Distributed Power Apportioning with Early Dispatch for Ancillary Services in Renewable Grids

Sourav Patel, Student Member, IEEE, Blake Lundstrom, Senior Member, IEEE, Govind Saraswat, Member, IEEE, Murti V. Salapaka, Fellow, IEEE

Abstract—This article develops a distributed framework for coordinating distributed energy resources (DERs) in a power network to provide secondary frequency response (SFR) as an ancillary service to the bulk power system. A distributed finite-time protocol-based solution is adopted that allows each DER in the network to determine power reference commands. The distributed protocol respects information exchange constraints posed by a communication network layer while being robust to delays in the communications channels. The proposed framework enables coordinated response and control of the aggregated DERs by apportioning the share of generation that each DER needs to provide towards meeting any specified global SFR command while allowing for adjustments due to variability in generation and demand in order to prioritize renewable energy sources in the network. A novel early dispatch mechanism with brown start is synthesized to achieve initial DER response to changing SFR commands that is faster than state-of-the-art distributed approaches. The proposed power apportioning protocol is validated using an end-to-end power hardware-in-the-loop configuration at a distribution system scale with 40+ physical hardware DERs, underlying 7-MW power system model, a 250-DER communication topology with physical and simulated distributed controller nodes, varied communication protocols, and an underlying real-world power system model. Experimental results demonstrate the efficacy of the proposed method toward distributed coordination of hundreds of DERs for providing fast response at SFR timescales.

Index Terms—Distributed power apportioning, finite-time ratio consensus, secondary frequency response, ancillary services

I. INTRODUCTION

Dispatch methodologies for deploying ancillary services for reliable operation of the modern grid are changing significantly as conventional baseload generating units are being replaced by a large number of smaller DERs scattered throughout the network (termed as decentralization). The addition of renewable energy sources (RES) such as, wind and photovoltaic (PV) generation, with increased generation variability necessitates increased flexibility of dispatch operations [1], [2]. As a result, providing essential reliability services, such as, secondary frequency response (SFR), which corrects for imbalances between total generation and load in the system to restore frequency to its nominal value needs to emphasize DERs and RESs as a focus. An increase in power electronics-interfaced DERs, energy storage systems (ESS) and flexible loads interacting with the grid have shown promise in providing SFR, (for example, the Electric Reliability Council of Texas’s (ERCOT’s) responsive reserve services (RRS)), thereby demonstrating the viability of wide-scale adoption of DERs for such applications [3], [4]. A tighter integration of the distribution system operators (DSO) with aggregators, independent power producers and a large number of residential/commercial units with DER assets (‘prosumers’) as emerging potential participants is exigent.

To achieve distribution-level SFR with a decentralized framework, a large number of DERs must be coordinated on a fast timescale. Earlier works in the literature, such as [5], focused on implementing a centralized control approach for coordinating DERs, wherein a secondary centralized controller at the distribution level collects states of DERs via a communication network and sends dispatch commands to local actuators. Such centralized approaches lack flexibility and scalability in providing SFR, and they require expensive high-performance computing and high-speed communication networks to meet SFR requirements satisfactorily. Also, centralized control approaches have added challenges of reduced resiliency. In order to mitigate these challenges, distributed control approaches are proposed with DERs as a multi-agent system (MAS) [6].

Advantages of distributed approaches include coordination using only local computations, plug-and-play capability, and resiliency to node failures. A gather-broadcast method was presented in [7], and [8] developed a distributed average integral method; however, these methods are often sensitive to gain coefficients and can result in slow convergence to dispatch outputs. Reference [9] presented a distributed SFR approach but did not consider RESs and relies on heuristics to achieve dispatch requests. Distributed consensus-based algorithms form a primary approach for many applications where distributed decision making is needed; in this article, achieving consensus between agents on decision variables, forms a main thrust for realizing fast distributed SFR. Major challenges faced by consensus-based distributed approaches include: (i) inherent delays in communication channels that has significant impact on accuracy [10] and (ii) the asymptotic nature of co-
vergence of the algorithms entails in principle that the result is known only with a infinite horizon. Here, for coordination for SFR, the result of the distributed algorithm will determine the power dispatch command of each individual DER. Thus there is a need for a stopping criterion at each DER, which can be utilized to terminate the consensus algorithm and for determining its dispatch decision.

We now briefly describe the analytical frameworks on consensus based distributed approaches without any specific emphasis on the problem of distributed power apportioning. Reference [11] utilized the ratio-consensus algorithm to optimally coordinate DERs over time-varying directed communication networks for providing slower timescale tertiary support to the grid. However, [11] does not address communication delays that plague any practical implementation. Moreover, as alluded to earlier, a challenge with distributed algorithms that exchange information with neighbors and update their state multiple times over many iterations is the need to ascertain when to stop iterating and use the decision parameter for a subsequent action, such as determining how much power a DER needs to dispatch. Here, if the detection of convergence can be achieved (within a pre-specified tolerance value) by the nodes distributedly, algorithm run-times longer than necessary can be avoided, making it possible to employ low computational footprint, low cost devices. To circumvent the issue of asymptotic convergence, [12], [13] proposed finite-time algorithms to compute the consensus value using network observability; however, here, limitations of high computational footprint and large storage requirements at each node, render the approach unsuitable for applications, such as fast distributed SFR, where coordination with a fast response time is required. Reference [14] proposed a distributed finite time termination of ratio consensus, which built on [15] and [16], for frequency regulation in a network of islanded ac microgrids, but it did not consider communication delays in its formulation. Distributed finite-time termination of ratio consensus in the presence of bounded delays are presented in our earlier work [17].

