t; \omega) : \Omega \to \{0, 1\}$ to denote if the communication over $(j, i) \in \mathcal{E}$ is successful or not: let $r_{ji}(t; \omega) = 1$ if the information from agent i is received by agent j at time t; otherwise let $r_{ji}(t; \omega) = 0$. Notice that for each link $(j, i) \in \mathcal{E}$, $r_{ji}(t; \omega)$ can be defined accordingly at time t. We let $r(t; \omega)$ be a random vector containing all random variables of $\{r_{ji}(t; \omega) : (j, i) \in \mathcal{E}\}$ in a fixed order and denote $p_{ji}(t) := \mathbf{E}[r_{ji}(t)]$. We make the following assumption on the sequence $(r(t))_{t \in \mathbb{N}}$, where $\mathbb{N} = \{0, 1, 2 \ldots\}$.

Assumption 4. *The binary random vectors $r(0), r(1), \ldots$ has the following property:*

(i). *For any time t, any two elements $r_{ji}(t)$ and $r_{lk}(t)$ of $r(t)$ are independent.*

(ii). *The link failure rate is strictly less than one, i.e., $1 - p_{ji}(t) < 1$.*

(iii). *The random vectors $r(0), r(1), \ldots$ are independently and identically distributed (i.i.d.).*

According to (ii) of Assumption 4, we can simplify the notation of $p_{ji}(t)$ by discarding the time index "t" into p_{ji}. In particular, at any time instant t, for $(j, i) \in \mathcal{E}$, let $r_{ji}(t)$ be indicator variables which take value $r_{ji}(t) = 1$ if the message from agent i is received by agent j, otherwise $r_{ji}(t) = 0$. The goal of this paper is to propose a distributed algorithm for the DER coordination problem that is able to overcome packet drops.

IV. Main Results

To cope with the effects of unreliable communication networks to the distributed algorithms for DER coordination, in this section, we present a robustified extension of the distributed algorithm (10). We then show that this robustified distributed algorithm is capable to solve the DER coordination problem even in the presence of packet-dropping communication links as long as the underlying communication network is strongly connected with a positive probability.

A. Resilient Distributed DER Coordination Algorithm against Packet Drops

To propose a resilient DER coordination algorithm against packet drops, we integrate the algorithm (10) with the running-sum method proposed in [19], [20] for handling packet drops. The proposed algorithm is given in Algorithm 1. Intuitively, compared to the distributed algorithm (10), in Algorithm 1, each agent j keeps track of certain additional variables, includes them in the message it broadcasts, and uses them in the update equations. In particular, besides variables $w_j(t+1)$, $y_j(t+1)$, $\lambda_j(t+1)$, $x_j(t+1)$ and $v_j(t+1)$, each agent j at time instant $t+1$ also maintains additional variables $\sigma_j(t+1) = \sum_{k=0}^{t} \frac{v_j(k)}{1+d_j^{\text{out}}}$ and $\eta_j(t+1) = \sum_{k=0}^{t} \frac{y_j(k)}{1+d_j^{\text{out}}}$, which are the running sums of v_j and y_j respectively, and $\rho_{ji}(t+1)$ and $\upsilon_{ji}(t+1)$ for $i \in \mathcal{N}_j^{\text{in}}$ which keep track of the running sum of v_j and y_j received at agent j from agent i. These variables are updated according to Algorithm 1. Notice that each agent j computes the running sums $\sigma_j(t+1)$ and $\eta_j(t+1)$ according to (12) and sends them to all its outgoing neighboring agents. The running sums are initialized to $\sigma_j(0) = 0$ and $\eta_j(0) = 0$ for all $j \in \mathcal{V}$. The variables ρ_{ji} and υ_{ji} remain unchanged until a transmission is successfully received on link $(j, i) \in \mathcal{E}$, i.e., $r_{ji}(t) = 1$. It is clear that each agent knows the running sum of itself, i.e., $\rho_{jj}(k+1) = \sigma_j(t+1)$ and $\upsilon_{jj}(t+1) = \eta_j(t+1)$. Finally, agent j updates the values of w_j and y_j to be the sum of the differences between the two most recently received running sum values according to (14a) and (14b) while other update equations (14c) to (14e) are the same as (10c)-(10e).

Remark 2. *We compare Algorithm 1 with the existing distributed algorithms in the literature. Without the running sum variables $\sigma_j(t)$ and $\eta_j(t)$, the algorithm reduces to the subgradient-push algorithm in [17]. The distributed DER coordination algorithm in [16] can be treated as a particular version of Algorithm 1 without the running sum variables. In this case, all $\lambda_j(t+1)$ converge to an optimal incremental cost λ^* in the absence of lossy communication links. Without (14d) and the gradient term in (14e), the algorithm is reduced to the algorithm proposed in [20] which converges to the average of the initial values almost surely in the presence of packet-dropping communication links. The inclusion of the gradient term in the update of $\upsilon_j(t+1)$ is to ensure that all $\lambda_j(t+1)$ converge to an optimal incremental cost λ^* almost surely.*

Remark 3. *In the absence of lossy communication links (i.e., $r_{ji}(t) \equiv 1$), Algorithm 1 reduces to the algorithm (10). This can be seen from the following observations: 1) When $r_{ji}(t) \equiv 1$, there hold $\rho_{ji}(t) \equiv \sigma_i(t)$ and $\upsilon_{ji}(t) \equiv \eta_i(t)$. 2) According to (12), (14a), and (14b),*

$$w_j(t+1) = \sum_{i \in \mathcal{N}_j^{\text{in}} \cup \{j\}} \frac{\upsilon_i(t)}{d_i^{\text{out}} + 1}$$

and

$$y_j(t+1) = \sum_{i \in \mathcal{N}_j^{\text{in}} \cup \{j\}} \frac{y_i(t)}{d_i^{\text{out}} + 1}.$$

Hence, Algorithm 1 is a robustified extension of the distributed algorithm (10) for the case with packet-dropping communication links.

Algorithm 1 Distributed algorithm for the DER coordination problem over networks with packet-dropping communication links

1: **Input:** $\upsilon_j(0)$, $\sigma_j(0) = 0$, $\rho_{ji}(0) = 0$, $\forall i \in \mathcal{N}_j^{\text{in}}$, $y_j(0) = 1$, $\eta_j(0) = 0$, $\upsilon_{ji}(0) = 0$, $\forall i \in \mathcal{N}_j^{\text{in}}$.

2: **for** $t \geq 0$:

3: **Compute:**

$$\sigma_j(t+1) = \sigma_j(t) + \upsilon_j(t)/(1 + d_j^{\text{out}}), \quad (12a)$$
$$\eta_j(t+1) = \eta_j(t) + y_j(t)/(1 + d_j^{\text{out}}). \quad (12b)$$

4: **Broadcast:** $\sigma_j(t+1)$ and $\eta_j(t+1)$ to all $\ell \in \mathcal{N}_j^{\text{out}}$.

5: **Receive:** From each $i \in \mathcal{N}_j^{\text{in}}$ receive $\sigma_i(t+1)$ and $\eta_i(t+1)$ if $r_{ji}(t) = 1$.