All of the analytical works reported above are not instantiated to large scale coordination of DERs; thus an effective framework for distributed aggregation based SFR is currently absent. We remark that in [15], distributed apportioning was achieved for a few DERs. Here, the algorithms used did not address the issues of non-ideal nature of communication, nor any guarantees on the finite-time distributed stopping criterion are utilized. Without addressing these issues, the consensus based strategies remain inapplicable for practical sized distributed DER aggregation goals. Indeed, to the best of the authors’ knowledge, no prior work or framework has demonstrated a provably guaranteed distributed method, with instantiation of a large scale coordination to achieve DER based SFR. This article develops a scalable DSO-centric framework towards coordinating large numbers (1000+) of DERs to provide support to system operators via aggregators in dispatching SFR. The framework meets the SFR ancillary demand (global objective) of the system operator by aggregating distribution-level DERs in a distributed manner while respecting local capacity constraints of each DER. Coordinated response and control of aggregated DERs are achieved at SFR timescales: initial response times of less than 5 seconds and a ramp to response set point within 1 minute are demonstrated. Moreover, the distributed power apportioning framework comes with guarantees on reaching desired and feasible dispatch decisions in finite-time, even when the communication suffers from uncertainties such as delays. The developed framework’s efficacy is demonstrated using low-cost Raspberry Pi (Rpi) devices (with local communication and computational intelligence capabilities) interfacing with an underlying power controller layer. The framework here is robust to the presence of bounded delays in the communication channels. Further, early dispatch and brown-start mechanisms to enable participating DERs to respond to an SFR signal at timescales faster than the state-of-the-art distributed approaches are developed. We now summarize the major contributions of this article:

1) A distributed framework for coordinated power apportioning in the presence of communication delays that preserves the privacy of private values (capacity, output power) of the participating DERs.

2) An extension of distributed stopping criteria [16] where each DER in the network can detect convergence within a tolerance independently to formulate a novel early dispatch mechanism. This allows the network to achieve SFR faster than state-of-the-art distributed approaches (initial response time of less than 5 s and ramp time of 1 minute).

3) A brown-start approach whereby DERs, already participating in SFR, smoothly transition to new states in order to meet new system operator commands in the presence of changes in generation capacities, is developed. This contributions extends authors’ work in [18].

4) A large scale validation employing 40+ physical hardware DERs and 250 nodes on a real-world distribution system model. To the best of our knowledge, distributed DER coordination and control at such a scale is not demonstrated earlier with a large-scale validation.

II. SYSTEM DESCRIPTION

This section describes the system under study to develop the resource apportioning problem. A detailed description of graph theory and linear algebra notions that are used in the subsequent development are available in [19]. We consider DERs (such as a PV array and battery ESS), each interfaced with DC-AC inverter connected to local ac loads and a grid connection and to other units through a point of common coupling (PCC) in a microgrid or to the grid through an aggregator (see Fig. 1).

A. Communication Network of DER Units

In order to facilitate exchange of information to arrive at viable power commands for meeting ancillary demand services of the grid, we consider a network of DER units as a multi-agent system (MAS) with agents/nodes interacting with their neighbors, over a communication infrastructure (Communication Layer in Fig. 1). We consider a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where vertices $\mathcal{V} = \{1, 2, \ldots, N\}$ denote the DER nodes in the network, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges representing the communication topology overlay.
on the underlying power system infrastructure. DER units communicate with each other as dictated by the network topology where each node communicates with its neighbors using multiple allowable communication modalities (wired, wireless, hybrid). Communication can be bidirectional or directional.

In the network representation (see Fig. 1), a bidirectional channel is represented using a bi-directed edge and a directed channel with a directed edge. The underlying communication (implemented here using Rpi devices) is uncertain suffering from delays.

**Assumption 1.** $G = \{V,E\}$ is connected.

**Assumption 2.** For any node pair $i, j \in V$ and $(i,j) \in E$, the delay on the edge from node $j$ to node $i$, denoted as $\tau_{ij}$, satisfies $\tau_{ij} \leq \bar{\tau} < \infty$, where $\bar{\tau} > 0$.

Assumption 2 reflects the *partial synchrony* condition for a real-world communication network in distributed computing. This framework also adheres to communication protocols that guarantee no packet loss, such as Transmission Control Protocol-based websockets protocols.

We consider, an *aggregator* to be an entity interfacing with the DSO on one end and prosumers on the other. The aggregator is responsible for accumulating DERs available at the distribution level to provide a grid ancillary power command, $\rho_d(t_m)$, to the network of DERs. Here, the aggregator can communicate the command $\rho_d$ only to nodes in its communication neighborhood; here the aggregator is assumed to have $l \geq 1$ neighbors. The demand signal considered here is the SFR signal where the following specifications have to be met (i) initial response time of less than 5 s, (ii) ramp response time to set point within 1 minute, (iii) maintaining the desired SFR command for a maximum time period of at least 30 minutes or as required by the system operator. DERs have the capability to rapidly adjust their dispatch set points to output the SFR commands.

**B. Problem Formulation**

An aggregator, upon receiving a system operator’s SFR signal (can be commanded manually or automatically with less than 6 s refresh rates) at time instants $t_0 < t_1 < t_2 < \ldots$, sends a power command, $\rho_d(t_m)$, at time instant $t_m$, where $m \in \{0, 1, 2, \ldots\}$, to one or more aggregator-facing DER in the network. The $N$ participating DERs with a given graph topology, $G = \{V,E\}$, need to collectively meet the aggregator command, $\rho_d(t_m)$, while communicating only to their neighboring (neighbors determined by the communication network layer) DER units and respecting their individual resource (generation) constraints. Let $\pi_i^{\max}(t_m)$ and $\pi_i^{\min}(t_m)$ be the maximum and minimum generation capacities of the DERs at the $i^{th}$ DER, respectively, at instant $t_m$. Let $\pi_i^*(t_m)$ denote the steady-state reference power command to be supplied by the $i^{th}$ DER in response to an aggregator command, $\rho_d(t_m)$. The resource apportioning problem can be formulated mathematically as: for any time instant $t_m$ determine:

$$\left\{ \pi_i^*(t_m) \right\}_{i=1}^N$$

such that

$$\sum_{i \in V} \pi_i^*(t_m) = \rho_d(t_m)$$

$$\pi_i^{\min}(t_m) \leq \pi_i^*(t_m) \leq \pi_i^{\max}(t_m) \quad \text{for all } i \in V. \tag{1}$$

Furthermore, the viability of the resource apportioning problem holds when $\sum_{i=1}^N \pi_i^{\min}(t_m) \leq \rho_d(t) \leq \sum_{i=1}^N \pi_i^{\max}(t)$; otherwise for $\rho_d(t) \geq \sum_{i=1}^N \pi_i^{\max}(t)$, we set $\pi_i^*(t) = \pi_i^{\max}(t)$ for all $i \in V$.