6: **Set:**

$$\rho_{ji}(t+1) = \begin{cases} \sigma_i(t+1), & \text{if } r_{ji}(t) = 1 \text{ or } i = j, \\ \rho_{ji}(t), & \text{if } r_{ji}(t) = 0. \end{cases} \quad (13a)$$

$$\upsilon_{ji}(t+1) = \begin{cases} \eta_i(t+1), & \text{if } r_{ji}(t) = 1 \text{ or } i = j, \\ \upsilon_{ji}(t), & \text{if } r_{ji}(t) = 0. \end{cases} \quad (13b)$$

7: **Compute:**

$$w_j(t+1) = \sum_{i \in \mathcal{N}_j^{\text{in}} \cup \{j\}} (\rho_{ji}(t+1) - \rho_{ji}(t)), \quad (14a)$$
$$y_j(t+1) = \sum_{i \in \mathcal{N}_j^{\text{in}} \cup \{j\}} (\upsilon_{ji}(t+1) - \upsilon_{ji}(t)), \quad (14b)$$
$$\lambda_j(t+1) = \frac{w_j(t+1)}{y_j(t+1)}, \quad (14c)$$
$$x_j(t+1) = \text{proj}_{X_j}\left(\nabla C_j^{-1}(\lambda_j(t+1))\right), \quad (14d)$$
$$\upsilon_j(t+1) = w_j(t+1) - \gamma(t+1)(x_j(t+1) - D_j). \quad (14e)$$

B. Convergence Results

In this section, we present the convergence results for the proposed Algorithm 1. We first show that, for each agent j, a subsequence of $(\lambda_j(t))_{t \in \mathbb{N}}$ almost surely (a.s.) converges to the same optimal incremental cost λ^*, which is an optimal solution to the dual problem (2). By doing so, we then show that the proposed distributed Algorithm 1 is able to solve the DER coordination problem over networks with packet-dropping communication links. Here we focus on the almost sure convergence analysis (i.e., pointwise convergence on the sample space Ω). Our main result is obtained with the help of results from the weak ergodicity theory and the supermartingale convergence theorem.

In order to present the main results, the following property from [20] is needed. The presentation of the property will be adopted to the context of this paper.

Proposition 1 ([20] Lemma 2)**.** *Assume that Assumptions 1 and 4 are satisfied. For Algorithm 1, we have $\mathbf{P}(y_j(t) \geq C \text{ i.o.}) = 1$, where $C := \frac{1}{N^N}$ and "i.o." is short for "infinitely often".*

By virtue of Proposition 1, we can define a sequence of time instants for each agent j, at which $y_j(t) \geq C$ is satisfied, as follows:

$$t_{j,1} = \min\{t : y_j(t) \geq C\},$$
$$t_{j,2} = \min\{t : y_j(t) \geq C, t > t_{j,1}\},$$
$$\vdots$$
$$t_{j,k} = \min\{t : y_j(t) \geq C, t > t_{j,k-1}\}.$$

Proposition 1 implies that the sequence $\mathcal{T}_j := (t_{j,1}, \ldots, t_{j,k}, \ldots)$ has countably infinite elements a.s. for all $j \in \mathcal{V}$. We are now ready to present our main result, which states that $\lambda_j(t_{j,k})$, where $t_{j,k} \in \mathcal{T}_j$, converges to λ^* almost surely.

Lemma 2 (Almost Sure Convergence)**.** *Assume that Assumptions 1, 2, 3 and 4 are satisfied. Then the sequence $(\lambda_j(t_{j,k}))_{t_{j,k} \in \mathcal{T}_j}$ for any $j \in \mathcal{V}$ converges to the same random optimal incremental cost $\lambda^*(\omega)$ almost surely, i.e., $\mathbf{P}\left(\lim_{k \to \infty} \|\lambda_j(t_{j,k}; \omega) - \lambda^*(\omega)\| = 0\right) = 1$ for all $j \in \mathcal{V}$.*

The proof of Lemma 2 is somewhat involved and is given in Section IV-C. Basically, it contains two main steps. In the first step, we show that $\lambda_j(t+1)$ for all $j \in \mathcal{V}$ almost surely converges to a time-varying function (to be specified in the proof) increasingly well as time goes on. In the second step, we show that this time-varying function almost surely converges to an optimal incremental cost λ^*.

Remark 4. *Lemma 2 presents an almost sure convergence on the subsequence of $\lambda_j(t)$ over the time instants when $y_j(t)$ exceeds C. Such a convergent subsequence may not imply the convergence of the whole sequence. This is because when $y_j(t)$ is small, the deviation of $w_j(t)$'s cannot be ignored compared with $y_j(t)$. Therefore, the current proof can only characterize the convergence behavior over the time instants when $y_j(t)$ exceeds C. Notice that this convergence definition is consistent with the existing literature, see, e.g., [20], [29].*

Remark 5. *It should be emphasized that, in the consensus literature, the equivalence between consensus in mean square and almost sure consensus over random networks generated by i.i.d. stochastic matrices, can be readily established via a monotonicity argument on a sequence $\max_{i,j\in\mathcal{V}} \|z_i(t) - z_j(t)\|$, where $z_j(t)$ denotes the state of agent j at time t [30]. However, such an equivalent relation does not hold when studying our Algorithm 1 in mean square sense and in almost sure sense because the monotonicity property does not hold when the running-sum and the (sub)gradient-push protocols are used.*

Lemma 2 together with the update equation for the generation in (14d) and the zero duality between the primal problem (1) and the dual problem (2) leads to the following theorem.

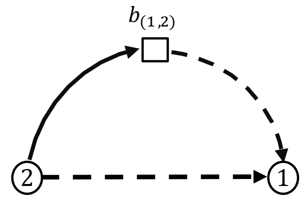
Theorem 1. *Assume that Assumptions 1, 2, 3 and 4 are satisfied. Then the distributed Algorithm 1 solves the optimal DER coordination problem in the sense that $\lambda_j(t_{j,k}) \to \lambda^*$ and $x_j(t_{j,k}) \to x_j^*$ a.s. as $k \to \infty$, where $t_{j,k} \in \mathcal{T}_j$, λ^* and x_j^* are respectively an optimal incremental cost and the optimal generation for generator j.*

C. Proof of Lemma 2

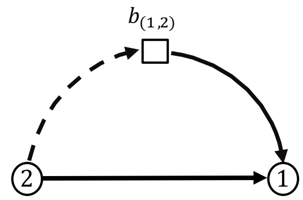
We will build our analysis by using the augmented graph idea from [20]. In particular, for each communication link $(j,i) \in \mathcal{E}$, we add a virtual buffer agent $b_{(j,i)}$ which stores the mass that may have otherwise been lost due to packet drops over the link (j,i). In doing so, we define the augmented graph $\mathcal{G}^a := (\mathcal{V}^a, \mathcal{E}^a)$ with

$$\mathcal{V}^a = \mathcal{V} \cup \{b_{(j,i)} | (j,i) \in \mathcal{E}\},$$
$$\mathcal{E}^a = \mathcal{E} \cup \{(b_{(j,i)}, i) | (j,i) \in \mathcal{E}\} \cup \{(j, b_{(j,i)}) | (j,i) \in \mathcal{E}\}.$$

Let $E := |\mathcal{E}|$. The augmented graph $\mathcal{G}^a$ has $\tilde{N} := N + E$ agents, where the first N agents are the ones in the original graph and the last E ones are the virtual buffer agents. Fig. 5 illustrates the augmented graph idea for a line graph consisting of two agents and one virtual agent. When there is a packet drop on $(1,2)$, with the link $(b_{(1,2)}, 2)$, the virtual buffer agent $b_{(1,2)}$ holds the mass that may otherwise been lost according to Fig. 5a. When the communication link $(1,2)$ becomes reliable,



(a) Packet drop on the communication link $(1,2)$.



(b) No packet drop on the communication link $(1,2)$.

Fig. 5. An augmented graph with a virtual buffer agent $b_{(1,2)}$, where the dashed lines have packet drops while the solid lines do not.

the information from agent 2 and the mass in the virtual buffer agent $b_{(1,2)}$ are transmitted to agent 1 according to Fig. 5b.

We next introduce the variables $\tilde{w}_\ell$, $\tilde{y}_\ell$, and $\tilde{v}_\ell$ for the virtual agents $\ell = b_{(j,i)}$ in the augmented graph, with initial conditions 0. The updates of $\tilde{w}_\ell$, $\tilde{y}_\ell$, and $\tilde{v}_\ell$ for $\ell = b_{(j,i)}$ are as follows:

$$\tilde{w}_l(t+1) = \begin{cases} \tilde{w}_l(t) + v_i(t)/(1+d_j^{\text{out}}), & \text{if } r_{ji}(t) = 0, \\ 0, & \text{otherwise}; \end{cases}$$

$$\tilde{y}_l(t+1) = \begin{cases} \tilde{y}_l(t) + y_i(t)/(1+d_j^{\text{out}}), & \text{if } r_{ji}(t) = 0, \\ 0, & \text{otherwise}; \end{cases}$$

and $\tilde{v}_l(t+1) = \tilde{w}_l(t+1)$. Given an arbitrarily given order $l_1, l_2, \ldots, l_E$ for the elements of $\mathcal{E}$, we define $\tilde{w} = [w_1, \ldots, w_N, \tilde{w}_{b_{l_1}}, \ldots, \tilde{w}_{b_{l_E}}]^\top$, $\tilde{y} = [y_1, \ldots, y_N, \tilde{y}_{b_{l_1}}, \ldots, \tilde{y}_{b_{l_E}}]^\top$, $\tilde{v} = [v_1, \ldots, v_N, \tilde{v}_{b_{l_1}}, \ldots, \tilde{v}_{b_{l_E}}]^\top$. With these notations, Algorithm 1 can be rewritten into a matrix form as

$$\tilde{w}(t+1) = M(t)\tilde{v}(t), \tag{15a}$$
$$\tilde{y}(t+1) = M(t)\tilde{y}(t), \tag{15b}$$
$$\lambda_j(t+1) = \frac{w_j(t+1)}{y_j(t+1)}, \; j \in \mathcal{V}, \tag{15c}$$
$$x_j(t+1) = \text{proj}_{X_j}\left(\nabla C_j^{-1}(\lambda)\right), \tag{15d}$$
$$\tilde{v}(t+1) = \tilde{w}(t+1) - \gamma(t+1)[x^\top(t+1) - \tilde{D}^\top, \mathbf{0}]^\top, \tag{15e}$$

where $\tilde{D} = [D_1, \ldots, D_N]^\top$. Some immediate observations from (15) are as follows: $M(t) \in \mathbb{R}^{\tilde{N}\times\tilde{N}}$ is a random matrix, depending on a set of random variables $\{r_{ji}(t) | (j,i) \in \mathcal{E}\}$ and is column stochastic.

To prove Lemma 2, we need the following lemma whose proof can be found in Appendix A.

Lemma 3. *Consider a sequence $(M(t))_{t\in\mathbb{N}}$ of column stochastic matrices. which are random and given by (15a). Assume that Assumption 1 is satisfied. Then there exists a uniform bound $\beta \in (0,1)$ and a sequence $(h(t))_{t\in\mathbb{N}}$ of stochastic vector such that $\limsup_{t\to\infty} \left|[M(t)M(t-1)\cdots M(s)]_{ij} - h_i(t)\right|^{1/(t-s)} \le \beta$ a.s., for all $i, j \in \mathcal{V}$ and $s \le t$.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2720761, IEEE Transactions on Smart Grid

We are now ready to prove Lemma 2. The proof is carried out into two steps: in the first step, we show that $\lim_{k\to\infty} |\lambda_j(t_{j,k}) - \frac{1}{N}\sum_{i=1}^{\tilde{N}} \tilde{v}_i(t_{j,k})| = 0$ a.s. for all $j \in \mathcal{V}$ (Recall that $\tilde{v}_i(t_{j,k})$ is updated following (15e).); in the second step, we show that $\lim_{t\to\infty} |\frac{1}{N}\sum_{i=1}^{\tilde{N}} \tilde{v}_i(t) - \lambda^*| = 0$ a.s.. The result then follows from the combination of these two steps. In what follows, the notation "$t_{j,k}$" will be written as "t_k" for short when the agent index j is specified in the context.

Denote $M(t : s) = M(t)\cdots M(s)$ (where $t \geq s$), $\bar{v}(t) = \frac{1}{N}\sum_{i=1}^{\tilde{N}} \tilde{v}_i(t)$ and $B = \max_{j\in\mathcal{V}} B_j$, with $B_j = \max_{x_j\in X_j} |x_j - D_j|$, for shorthand. With some algebra, we get

$$\tilde{w}_j(t+1) = [M(t:0)\tilde{v}(0)]_j + \sum_{s=1}^{t}[M(t:s)\epsilon(s)]_j, \quad (16)$$

$$\tilde{y}_j(t+1) = [M(t:0)\tilde{y}(0)]_j, \quad (17)$$

$$\mathbf{1}^\top\tilde{v}(t) = \mathbf{1}^\top\tilde{v}(0) + \sum_{s=1}^{t}\mathbf{1}^\top\epsilon(s), \quad (18)$$

where $\epsilon(t) = -\gamma(t)[x^\top(t) - \tilde{D}^\top, \mathbf{0}]^\top$. Then, for agent j we obtain

$$\begin{aligned}
&|\lambda_j(t_k+1) - \bar{v}(t_k)|\\
=&\left|\frac{w_j(t_k+1)}{y_j(t_k+1)} - \frac{\mathbf{1}^\top\tilde{v}(0) + \sum_{s=1}^{t_k}\mathbf{1}^\top\epsilon(s)}{N}\right|\\
\leq&\left|\frac{[M(t_k:0)\tilde{v}(0)]_j}{[M(t_k:0)\tilde{y}(0)]_j} - \frac{\mathbf{1}^\top\tilde{v}(0)}{N}\right|\\
&+\left|\frac{\sum_{s=1}^{t_k}[M(t_k:s)\epsilon(s)]_j}{[M(t_k:0)\tilde{y}(0)]_j} - \frac{\sum_{s=1}^{t_k}\mathbf{1}^\top\epsilon(s)}{N}\right|\\
:=&\xi_1(t_k) + \xi_2(t_k).
\end{aligned}$$

First we show that $\lim_{t_k\to\infty}\xi_1(t_k) = 0$ a.s.. To do so, notice that

$$\begin{aligned}
&\xi_1(t_k)\\
=&\left|\frac{[(M(t_k:0) - h(t_k)\mathbf{1}^\top)\tilde{v}(0)]_j}{[M(t_k:0)\tilde{y}(0)]_j}\right.\\
&\left.- \frac{[(M(t_k:0) - h(t_k)\mathbf{1}^\top)\tilde{y}(0)\mathbf{1}^\top\tilde{v}(0)]_j}{N[M(t_k:0)\tilde{y}(0)]_j}\right|,
\end{aligned}$$

where $h(t) := [h_1(t), \ldots, h_N(t), \mathbf{0}^\top]^\top$ is defined in (32) in the proof of Lemma 3. The equality follows because $\mathbf{1}^\top\tilde{y}(0) = N$. Denoting $M(t_k : s) - h(t_k)\mathbf{1}^\top := T(t_k : s)$, we further have

$$\begin{aligned}
\xi_1(t_k) \leq&\left|\frac{[T(t_k:0)\tilde{v}(0)]_j}{[M(t_k:0)\tilde{y}(0)]_j}\right| + \left|\frac{[T(t_k:0)\tilde{y}(0)\mathbf{1}^\top\tilde{v}(0)]_j}{N[M(t_k:0)\tilde{y}(0)]_j}\right|\\
\leq&\frac{2}{C}\max_{i\in\mathcal{V}}|[T(t_k:0)]_{ji}|\,\|v(0)\|_1, \quad (19)
\end{aligned}$$

where $C = \frac{1}{N^N}$ is defined in Proposition 1. From Lemma 3, there exist constants ϵ and L such that $\beta + \epsilon \in (0, 1)$ and

$$|[T(t:s)]_{ji}| \leq L(\lambda+\epsilon)^{t-s} \quad (20)$$

for all $i, j \in \mathcal{V}$. Then (19) and (20) together lead to

$$\xi_1(t_k) \leq \frac{2L}{C}(\lambda+\epsilon)^{t_k}\|\tilde{v}(0)\|_1. \quad (21)$$