We remark that the resource apportioning problem needs to be solved faster than the time interval between $t_{m-1}$ and $t_m$ when a new dispatch command is placed. In real-world power networks the maximum time to respond required by the system operator is in the order of minutes [20]. The approximate solution to [1] is found based on a finite-time termination criteria using a user-selected tolerance parameter which provides some flexibility wherein faster convergence results with a larger tolerance.
III. APPOINTING USING DISTRIBUTED AVERAGING

The resource apportioning problem with constraints can be solved using average consensus protocols. We first summarize the distributed averaging protocol \[16\], \[17\], \[21\], and its extension toward solving the resource apportioning problem for the asymptotic case.

A. Distributed Averaging Protocol

**Assumption 3.** Let \( p_{ij} \) denote the weight on information coming from node \( j \) to node \( i \). The weight matrix \( P(i, j) = p_{ij} \geq 0 \) associated with \( G \) is primitive (if it is irreducible and has only one eigenvalue of maximum modulus) and column stochastic (all elements of every column of the matrix sum to one).

**Assumption 4.** Any node \( i \in V \) in \( G = (V, E) \) has access to its own value at any instant without any delay.

**Definition 1.** (In-neighbor Set and Out-neighbor Set) In a graph \( G = (V, E) \), the in-neighbor set, \( N^{-}_i \), and out-neighbor set, \( N^{+}_i \), of node \( i \in V \), are given as \( N^{-}_i := \{ j \mid (i, j) \in E, i \neq j \} \) and \( N^{+}_i := \{ j \mid (j, i) \in E, i \neq j \} \), respectively. The cardinality of the out-neighbor set of a node \( i \in V \) is called the out-degree of the node denoted by \( D_i \).

Consider the following update iterations for states \( x \) and \( y \) maintained by all nodes in the network:

\[
x_{i}(k+1) = p_{ii}x_{i}(k) + \sum_{j \in N^{-}_{i}} p_{ij}x_{j}(k - \tau_{ij}), \quad (2)
\]

\[
y_{i}(k+1) = p_{ii}y_{i}(k) + \sum_{j \in N^{-}_{i}} p_{ij}y_{j}(k - \tau_{ij}), \quad (3)
\]

where \( \tau_{ij} \) is the delay in receiving data from node \( j \) to node \( i \).

**Theorem III.1.** (Ratio Consensus \[21\]) Suppose Assumptions \[7\] and \[1\] hold. Let the initial conditions for the numerator states be given as \( x(0) = [x_{1}(0) \ x_{2}(0) \ldots x_{N}(0)]^T \) and \( y(0) = [y_{1}(0) \ y_{2}(0) \ldots y_{N}(0)]^T \). Then the ratio \( x_{i}(k)/y_{i}(k) \) asymptotically converges to \( \sum_{i=1}^{N} \frac{x_{i}(0)}{y_{i}(0)} \) for all \( i = 1, \ldots, N \).

**Remark 1.** The ratio \( x_{i}(k)/y_{i}(k) \) is only well-defined when \( y_{i}(0) > 0 \), which is guaranteed when \( y_{i}(0) > 0 \) for all \( i \in V \).

**Remark 2.** (Distributed Synthesis) The weight matrix being column stochastic enforces the weights \( p_{ij} \) to be chosen in a purely distributed manner. A simple scheme is that \( i^{th} \) node sets the weights \( p_{ij} = \frac{1}{D_{j} + 1} \) for all \( j \in \{ N^{+}_{i} \cup i \} \) and communicates to node \( j \), \( \sigma_{ij}(k) = p_{ij}x_{i}(k) \). Each node \( j \) executes \( x_{j}(k) = p_{jj}x_{j}(k) + \sum_{i} \sigma_{ji}(k - \tau_{ji}) \) thus realizing \[1\] and \[3\].

B. Power Apportioning Protocol

Let the DER communication network be represented by a graph \( G(V, E) \). Let \( N_{d} \) denote the set, \( l = |N_{d}| \), of nodes directly communicating with the aggregator, referred to as command circulating nodes. Upon receiving a DSO signal for the dispatch of DER as part of the SFR at any time instant \( t_{m} \), the aggregator communicates the ancillary service request, \( \rho_{d}(t_{m}) \), to the network of DERs by communicating to the command circulating nodes in \( N_{d} \). At each time instant \( t_{m} \), each node \( i \) initializes two states \([r_{i}(0), s_{i}(0)]^T\) such that:

\[
r_{i}(0) = \left\{ \begin{array}{ll}
\frac{\rho_{d}(t_{m})}{l} - \pi_{i}^{\min}(t_{m}), & \text{if } i \in N_{d}, \\
-\pi_{i}^{\min}(t_{m}), & \text{if } i \notin N_{d},
\end{array} \right.
\]

\[
s_{i}(0) = \pi_{i}^{\max}(0) - \pi_{i}^{\min}(t_{m}), \quad \text{for all } i \in V. \tag{5}
\]

Each node executes versions of (2) and (3) as described below:

\[
r_{i}(k+1) = p_{ii}r_{i}(k) + \sum_{j \in N^{-}_{i}} p_{ij}r_{j}(k - \tau_{ij}), \tag{6}
\]

\[
s_{i}(k+1) = p_{ii}s_{i}(k) + \sum_{j \in N^{-}_{i}} p_{ij}s_{j}(k - \tau_{ij}), \tag{7}
\]

where, \( \pi_{i}^{\max}(t_{m}) \) and \( \pi_{i}^{\min}(t_{m}) \) denote the maximum and minimum power capacity of the \( i^{th} \) DER at time \( t_{m} \). We remark that, \( s_{i}(0) > 0 \) for all \( i \in V \) as \( \pi_{i}^{\max}(t_{m}) > \pi_{i}^{\min}(t_{m}) \).