Then we have $\xi_1(t_k) \to 0$ a.s. when $t_k \to \infty$.

To complete the first step, we only need to show that $\lim_{t_k\to\infty}\xi_2(t_k) = 0$ a.s.. To do so, we rewrite $\xi_2(t_k)$ as follows:

$$\begin{aligned}
&\xi_2(t_k) \quad (22)\\
=&\left|\frac{\sum_{s=1}^{t_k}[(M(t_k:s) - h(t_k)\mathbf{1}^\top)\epsilon(s)]_j}{[M(t_k:0)\tilde{y}(0)]_j}\right.\\
&\left.- \frac{\sum_{s=1}^{t_k}[(M(t_k:s) - h(t_k)\mathbf{1}^\top)\tilde{y}(0)\mathbf{1}^\top\epsilon(s)]_j}{N[M(t_k:0)\tilde{y}(0)]_j}\right|\\
=&\left|\frac{\sum_{s=1}^{t_k}[T(t_k:s)\epsilon(s)]_j}{[M(t_k:0)\tilde{y}(0)]_j} - \frac{\sum_{s=1}^{t_k}[T(t_k:0)\tilde{y}(0)\mathbf{1}^\top\epsilon(s)]_j}{N[M(t_k:0)\tilde{y}(0)]_j}\right|\\
\leq&\left|\frac{\sum_{s=1}^{t_k}[T(t_k:s)\epsilon(s)]_j}{[M(t_k:0)\tilde{y}(0)]_j}\right| + \left|\frac{\sum_{s=1}^{t_k}[T(t_k:0)\tilde{y}(0)\mathbf{1}^\top\epsilon(s)]_j}{N[M(t_k:0)\tilde{y}(0)]_j}\right|\\
\leq&\frac{1}{C}\sum_{s=1}^{t_k}(\max_{i\in\mathcal{V}}|[T(t_k:s)]_{ji}| + \max_{i\in\mathcal{V}}|[T(t_k:0)]_{ji}|)\,\|\epsilon(s)\|_1,
\end{aligned}$$

which with (20) together leads to

$$\xi_2(t_k) \leq \frac{2L}{C}\sum_{s=1}^{t_k}(\lambda+\epsilon)^{t_k-s}\|\varepsilon(s)\|_1. \quad (23)$$

From Assumption 3, we have $\lim_{t\to\infty}\gamma(t) = 0$. Then, $\xi_2(t_k) \to 0$ holds a.s. as $t_k \to \infty$ by [31, Lemma 3.1], which is presented as Lemma 4 in the Appendix B.

In the second step, we show that $\lim_{t\to\infty}|\bar{v}(t) - \lambda^*| = 0$ a.s.. From (15e), it follows that

$$\begin{aligned}
&\bar{v}(t) - \lambda^*\\
=&\frac{1}{N}\mathbf{1}^\top M(t-1)\tilde{v}_i(t-1) - \lambda^* - \frac{\gamma(t)}{N}\sum_{j=1}^{N}(x_j(t) - D_j)\\
=&\bar{v}(t-1) - \lambda^* - \frac{\gamma(t)}{N}\sum_{j=1}^{N}(x_j(t) - D_j),
\end{aligned}$$

which leads to

$$\begin{aligned}
&|\bar{v}(t) - \lambda^*|^2\\
=&|\bar{v}(t-1) - \lambda^*|^2 + \frac{\gamma^2(t)}{N^2}|\sum_{j=1}^{N}x_j(t) - D_j|^2\\
&- \frac{2\gamma(t)}{N}\sum_{j=1}^{N}(x_j(t) - D_j)(\bar{v}(t-1) - \lambda^*). \quad (24)
\end{aligned}$$

The cross term in (24) can be bounded as follows. Let $f_j(\lambda) = -\Psi_j(\lambda) - \lambda D_j$. By the concavity of (3), $f_j(\lambda)$ is a convex function of λ, and therefore

$$\begin{aligned}
f_j(\lambda^*) \geq& f_j(\lambda_j(t)) - (x_j(t) - D_j)(\lambda_j(t) - \lambda^*),\\
f_j(\lambda_j(t)) \geq& f_j(\bar{v}(t-1)) - B|\lambda_j(t) - \bar{v}(t-1)|,
\end{aligned}$$

which together yields

$$\begin{aligned}
&(x_j(t) - D_j)(\bar{v}(t-1) - \lambda^*)\\
=&(x_j(t) - D_j)(\lambda_j(t) - \lambda^*)\\
&+ (x_j(t) - D_j)\left(\bar{v}(t-1) - \lambda_j(t)\right)\\
\geq& f_j(\lambda_j(t)) - f_j(\lambda^*) - B|\lambda_j(t) - \bar{v}(t-1)|\\
\geq& f_j(\bar{v}(t-1)) - f_j(\lambda^*) - 2B|\lambda_j(t) - \bar{v}(t-1)|.
\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2720761, IEEE Transactions on Smart Grid

Combining the above inequality with (24), we have

$$\begin{aligned}|\bar{v}(t)-\lambda^*|^2 \leq & |\bar{v}(t-1)-\lambda^*|^2 + B^2\gamma^2(t) \\ & - \frac{2\gamma(t)}{N}\sum_{j=1}^{N}(f_j(\bar{v}(t-1)) - f_j(\lambda^*)) \\ & + \frac{4B\gamma(t)}{N}\sum_{j=1}^{N}|\lambda_j(t)-\bar{v}(t-1)|. \end{aligned} \quad (25)$$

We next show that $|\bar{v}(t)-\lambda^*|^2$ converges to a random variable a.s. by taking conditional expectations given $M(0),\ldots,M(t-1)$ at both sides of the inequality (25) and applying the supermartingale convergence theorem [32, Lemma 11], which is presented as Lemma 5 in the Appendix B for readers' convenience, with

$$z(t) = |\bar{v}(t-1)-\lambda^*|^2,\ \alpha_1(t) = 0,$$

$$u(t) = \frac{2\gamma(t)}{N}\sum_{j=1}^{N}(f_j(\bar{v}(t)) - f_j(\lambda^*)).$$

and

$$\alpha_2(t) = B^2\gamma^2(t) + \frac{4B\gamma(t)}{N}\sum_{j=1}^{N}|\lambda_j(t)-\bar{v}(t-1)|.$$

In order to apply the supermartingale convergence theorem (Lemma 5), the following conditions

$$\sum_{t=0}^{\infty}\alpha_1(t) < \infty \text{ a.s., and } \sum_{t=0}^{\infty}\alpha_2(t) < \infty \text{ a.s.}$$

need to be satisfied.

The first one is obvious since $\alpha_1(t) = 0$. To check the second condition, we first note that for the first term, $\sum_{t=0}^{\infty} B^2\gamma^2(t) < \infty$ since the step-size satisfies $\sum_{t=0}^{\infty}\gamma^2(t) < \infty$. Also note that by (21), (23) and Lemma 4(b), we can verify that $\sum_{t=0}^{\infty}\gamma(t)\sum_{j=1}^{N}|\lambda_j(t)-\bar{v}(t-1)| < \infty$ a.s.. Therefore, $\sum_{t=0}^{\infty}\alpha_2(t) < \infty$ a.s..

Hence, from the supermartingale convergence theorem (i.e., Lemma 4 in the Appendix B), we conclude: (i). $|\bar{v}(t;\omega)-\lambda^*|^2$ converges to a random variable a.s. for any given dual optimal solution λ^*, and (ii). $\sum_{t=0}^{\infty}\gamma(t)\sum_{j=1}^{N}(f_j(\bar{v}(t;\omega)) - f_j(\lambda^*)) < \infty$ a.s.. The rest of the proof is similar to that of [17, Lemma 7]. Since $\sum_{t=0}^{\infty}\gamma(t) = \infty$, we can show that with probability 1, there exists a convergent subsequence $(\bar{v}(t^l;\omega))$ such that $\bar{v}(t^l;\omega) \to v_*(\omega)$ and $f_j(\bar{v}(t^l;\omega)) \to f_j(\lambda^*)$. Therefore, $v_*(\omega)$ is a dual optimal solution by the continuity of f_j. Letting $\lambda^* = v_*(\omega)$ in (i), we have that $\bar{v}(t;\omega)$ converges to $v_*(\omega)$.

The proof is complete now.

Remark 6. *The proof technique used in the first step of the proof for Lemma 2 is motivated by [20]. In particular, we use the idea of virtual buffer agents to store the information that may have otherwise been lost due to the packet-dropping communication links. However, the situation here is much more complicated due to the additional gradient terms in (14e) which are needed to ensure that the algorithm converges to an optimal increment cost almost surely. Nevertheless, such gradient terms are well behaved in the sense that the multiplication of these gradient terms together with the diminishing step size asymptotically vanish. This nice property allows us to treat these additional terms as perturbations and show that the proposed distributed Algorithm 1 still converges under these perturbations. Of course, as shown in the proof of the first part of Lemma 2, it no longer converges to the average of the initial values almost surely, but converges to the average function $\bar{v}(t)$ increasing well as time goes on. In the second step, we show that $\bar{v}(t)$ converges almost surely to an optimal incremental cost.*

V. Case Studies

In this section, we present various case studies to illustrate and validate the proposed algorithm. We begin by revisiting the motivating example in Section III. We then show the performance of the proposed algorithm for the case where the communication links suffer from different probabilities of packet drops. Finally, we consider the effect of different splitting of total demand on the proposed algorithm.

A. Motivating Example Revisit

First, we return to the motivating example in Section III-B. This example shows that the previously proposed algorithm (10) always fails to converge, when each communication link $(j,i) \in \mathcal{E}$ suffers a packet drop with the same probability $q_{ji} = 0.1$, which are independent between communication links and between time instants. Let us consider the same scenario but with the newly developed Algorithm 1. Since the packet drops are random, the iteration results at each agent vary from one simulation to another. Nevertheless, the proposed Algorithm 1 always solves the DER coordination problem. The simulation results of a particular run are given in Fig. 6. As can been seen, even in the presence of packet-dropping communication links, each variable still converges to the optimal value as the case without packet drops shown in Fig. 3 yet with a slower convergence rate.

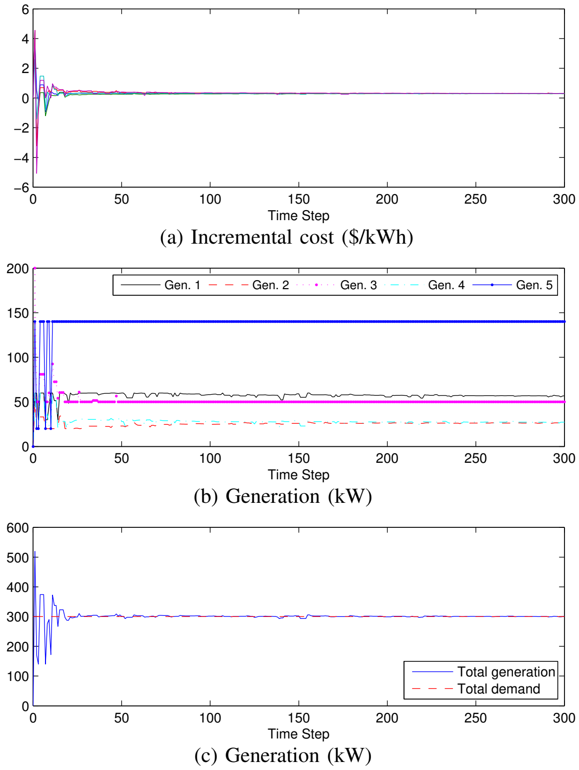
B. Different Probabilities for Packet-Dropping Communication Links

Notice that in the above case study, each communication link suffers a packet drop with the same probability. We now consider a more general case where different communication links have different probabilities of packet drops. In particular, the probabilities of packet drops in different communication links are $q_{14} = 0.1$, $q_{21} = 0.12$, $q_{31} = 0.08$, $q_{32} = 0.13$, $q_{35} = 0.03$, $q_{45} = 0.05$, $q_{52} = 0.15$, and $q_{53} = 0.09$. The simulation results of one realization are given in Fig. 7. As can been seen, even when communication links suffer packet drops with different probabilities, each variable still converges to the optimal value.

C. Different Splitting of Total Demand

We now consider the effect of different splitting of total demand. Recall that the virtual local demand at each bus is arbitrarily assignable as long as the summation is equal to the total demand, i.e., $\sum_{j=1}^{N} D_j = D$. In our previous case studies, we have chosen $D_1 = 40$ kW, $D_2 = 30$ kW,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2720761, IEEE Transactions on Smart Grid

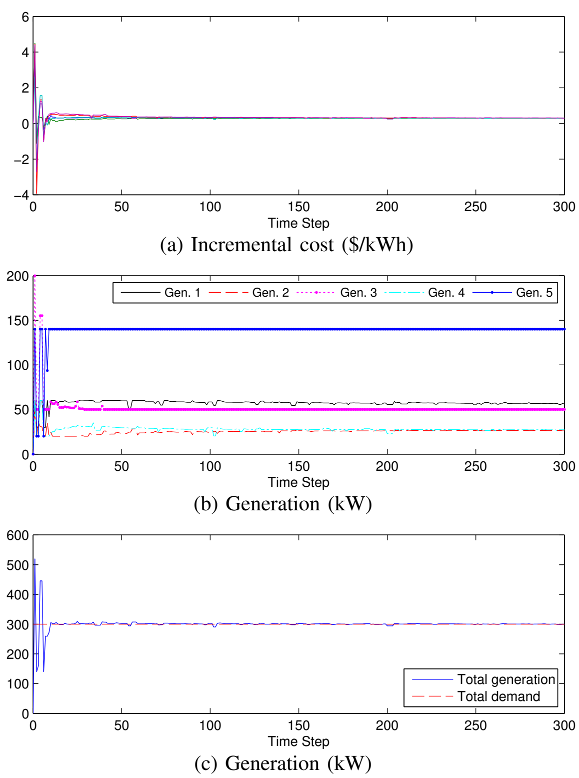


(a) Incremental cost ($/kWh)