**Lemma III.1.** Under Assumptions \[7\] and \[3\] the power apportioning protocol \[4\] converges asymptotically, i.e.:  

\[
\lim_{k \to \infty} r_{i}(k) = \frac{\sum_{i \in N_{d}} (\pi_{i}^{\max}(t_{m}) - \pi_{i}^{\min}(t_{m}))}{N\pi_{i}^{\min}(t_{m})}, \quad \text{for each } i \in V. \tag{8}
\]

**Proof.** The result can be obtained directly by noting that for a given time instant \( t_{m} \):

\[
\sum_{i=1}^{N} s_{i}(0) = \sum_{i \in N_{d}} \frac{\pi_{i}^{\min}(t_{m})}{l} + \sum_{i \notin N_{d}} \frac{\pi_{i}^{\min}(t_{m})}{l} \quad \text{and} \quad \sum_{i=1}^{N} s_{i}(0) = \sum_{i \in N_{d}} (\pi_{i}^{\max}(t_{m}) - \pi_{i}^{\min}(t_{m}))
\]

and applying Theorem III.1.

**Theorem III.2.** Let the power reference command for the \( i^{th} \) DER due to aggregator command at \( t_{m} \) be defined as, \( \pi_{i}^{*} := \pi_{i}^{\min}(t_{m}) + \lim_{k \to \infty} \frac{\rho_{d}(t_{m})}{l} (\pi_{i}^{\max}(t_{m}) - \pi_{i}^{\min}(t_{m})) \). Then \( \sum_{i=1}^{N} \pi_{i}^{*} \leq \sum_{i=1}^{N} \pi_{i}^{\max}(t_{m}) - \pi_{i}^{\min}(t_{m}) \). Therefore, \( \sum_{i=1}^{N} \pi_{i}^{*} = \rho_{d}(t_{m}) \) and \( \pi_{i}^{\min}(t_{m}) \geq \pi_{i}^{*} \) for all \( i \in V \).

**Proof.** See \[18\].

C. RES Prioritization

In order to demonstrate the dispatchability of available RES, this article also proposes an RES prioritization scheme. The motivation behind this objective is to dispatch spinning RES-based DERs as SFR before ESS for long-term reserve requirements. This is incorporated in the power apportioning protocol by setting \( \pi_{i}^{\min}(t_{m}) = \pi_{i}^{\min}(t_{m}) - \epsilon_{RES} \), where \( \epsilon_{RES} > 0 \) is a small. This envelope allows all available RES capacities at the initialization of the protocol to be used, allowing for RES prioritization as validated experimentally.

Theorem III.2 provides a distributed protocol to allocate resources to meet the demand \( \rho_{d}(t_{m}) \) by a DER network; however, it is clear that this protocol is not amenable to accommodating aggregator commands at subsequent time instants \( t_{m+1}, t_{m+2}, \ldots \) as the result in Theorem III.2 is asymptotic (where the desired dispatch, \( \pi_{i}^{*} \), of the \( i^{th} \) DER is determined only in the limit of iteration \( k \to \infty \)). In order to mitigate this issue, we now propose the formulation of the distributed finite-time termination protocol.
IV. DISTRIBUTED FINITE-TIME TERMINATION OF RESOURCE APportioning

In this section, we develop an algorithm using the Maximum and minimum consensus protocols for terminating the ratio consensus algorithm in finite-time based on a specified tolerance, $\rho$. We apply the results in [17] for distributed finite time termination of the ratio consensus algorithm. The resulting algorithm is presented in Algorithm 1.

Consider the maximum and minimum value of the ratio of consensus protocols (2–3) over all nodes within a time horizon $\bar{\tau}$ from any time instant $k$, given as:

$$ M(k) := \max_{j \in V} \frac{x_j(kr)}{y_j(kr)}, \quad y_j(k-r) \neq 0, \quad j \in V \tag{9} $$

$$ m(k) := \min_{j \in V} \frac{x_j(kr)}{y_j(kr)}, \quad y_j(k-r) \neq 0, \quad j \in V \tag{10} $$

Under assumptions [14], it can be shown that the global maximum (minimum) in the network is decreasing (increasing). Of particular interest are the maximum and minimum values over an “epoch” which is equal to $T(D, \bar{\tau}) = D(1+\bar{\tau})+\bar{\tau}$, which captures the upper bound on number of iterations required for any node in the network to communicate to any other node in the network.

**Theorem IV.1.** [17] Consider the initial ratio vector at the beginning of the $k$th epoch given by $x(kT):=\begin{bmatrix}x_1(kT) & \cdots & x_N(kT)\end{bmatrix}$ such that $\min_k x(kT) < \max_k x(kT)$, where, $k = 0, 1, 2, \ldots$. Then:

$$ M((k+1)T) < M(kT), \quad m((k+1)T) > m(kT). \tag{11} $$

The above result states the the maximum (minimum) over the network at the beginning of a epoch is strictly smaller (larger) than that at the beginning of a subsequent epoch. Thus the global maximum and minimum sampled at every epoch form strictly monotonic sequences. It can be further shown that these sequences converge to the consensus value, as specified in the next theorem.

**Theorem IV.2.** [17] \( \lim_{k \to \infty} M(kT) = \lim_{k \to \infty} m(kT) = \frac{\sum_{j=1}^{N} x_j(0)}{N}. \)

The theorem above establishes the asymptotic convergence of global maximum and minimum to the ratio of the sum of initial states. Thus, if we can determine the global maximum and minimum at each epoch distributedly, their difference can be used to bound how far each node’s ratio state is from the final consensus value.

We now introduce the maximum consensus and minimum consensus protocols that determine $M(kT)$ and $m(kT)$ in finite number of iterations.

A. Maximum and Minimum Consensus Protocols

The maximum consensus protocol (MXP) works as follows. Consider at time instant $\ell \bar{\tau}$, the value held by the $i^{th}$ node is $z_i(\ell \bar{\tau})$ with $z_i(0)$ being the initial value at $\ell = 0$. The $i^{th}$ node holds this value for $\bar{\tau}$ units of time; as $\bar{\tau}$ is the upper bound on delay, in $\bar{\tau}$ units of time node $i$ receives $z_j(\ell \bar{\tau})$ for all $j \in N_i$ and thus it can execute the update:

$$ z_i(\ell \bar{\tau} + \bar{\tau}) = \max_{j \in N_i} z_j(\ell \bar{\tau}), \quad z_i(\ell \bar{\tau} + k) = z_i(\ell \bar{\tau}), \quad \text{for} \quad k = 1, 2, \ldots, \bar{\tau} - 1. \tag{12a} $$

Similarly, the minimum consensus protocol (MNP) computes the minimum of the given initial node conditions $w(0) := [w_1(0), w_2(0), \ldots, \ w_N(0)]^T$ in a distributed manner. Consider at time instant $\ell \bar{\tau}$, the value held by the $i^{th}$ node is $w_i(\ell \bar{\tau})$ with $w_i(0)$ being the initial value at $\ell = 0$. Similar to the MXP protocol, $i^{th}$ node holds this value for $\bar{\tau}$ units of time and in $\bar{\tau}$ units of time node $i$ would have received $w_j(\ell \bar{\tau})$ for all $j \in N_i$ and thus it can execute the update:

$$ w_i(\ell \bar{\tau} + \bar{\tau}) = \min_{j \in N_i} w_j(\ell \bar{\tau}), \quad w_i(\ell \bar{\tau} + k) = w_i(\ell \bar{\tau}), \quad \text{for} \quad k = 1, 2, \ldots, \bar{\tau} - 1. \tag{12b} $$

**Remark 3.** MXP and MNP converge to the maximum and minimum of the initial conditions, respectively, within an epoch $T(D, \bar{\tau}) := D(1+\bar{\tau})+\bar{\tau}$, where $D$ is the upper bound on the diameter of the network [17]. Thus, each node can compute the global maximum and minimum distributedly in $D(1+\bar{\tau}) + \bar{\tau}$.

The MXP and MNP protocols at each node $i \in V$ are re-initialized at every epoch; here, for every $k = \theta(D(1+\bar{\tau}) + \bar{\tau})$, where $\theta = 1, 2, \ldots$, we initialize $z_i(0) = r_i(k)/x_i(0)$ and $w_i(0) = r_i(k)/x_i(0)$. In an epoch, each node $i$ can determine $M(\theta T) = \max_j z_j(\theta T)$ and $m(\theta T) = \min_j w_j(\theta T)$. Let $M(\theta T)$ and $m(\theta T)$ be the converged values from the MXP and MNP protocols after each epoch $T$. Then, it follows from Theorem IV.1 and Theorem IV.2 that $M(\theta T) \geq \sum_{i=1}^{N} z_i(0)$ and $m(\theta T) \geq \sum_{i=1}^{N} x_i(0)$. Thus, given any tolerance $\rho$, there exists a $\theta_0$ such that if $\theta \geq \theta_0$, $M(\theta T) - m(\theta T) \leq \rho$ where $M(\theta T) - m(\theta T)$ can be determined by every node in finite time. Thus, the consensus value $\sum_{i=1}^{N} z_i(0)$ can be determined with a tolerance $\rho$ in finite number of iterations. [17] establishes rigorously that given a threshold $\rho > 0$, Algorithm 1 terminates in a finite number of iterations.

**Remark 4.** The only global parameters needed by each node in order to implement the stopping criteria of (9) and (10) are upper bounds on both the maximum delay, $\bar{\tau}$, and diameter of the network, $D$, which can be enforced by design.

B. Early Dispatch Mechanism

In order to meet the requirement of initial response time of $< 5$ s for DERs participating in SFR support, we propose an early dispatch mechanism. In this mechanism, we use the ratio value from the most recent epoch to obtain reference power commands for each DER. The ratio value is guaranteed to stay within the global maximum and minimum ratios (computed via the MXP and MNP protocols) for the DER units running the power apportioning protocol due to the monotonicity property of the ratio consensus protocol (Theorem IV.2). The subsequences generated by the MXP-MNP protocols converging towards the final set point $\pi^*_i(t_m)$
within the tolerance, \( \rho \). Under the early dispatch mechanism the \( i \)th DER provides dispatch as:

\[
\pi^i(t_m, \theta) = \pi^i_{\text{min}}(t_m) + \frac{r_i(k)}{s_i(k)} (\pi^i_{\text{max}}(t_m) - \pi^i_{\text{min}}(t_m))
\]

where, \( k = \theta(D(1 + \tau) + \tau) \) (see Algorithm 1). Here \( \pi^i(t_m, \theta) \to \pi^i(t_m) \) for all \( i \in V \). For practical implementation, early dispatch is implemented for \( \theta > 3 \).