(b) Generation (kW)

(c) Generation (kW)

Fig. 6. Results for networks with packet-dropping communication links



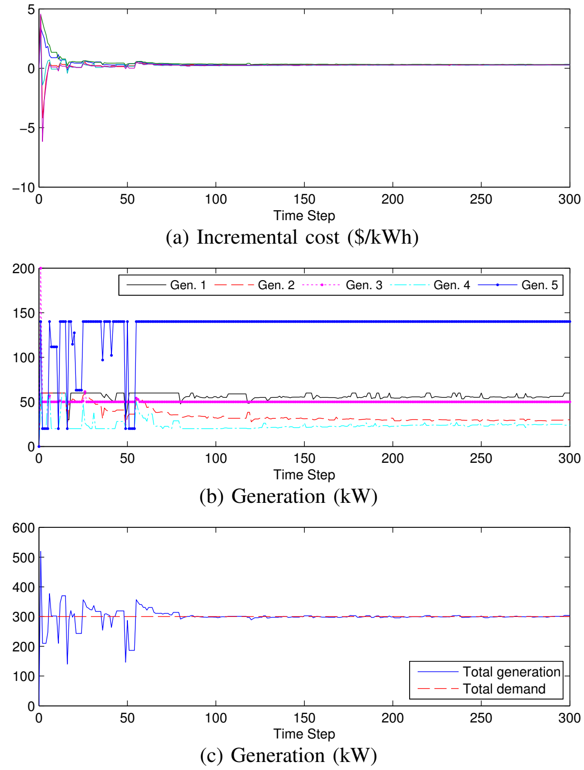
(a) Incremental cost ($/kWh)

(b) Generation (kW)

(c) Generation (kW)

Fig. 7. Results for networks with packet-dropping communication links

$D_3 = 100$ kW, $D_4 = 40$ kW, and $D_5 = 90$ kW. Now, let us assign the virtual local demand as $D_j = 60$ kW so that the total demand is also $D = \sum_{j=1}^{5} D_j = 300$ kW, The communication links have packet drops with the same probabilities as



(a) Incremental cost ($/kWh)

(b) Generation (kW)

(c) Generation (kW)

Fig. 8. Results for networks with packet-dropping communication links

those in Section V-B. We have tested the performance of the proposed algorithm by running the simulation various times—the proposed Algorithm 1 always solves the DER coordination problem. The simulation results of a particular run are given in Fig. 8. It shows that each variable still converges to the optimal value.

Remark 7. *In the above three case studies, the algorithm converges to the optimal values. However, the convergence rate are different, as shown in Fig. 6, Fig. 7, and Fig. 8. Intuitively speaking, the convergence rate depends on the probability of link failures, the splitting of D, and the step-size. However, the explicit relationship is difficult to obtain and is left as a future work.*

VI. Conclusions

This paper considers the distributed DER coordination problem over directed communication networks with packet-dropping links. We first showed by a motivating example that our previously developed distributed algorithm fails to solve the DER coordination problem in the presence of packet-dropping communication links. We then proposed a robustified extension of the distributed algorithm and showed that this robustified distributed algorithm is able to solve the DER coordination problem even in the presence of packet drops as long as the underlying communication network is strongly connected with a positive probability. One interesting direction is to explicitly characterize the convergence rate of the proposed algorithm. Another interesting direction is to extend the proposed distributed algorithm to accommodate additional physical models and constraints, such as transmission line loss, power flow, and transmission line flow constraints.

REFERENCES

[1] R. Lasseter, A. Akhil, C. Marnay, J. Stephens, J. Dagle, R. Guttromson, A. S. Meliopoulous, R. Yinger, and J. Eto, "Integration of distributed energy resources. The CERTS Microgrid concept," *Lawrence Berkeley National Laboratory*, 2002.

[2] S. Magnússon, P. C. Weeraddana, and C. Fischione, "A distributed approach for the optimal power-flow problem based on ADMM and sequential convex approximations," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 3, pp. 238–253, Sep. 2015.

[3] Z. Zhang and M.-Y. Chow, "Convergence analysis of the incremental cost consensus algorithm under different communication network topologies in a smart grid," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1761–1768, 2012.

[4] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[5] W. Ren, R. W. Beard, and E. Atkins, "Information consensus in multivehicle cooperative control," *IEEE Control Syst. Mag.*, vol. 27, no. 2, pp. 71–82, Apr. 2007.

[6] Z. Zhang and M.-Y. Chow, "Convergence analysis of the incremental cost consensus algorithm under different communication network topologies in a smart grid," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1761–1768, Nov. 2012.

[7] S. Kar and G. Hug, "Distributed robust economic dispatch in power systems: A consensus+innovations approach," in *Proc. IEEE Power Energy Soc. Gene. Meet.*, Jul. 2012.

[8] A. Dominguez-Garcia, S. Cady, and C. N. Hadjicostis, "Decentralized optimal dispatch of distributed energy resources," in *Proc. 51st IEEE Conf. Decision and Control (CDC)*, Dec. 2012, pp. 3688–3693.

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

[10] 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 Trans. Smart Grid*, vol. 6, no. 5, pp. 2836–2845, Nov. 2014.

[11] H. Xing, Y. Mou, M. Fu, and Z. Lin, "Distributed bisection method for economic power dispatch in smart grid," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3024–3035, Nov. 2015.

[12] T. Yang, D. Wu, Y. Sun, and J. Lian, "Minimum-time consensus based approach for power system applications," *IEEE Trans. Ind. Electron.*, vol. 63, no. 2, pp. 1318–1328, Feb. 2016.

[13] C. Zhao, J. He, P. Cheng, and J. Chen, "Consensus-based energy management in smart grid with transmission losses and directed communication," *IEEE Trans. Smart Grid*, 2016, to appear.

[14] Y. Xu, K. Cai, T. Han, and Z. Lin, "A fully distributed approach to resource allocation problem under directed and switching topologies," in *Proc. 10th Asian Control Conference (ASCC)*, May 2015.

[15] K. Cai and H. Ishii, "Average consensus on arbitrary strongly connected digraphs with time-varying topologies," *IEEE Trans. Autom. Control*, vol. 59, no. 4, pp. 1066–1071, Apr. 2014.

[16] T. Yang, J. Lu, D. Wu, J. Wu, G. Shi, Z. Meng, and K. H. Johansson, "A distributed algorithm for economic dispatch over time-varying directed networks with delays," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 5095–5106, Jun. 2017.

[17] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, Mar. 2015.

[18] L. Su and N. H. Vaidya, "Robust multi-agent optimization: Coping with packet-dropping link failures," 2016, available: arXiv:1606.08904v1.

[19] A. D. Dominguez-Garcia, C. N. Hadjicostis, and N. H. Vaidya, "Resilient networked control of distributed energy resources," *IEEE J. Select. Areas Commun.*, vol. 30, no. 6, pp. 1137–1148, Jul. 2012.

[20] C. N. Hadjicostis, N. H. Vaidya, and A. D. Dominguez-Garcia, "Robust distributed average consensus via exchange of running sums," *IEEE Trans. Autom. Control*, vol. 61, no. 6, pp. 1492–1507, Jun. 2016.

[21] R. Carli, G. Notarstefano, L. Schenato, and D. Varagnolo, "Analysis of newton-raphson consensus for multi-agent convex optimization under asynchronous and lossy communication," in *Proc. 54th IEEE Conf. on Decision and Control (CDC)*, Dec. 2015, pp. 418–424.