C. Brown Start: Changes in Power Demand as Input

In this section, we propose a brown-start mechanism where command circulating nodes require as input, the change in the requested power command from the previous time instant \( (\Delta \rho_d(t_m)) \) as the aggregator signal where for \( m > 0 \):

\[
\Delta \rho_d(t_m) = \rho_d(t_m) - \rho_d(t_{m-1}). \tag{14}
\]

The advantage of this mechanism is that the algorithm can be reinitialized with the converged state at the previous time instant rather than by following the initialization given by (4) with respect to an aggregator command \( \rho_d(t) \) that does not have regard for the operating state of the DER network. The latter approach results in larger deviations in output power between aggregator commands (see Fig. 2a). When \( \Delta \rho_d \) is small, the algorithm can converge very quickly as all the nodes will have initial values near the consensus state unlike the case where the algorithm is re-initialized with the new command, \( \rho_d(t_m) \). We propose a modified numerator update, \( r_i \), which is initialized as, \( r_i(0) := \)

\[
\begin{cases}
\frac{\Delta \rho_d(t_m)}{l} - \pi^i_{\text{min}}(t_m) + \pi^i_*(t_{m-1}), & \text{if } i \in \mathcal{N}_d, \\
-\pi^i_{\text{min}}(t_m) + \pi^i_*(t_{m-1}), & \text{if } i \notin \mathcal{N}_d
\end{cases}
\]

With the new initialization of \( r_i \) by (15), instead of (4), at a given instant \( t_m \) for \( m > 0 \), we have:

\[
\sum_{i=1}^{N} r_i(0) = \sum_{i \in \mathcal{N}_d} \left( \frac{\Delta \rho_d(t_m)}{l} - \pi^i_{\text{min}}(t_m) + \pi^i_*(t_{m-1}) \right) + \sum_{i \notin \mathcal{N}_d} \left( -\pi^i_{\text{min}}(t_m) + \pi^i_*(t_{m-1}) \right)
\]

Because the algorithm has converged in previous iteration, we have: \( \rho_d(t_{m-1}) = \sum_{i \in V} \pi^i_*(t_{m-1}) \). Thus, \( \sum_{i=1}^{N} r_i(0) = \sum_{i \in \mathcal{N}_d} \left( \frac{\rho_d(t_m)}{l} - \pi^i_{\text{min}}(t_m) \right) + \sum_{i \notin \mathcal{N}_d} \left( -\pi^i_{\text{min}}(t_m) \right) \), which is the same as when \( r_i \) was initialized by (4). Thus, Lemma III still holds.

D. Insufficient DER Capacity for SFR Dispatches

Typically in electricity markets, enough generation reserves and controllable load resources are committed such that North American Electric Reliability Corporation performance standards are always met; however, since in the proposed distributed framework the aggregator/DSO does not have access to real-time capacities of participating DERs, it is imperative that during insufficient capacity periods, i.e., when \( \sum_{i=1}^{N} \pi^i_{\text{max}}(t_m) < \rho_d(t_m) \), the maximum generation capacity be commanded in order to minimize the error between the commanded and dispatched output power. Algorithm 1 in such a scenario will command DER units to dispatch output power, \( \pi^i_*(t_m) = \pi^i_{\text{max}}(t_m) \). This is further validated in Test Case II.

Algorithm 1: Distributed finite-time termination of resource apportioning in the presence of communication delays (at each node \( i \in V \))

\[\begin{align*}
\text{Repeat:} & \quad \text{At each time instant } t_m, \ (m \in \{0, 1, 2, \ldots \}) \\
\text{Input:} & \quad \pi^i_{\text{min}}(t_m), \pi^i_{\text{max}}(t_m), s, t, D \\
\text{Initialize:} & \quad \text{if } m = 0 \quad \text{// Black start} \\
& \quad \text{Aggregator input: } \rho_d(t_m) \text{ if } i \in \mathcal{N}_d \\
& \quad \text{// Initialize } r_i(0) \text{ as in (4)} \\
& \quad \text{// Brown start} \\
& \quad \text{Aggregator input: } \rho_d(t_m) \text{ if } i \in \mathcal{N}_d \\
& \quad \text{// Initialize } r_i(0) \text{ as in (15)} \\
\text{end} \\
& \text{end} \\
& \text{end} \\
& \text{end} \\
\text{end}
\end{align*}\]

V. EXPERIMENTAL CONFIGURATION

A novel PHIL experimental configuration has been developed to validate the proposed distributed protocol containing 40+ physical hardware generation devices, 48 physical distributed controller nodes (DCNs), 202 simulated DCNs, and a real-time power system model. The components of the experimental configuration (Fig. 1) are described next.

A. Real-Time Distribution System Model (RT-DSM) Layer

The underlying RT-DSM used to represent the power network is based on a distribution network model from Aurora, Colorado, USA, which represents around 2000 customers and is augmented in this study to have 50.2% distributed PV RES penetration (both residential and commercial DER units —total PV 8.06 MW, battery ESS (BESS) capacity of 1.203 MW). The network has a peak load of 7.1 MW and has 5.285 MW of controllable DERs. Irradiance and load profiles
from the real-world system were used. The RT-DSM was executed on a 12-core OPAL-RT OP5707 using ePhasorSim at a time step of 10 ms. The capacities of participating DERs are shown in Table I. PHIL interfaces, including power amplifiers and feedback sensors for measurement, are provided to interconnect the physical hardware under test of Fig. 1 at five independent PCCs, $P_1$-$P_5$, in the RT-DSM.

B. Simulated Controller (SC) Layer

The simulated controller layer further consists of: 1) simulated local DER controllers (simLCs) that provide control signals to distributed PV inverters in the RT-DSM. simLC provides real-time measurements of minimum and maximum DER capacity to its DCN interface; 2) simulated distributed controller nodes (simDCNs) formulate the communication interfaces to simLC and provide power reference commands $(P_k^*, Q_k^*)$, see Fig. 1 based on the power apportioning protocol.

C. Communication Layer

Rpi-based DCNs, with inherent bounded time delays, are deployed to communicate with phLIS devices and simLCs. A communication topology is generated with the diameter and average node degree corresponding to the distribution network with 250 nodes overlayed on DCNs (simulated and Rpi-based) to represent the weakly-connected neighborhoods of RT-DSM. The WebSocket-based communication protocol between DCNs was implemented over Ethernet.