[22] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato, "Newton-raphson consensus for distributed convex optimization," *IEEE Trans. Autom. Control*, vol. 61, no. 4, pp. 994–1009, Apr. 2016.

[23] C. Godsi and G. Royle, *Algebraic Graph Theory*, ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 2001, vol. 207.

[24] D. P. Bertsekas, A. Nedić, and A. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.

[25] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.

[26] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *in Procs. 44th Annu. IEEE Symp. Found. Comp. Sci.*, Oct. 2003, pp. 482–491.

[27] A. D. Dominguez-Garcia and C. N. Hadjicostis, "Distributed strategies for average consensus in directed graphs," in *Proc. 50th IEEE Conf. Decision and Control (CDC)*, Dec. 2011, pp. 2124–2129.

[28] T. Charalambous, Y. Yuan, T. Yang, W. Pan, C. N. Hadjicostis, and M. Johansson, "Distributed finite-time average consensus in digraphs in the presence of time-delays," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 4, pp. 370–381, Dec. 2015.

[29] F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: distributed averaging using non-doubly stochastic matrices," in *Proc. IEEE Info. Theory*, Jun. 2010, pp. 1753–1757.

[30] J. Wu, Z. Meng, T. Yang, G. Shi, and K. H. Johansson, "Sampled-data consensus over random networks," *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4479–4492, Sep. 2016.

[31] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms forconvex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.

[32] B. T. Polyak, *Introduction to Optimization (Translation Series in Mathematics and Engineering)*. Optimization Software, 1987.

[33] E. Seneta, *Non-negative Matrices and Markov Chains*. Springer Science & Business Media, 2006.

[34] J. Hajnal and M. S. Bartlett, "Weak ergodicity in non-homogeneous markov chains," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 54, no. 02, pp. 233–246, 1958.

[35] W. Feller, *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, 2008, vol. 2.

APPENDIX A
PROOF OF LEMMA 3

Consider a sequence $(x(t))_{t\in\mathbb{N}}$, where $x(t)$ is defined as

$$x(t) = M(t)^\top \cdot \dots \cdot M(s)^\top x(s), \tag{26}$$

where $t \geq s$. Consider a specific event for time t, which is defined as

$$A(t) = \{\omega \in \Omega : r_{ij}(t;\omega) = 1,\ (j,i) \in \mathcal{E}\}.$$

It contains all situations, for which the links of $\mathcal{E}$ are all reliable. The outcome of $M(t)$ corresponding to the realizations in $A(t)$ is denoted as M. The probability of $A(t)$ is time-invariant and can be computed as

$$\mathbf{P}(A(t)) = \prod_{(j,i)\in\mathcal{E}} p_{ji} := p > 0.$$

When there is no packet drop over $\mathcal{G}$, there are paths from each virtual buffer agent to the actual agents and paths from any actual agents to the other actual agents (see Fig. 5b). Also observe that in M the diagonal elements corresponding to the actual agents are all strictly positive. Therefore, if we look at N consecutive matrices M jointly, there are at least one path from any actual or virtual agent to any actual ones. A mathematical formulation of this property is that: letting $T = M^N$ for $i \in \mathcal{V}$, there holds $T(i,:) > (1/N)^N$.

Then, following results on coefficients of ergodicity [33], [34], we have that: if

$$z' = T^\top z,$$

then

$$\max_{i\in\mathcal{V}} z' - \min_{i\in\mathcal{V}} z' \le \big(1-(1/N)^N\big)\big(\max_{i\in\mathcal{V}} z - \min_{i\in\mathcal{V}} z\big). \quad (27)$$

For ease of notation, in the rest of the paper we write $\beta_1 := 1-(1/N)^N$ for short. Evidently $\beta_1 \in (0,1)$. In contrast, for a generic column stochastic matrix, the equality (27) does not hold in general, while a weaken version truly holds: if $\tilde{M}$ is a generic column stochastic matrix and $z' = \tilde{M}^\top z$, then

$$\max_{i\in\mathcal{V}} z' - \min_{i\in\mathcal{V}} z' \le \max_{i\in\mathcal{V}} z - \min_{i\in\mathcal{V}} z. \quad (28)$$

Having obtaining the above properties for $M(t)$, next we are in a position to prove the conclusion. To this end, we let $T(k) = M((k-1)N+s)\cdots M(kN+s-1)$ and $z(k+1) = T(k)^\top z(k)$ with $z(1) = e_j$ for $j \in \mathcal{V}$. Then it can be seen that $z(k) = x((k-1)N+s)$. We further define the following notations based on $T(k)$. Let

$$\mathfrak{X}(k) = \max_{i\in\mathcal{V}}[z(k+1)]_i - \min_{i\in\mathcal{V}}[z(k+1)]_i$$

and

$$\xi(k) = \frac{\mathfrak{X}(k+1)}{\mathfrak{X}(k)}.$$

By (27) and (28), we get

$$\mathbf{E}[\log \xi(k)] \le \log(p^N \beta_1) := \beta_2 < 0. \quad (29)$$

By writing out common elements of fractions, we can represent the ratio of $\mathfrak{X}(k+1)$ and $\mathfrak{X}(1)$ by products of $\xi(k), \xi(k-1), \ldots, \xi(1)$, i.e.,

$$\frac{\mathfrak{X}(k+1)}{\mathfrak{X}(1)} = \xi(k)\cdots\xi(1) \quad (30)$$

Taking the logarithm over both sides of the above equation yields

$$\log \mathfrak{X}(k+1) - \log \mathfrak{X}(1) = \sum_{i=1}^{k} \log \xi(i).$$

Since $\log\xi(i)$'s have uniformly bounded covariances, Kolmogorovs strong law of large numbers [35] shows that, with probability 1,

$$\lim_{t\to\infty} \frac{1}{k}\sum_{i=s}^{k} \log \xi(i) = \mathbf{E}[\log \xi(k)],$$

which together with (29) implies that, with probability 1,

$$\lim_{k\to\infty} \frac{1}{k} \log \frac{\mathfrak{X}(k+1)}{\mathfrak{X}(1)} \le \beta_2.$$

Since $\mathfrak{X}(1) = 1$, the above equality further implies

$$\lim_{k\to\infty} \mathfrak{X}(k)^{1/k} \le e^{\beta_1}, \text{ with probability } 1.$$

Additionally, since $\max_{i\in\mathcal{V}} x(t) - \min_{i\in\mathcal{V}} x(t)$ is non-increasing by (28) and $z(k) = x((k-1)N+s)$, there hold with probability 1 that

$$\big[\max_{i\in\mathcal{V}} x(t) - \min_{i\in\mathcal{V}} x(t)\big]^{1/(t-s)} \le \big[\max_{i\in\mathcal{V}} z(\lceil\frac{t-s}{N}\rceil) - \min_{i\in\mathcal{V}} x(\lceil\frac{t-s}{N}\rceil)\big]^{1/(t-s)}$$

and

$$\big[\max_{i\in\mathcal{V}} x(t) - \min_{i\in\mathcal{V}} x(t)\big]^{1/(t-s)} \ge \big[\max_{i\in\mathcal{V}} z(\lceil\frac{t-s}{N}\rceil+1) - \min_{i\in\mathcal{V}} z(\lceil\frac{t-s}{N}\rceil+1)\big]^{1/(t-s)}.$$