D. Physical Hardware Local Inverter System (phLIS) Layer

1) Residential-scale DER: This 2-kVA custom-built DER (Fig. 3 (a)) is a Type-4 Inverter connected to a PV simulator, residential loads, and is interfaced at P1 - 240-V PCC.
2) Utility-scale DER (Type-6 Inverter): This custom-built DER consists of a BESS (100 kW three-phase, Lithium-ion based 32.8 kWh) and a commercial PV inverter (100 kW, three-phase) (Fig. 3 (b)) coordinated via an integrated controller and interfaced at the three-phase PCC P2.
3) DER racks: Each DER rack (Fig. 3 (c),(d)) consists of two commercial-off-the-shelf residential PV string inverters (Type 1, 3/5 kVA and Type 2 3.8 kVA), a BESS inverter (Type 5, 5 kW) and three racks of 12 PV microinverters (Type 3, 320 W each). Each DER rack is capable of receiving a power reference command via MODBUS protocol, and the racks are interfaced at 240 V PCCs P3 - P5 as shown in Table I. Note that the phLIS DERs are scaled through PHIL interfaces to emulate large-scale hardware devices in the RT-DSM. These scalings are presented in Table I.

VI. RESULTS

The developed PHIL system is used for demonstrating the capability of the finite-time power apportioning protocol toward providing SFR as an ancillary service to the grid. Each phLIS is facilitated with a RPi-based module for communication with other network DER units. For validation, $\tau$ of 50 ms for Test Case I and 20 ms for Test Case II are chosen which are typical round trip times observed. The communication topology of Fig. 2 (b) is chosen.

![Figure 2](image_url)

Test Case I: Black Start: DER Aggregation and Early Dispatch with Fixed Capacities

We consider the scenario where the SFR commanded by the aggregator is $\rho_d$ (instead of $\Delta\rho_d$) and participating DERs have fixed capacities. The RT-DSM follows a net-load profile where the feeder consumption supplied by the bulk power system is 200 kW (Fig. 6 (a)) and DER units collectively provide 225 kW. A fixed irradiance of 800 W/m$^2$ is considered for Test Case I. Distributed PV inverters in the network have headroom available to provide SFR. At

| Type | Capacity (kW) | Device Count | PHIL Interface point |
|------|--------------|--------------|---------------------|
|      | physical     | sim.         | physical | sim.   |                 |
|      | actual       |              | 3.6       | 105/7  | P3-P5          |
| Type 1 | 3/5/5       | 100/75       | 3.6       | 105/7  | P3-P5          |
| Type 2 | 3.8         | 375          | 100       | 10     | P3-P5          |
| Type 3 | 0.32 each   | 6.25/8.33    | 12/12/12 | -      | P3-P5          |
| Type 4 | 2            | 2            | 1         | -      | P1             |
| Type 5 | 5/100       | 7-10         | 1/2       | 65     | P3-P5          |
| Type 6 | 100          | 1000         | -         | -      | P2             |
| Total  | 2,410 (MW)   | 2,864 (MW)   | 46        | 237    |                |

![Figure 3](image_url)
$8$

(a) Power reference commands at Rpi-DCNs, (b) dispatched output power for DER Rack 1 inverters.

Figure 4. (a) Power reference commands at Rpi-DCNs, (b) dispatched output power for DER Rack 1 inverters.

Figure 5. (a) Reference power command updates for 250 DER units in the network, (b) Total reference power command from the DER network with $\Delta \rho_d = 275 \text{ kW}$ and $\rho_d = 500 \text{ kW}$.

Figure 5 shows the power reference updates and the total active power reference for all the 250 participating DERs. These results establish that the desired SFR command is met and the response times achieved by the simulated and phLIS DERs satisfy the initial response time requirement of 5 seconds and ramp time to response of 1 minute. Figure 6(a), compares the case with and without DER' dispatch to meet active power in the feeder. The response of the feeder with respect to the increase in aggregated DER' output is presented in Fig. 6(b). The red line here corresponds to the change in dispatch command received by $N_d$. The secondary feeder voltages were remain well within $\pm 1\%$ of nominal (ANSI C84.1).

Test Case II: Brown Start: DER Aggregation with Early Dispatch for Time-Varying Capacities and Aggregator command

This test case emphasizes SFR due to continuous dispatches and variability in PV generation capability, requiring adjustments of minimum and maximum capacities of network DERs. In this scenario, the brown-start mechanism is implemented to achieve power apportioning due to changes in the system operator’s command and/or changes in generation capacities of participating PV and BESS DERs. Here, each PV inverter follows an irradiance profile sampled every 30 seconds. Initially, at $t = 0 \text{ s}$, $N_d = 2$ (here Mobile DER Rack 1: Type 1 and Type 2 inverters) receive the initial aggregator demand $\rho_d(t_0) = 3.62 \text{ MW}$ and within 5 s all the units in the network respond initially to meet the net command. A $\Delta \rho_d(t)$ of -0.82 MW at $t = 900 \text{ s}$ and +1.01 MW at $t = 1080 \text{ s}$ are provided as aggregator commands to the network via $N_d$ (see Fig. 7). At time periods, $0 < t < 900 \text{ s}$, $900 < t < 1080 \text{ s}$, and $t > 1080 \text{ s}$, the brown-start and early dispatch mechanisms allow DERs to be dispatched to achieve feasible solutions to (1) in the presence of variable solar irradiance and discharging/charging BESS DERs. The ramp response time (within $\pm 5\%$ of the steady-state value) to meet these commands by the aggregate DER network in the RT-DSM is $\sim 21 - 44 \text{ s}$ (to within SFR dispatch timescales) as shown by the purple dotted lines in Fig. 7. Fig. 8 shows the commanded power from physical DERs and simulated DERs in RT-DSM as well as the output response of selected physical hardware and simulated devices in the system. From $0 \leq t \leq 180 \text{ s}$, because the total demand can be solely met by DERs with RES prioritization, due to the high irradiance that allows the distributed PV units to operate at their full capacity, the BESS units are not dispatched. For $180 < t < 800 \text{ s}$, however, because of the
decrease in irradiance, the net output power of all distributed PV units decreases, causing the algorithm to command BESS units in the network to dispatch their cumulative maximum capacity of 1200 kW in order to minimize the difference in commanded and generated power. Similarly, at $t = 800$ s, when RES generation output was higher, BESS dispatch was adjusted back to meet the deficit. At $t = 900$ s, because the RES generation is higher than demanded $\rho_2(t)$, surplus PV generation is used to charge the BESS units. Finally, at $t = 1080$ s, both RES and BESS units are dispatched to meet and sustain the changed power commands of the aggregator with a ramp response time of less than 50 s. Finally, the secondary-side feeder voltages were observed to have voltage deviations to be within $\pm 5\%$.