The above upper and lower bounds converge to $(\lim_{k\to\infty}\mathfrak{X}(k)^{1/k})^{1/N}$, therefore

$$\lim_{t\to\infty} \big[\max_{i\in\mathcal{V}} x(t) - \min_{i\in\mathcal{V}} x(t)\big]^{1/(t-s)} \le e^{\beta_1/N} := \beta. \quad (31)$$

Due to the above relation, we further have

$$\lim_{t\to\infty} \big[\max_{i\in\mathcal{V}} x(t) - \min_{i\in\mathcal{V}} x(t)\big] = 0, \quad \text{a.s..}$$

When the above relation holds, we can define the following limit

$$h_i(s) = \lim_{t\to\infty} \frac{1}{N}\sum_{i\in\mathcal{V}} [M(t)^\top \cdots M(s)^\top]_{ij} \quad (32)$$

and have

$$\begin{aligned} &|[M(t)^\top\cdots M(s)^\top]_{ij} - h_i(s)| \\ &\le \left|[[M(t)^\top\cdots M(s)^\top]_{ij} - \frac{1}{N}\sum_{j\in\mathcal{V}}[M(t)^\top\cdots M(s)^\top]_{ij}\right| \\ &\quad + \left|h_i(s) - \frac{1}{N}\sum_{j\in\mathcal{V}}[M(t)^\top\cdots M(s)^\top]_{ij}\right| \\ &\le \max_{i\in\mathcal{V}} x(t) - \min_{i\in\mathcal{V}} x(t) + \sup_{t'\ge t}\Big|\frac{1}{N}\sum_{j\in\mathcal{V}}[M(t')^\top\cdots M(s)^\top]_{ij} \\ &\qquad - \frac{1}{N}\sum_{j\in\mathcal{V}}[M(t)^\top\cdots M(s)^\top]_{ij}\Big| \\ &\le 2\left(\max_{i\in\mathcal{V}} x(t) - \min_{i\in\mathcal{V}} x(t)\right). \end{aligned}$$

Here last inequality holds because, for any $t' \ge t$, $|\frac{1}{N}\sum_{j\in\mathcal{V}}[M(t')^\top\cdots M(s)^\top]_{ij}$ is a convex combination of $[M(t)^\top\cdots M(s)^\top]_{ij}$ for $j\in\mathcal{V}$. It together with (31) leads to that $\lim_{t\to\infty}|[M(t)^\top\cdots M(s)^\top]_{ij} - h_i(s)|^{1/(t-s)} \le \beta$ a.s. with $\beta \in (0,1)$.

When taking transpose of $M(t)^\top\cdots M(s)^\top$, we get $M(s)\cdots M(t)$. Since $M(t)$'s are i.i.d. distributed, reversing the order of the matrices does not change the probability, the result follows.

The proof is now complete.

Appendix B
Useful Lemmas

Lemma 4 ([31] Lemma 3.1)**.** *Let $(\gamma(t))_{t\in\mathbb{N}}$ be a scalar sequence.*

(a) If $\lim_{t\to\infty}\gamma(t) = \gamma$ and $0 < \beta < 1$, then $\lim_{t\to\infty}\sum_{\ell=0}^{t}\beta^{t-\ell}\gamma(\ell) = \frac{\gamma}{1-\beta}$.

(b) If $\gamma(t) \ge 0$ for all t, $\sum_{t=0}^{\infty}\gamma(t) < \infty$ and $0 < \beta < 1$, then $\sum_{t=0}^{\infty}(\sum_{\ell=0}^{t}\beta^{t-\ell}\gamma(\ell)) < \infty$.

Lemma 5 ([32] Lemma 11)**.** *Let $z(t)$, $u(t)$, $\alpha_1(t)$ and $\alpha_2(t)$ be nonnegative random variables and let*

$$\mathbf{E}[z(t+1)|\mathcal{F}_t] \le (1+\alpha_1(t))z(t) - u(t) + \alpha_2(t) \text{ a.s.},$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2720761, IEEE Transactions on Smart Grid

and

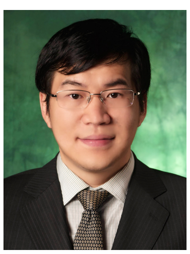
$$\sum_{t=0}^{\infty} \alpha_1(t) < \infty \text{ a.s., and } \sum_{t=0}^{\infty} \alpha_2(t) < \infty \text{ a.s.,}$$

where $\mathbf{E}[z(t+1)|\mathcal{F}_t]$ *denotes the conditional expectation for the given* $z(0), \ldots, z(t)$, $u(0), \ldots, u(t)$, $\alpha_1(0), \ldots, \alpha_1(t)$ *and* $\alpha_2(0), \ldots, \alpha_2(t)$. *Then*

$$z(t) \to z \text{ a.s., and } \sum_{t=0}^{\infty} u(t) < \infty \text{ a.s.,}$$

where $z \geq 0$ *is some random variable.*

Junfeng Wu received the B.Eng. degree from the Department of Automatic Control, Zhejiang University, Hangzhou, China, in 2009 and the Ph.D. degree in Electrical and Computer Engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2013. From September to December 2013, he was a Research Associate in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology, Hong Kong. He is currently a Postdoctoral Researcher at ACCESS (Autonomic Complex Communication nEtworks, Signals and Systems) Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden. His research interests include networked control systems, state estimation, and wirelesssensor networks, multi-agent systems.



Tao Yang (M'12) received the B.S. degree in Computer Science from Harbin University of Science and Technology in 2003, the M.S. degree with distinction in control engineering from City University, London in 2004, and the Ph.D. degree in electrical engineering from Washington State University in 2012. Between August 2012 and August 2014, he was an ACCESS Post-Doctoral Researcher with the ACCESS Linnaeus Centre, Royal Institute of Technology, Sweden. He is currently an Assistant Professor at the Department of Electrical Engineering, University of North Texas (UNT). Prior to joining the UNT, he was a Scientist/Engineer II with Energy & Environmental Directorate, Pacific Northwest National Laboratory. His research interests include distributed control and optimization in power systems, Cyber Physical Systems, networked control systems, vehicle platoon, and multi-agent systems.

Di Wu (M'12) received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiao Tong University in 2003 and 2006, respectively, and the Ph.D. degree in electrical and computer engineering from Iowa State University, Ames, in 2012. Dr. Wu is currently a Senior Engineer in Energy and Environment Directorate at Pacific Northwest National Laboratory. His research focuses on control and optimization of distributed energy resources and demand response, assessment of energy storage for grid applications, water-energy nexus, and power system dynamic simulation. He serves as an Associate Editor for the *IEEE Power and Energy Technology Systems Journal*.

Karanjit Kalsi (M'10) received the M.Eng. degree from the University of Sheffield, Sheffield, U.K., in 2006, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2010. He is currently a Power Systems Research Engineer at the Pacific Northwest National Laboratory, Richland, WA, USA.

Karl Henrik Johansson (SM'08/F'13) is Director of the Stockholm Strategic Research Area ICT The Next Generation and Professor at the School of Electrical Engineering, KTH Royal Institute of Technology. He received MSc and PhD degrees in Electrical Engineering from Lund University. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests are in networked control systems, cyber-physical systems, and applications in transportation, energy, and automation. He is a member of the IEEE Control Systems Society Board of Governors and the European Control Association Council. He has received several best paper awards and other distinctions, including a ten-year Wallenberg Scholar Grant, a Senior Researcher Position with the Swedish Research Council, the Future Research Leader Award from the Swedish Foundation for Strategic Research, and the triennial Young Author Prize from IFAC. He is member of the Royal Swedish Academy of Engineering Sciences, Fellow of the IEEE, and IEEE Distinguished Lecturer.