VII. CONCLUSION

This article develops a scalable distributed framework for coordinating and aggregating large numbers of DERs to provide ancillary service support in the form of SFR to a bulk power system. The results of finite-time termination of ratio consensus were extended to propose a distributed power apportioning protocol to meet the time-varying aggregator command within a specified tolerance by participating DERs in the presence of bounded communication delays while prioritizing RES in the network. DER responses faster than state-of-the-art distributed approaches required for SFR services were achieved by implementing an early dispatch mechanism along with a brown-start approach to further improve DERs’ ramping performance. Experimental results, validated on a unique PHIL testbed at scale, show that the required performance metrics were met (initial and ramp responses were achieved to be less than 5 seconds and < 50 seconds for the setup, respectively).

REFERENCES

[1] M. Hedayati-Mehdiabadi, K. W. Hedman, and J. Zhang, “Reserve policy optimization for scheduling wind energy and reserve,” IEEE Transactions on Power Systems, vol. 33, no. 1, pp. 19–31, 2017.
[2] E. Ela and M. O’Malley, “Studying the variability and uncertainty impacts of variable generation at multiple timescales,” IEEE Transactions on Power Systems, vol. 27, no. 3, pp. 1324–1333, 2012.
[3] P. Moeller, J. Norris, and C. A. LaFleur, “Compliance filing of the midwest independent transmission system operator, inc. regarding dispatchable intermittent resources docket no. ER11-1991-00,” 2011.
[4] R. Konidena, “FERC order 811 levels the playing field for energy storage,” MRS Energy & Sustainability, vol. 6, 2019.
[5] K. Xi, J. L. Dubbledam, H. X. Lin, and J. H. van Schuppen, “Power-imbalance allocation control of power systems-secondary frequency control,” Automatica, vol. 92, pp. 72–85, 2018.
[6] A. Rassa, C. van Leeuwen, R. Spaans, and K. Kok, “Developing local energy markets: A holistic system approach,” IEEE Power and Energy Magazine, vol. 17, no. 5, pp. 59–70, 2019.
[7] F. Dörfler and S. Grammatico, “Gather-and-broadcast frequency control in power systems,” Automatica, vol. 79, pp. 296–305, 2017.
[8] C. Zhao, E. Mallada, and F. Dörfler, “Distributed frequency control for stability and economic dispatch in power networks,” in ACC, 2015.
[9] O. Méigel, T. Liu, D. J. Hill, and G. Andersson, “Distributed secondary frequency control algorithm considering storage efficiency,” IEEE Transactions on Smart Grid, 2017.
[10] L. Fang and P. J. Antsaklis, “Information consensus of asynchronous discrete-time multi-agent systems,” American Control Conference, 2005. Proceedings of the 2005, pp. 1883–1888, 2005.
[11] T. Yang, D. Wu, H. Fang, W. Ren, H. Wang, Y. Hong, and K. H. Johansson, “Distributed energy resource coordination over time-varying directed communication networks,” IEEE Transactions on Control of Network Systems, vol. 6, no. 3, pp. 1124–1134, 2019.
[12] S. Sundaram and C. N. Hadjicostis, “Finite-time distributed consensus in graphs with time-invariant topologies,” in 2007 American Control Conference. IEEE, pp. 711–716.
[13] T.-M. D. Tran and A. Y. Kibangou, “Distributed design of finite-time average consensus protocols,” in 4th IFAC Workshop on Distributed Estimation and Control in Networked Systems (NecSys), 2013.
[14] S. T. Cady, A. D. Domínguez-García, and C. N. Hadjicostis, “Finite-time approximate consensus and its application to distributed frequency regulation in islanded ac microgrids,” in System Sciences (HICSS), 2015 48th Hawaii International Conference on. IEEE, 2015, pp. 2664–2670.
[15] A. D. Domínguez-García and C. N. Hadjicostis, “Coordination and control of distributed energy resources for provision of ancillary services,” in Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on. IEEE, 2010, pp. 537–542.
[16] V. Yadav and M. V. Salapaka, “Distributed protocol for determining when averaging consensus is reached,” in 45th Annual Allerton Conf, 2007, pp. 715–720.
[17] M. Prakash, S. Talukdar, S. Attree, V. Yadav, and M. V. Salapaka, “Distributed stopping criterion for consensus in the presence of delays,” IEEE Transactions on Control of Network Systems, vol. 7, 2019.
[18] S. Patel, S. Attree, S. Talukdar, M. Prakash, and M. V. Salapaka, “Distributed apportioning in a power network for providing demand response services,” in 2017 IEEE International Conference on Smart Grid Communications. IEEE, 2017, pp. 38–44.
[19] R. Diestel, Graph theory. Springer-Verlag Berlin and Heidelberg, 2000.
[20] J. Ellison, V. W. Loose, and L. Tesfatsion, “A survey of operating reserve markets in us iso/ito-managed electric energy regions.” Sandia National Laboratories, Tech. Rep., 2012.
[21] C. N. Hadjicostis and T. Charalambous, “Average consensus in the presence of delays in directed graph topologies,” IEEE Transactions on Automatic Control, vol. 59, no. 3, pp. 763–768, 2014.