Stochastic Event-Triggered Algorithm for Distributed Convex Optimization

Kam Fai Elvis Tsang, Mengyu Huang, Ling Shi, *Senior Member, IEEE*, and Karl Henrik Johansson, *Fellow, IEEE*

***Abstract*—This article investigates the problem of distributed convex optimization under constrained communication. A novel stochastic event-triggering algorithm is shown to solve the problem asymptotically to any arbitrarily small error without exhibiting Zeno behavior. A systematic design of the stochastic event processes is then derived from the analysis on the optimality and communication rate with the help of a meta-optimization problem. Finally, a numerical example on distributed classification is provided to visualize the performance of the proposed algorithm in terms of convergence in optimization error and average communication rate with comparison to other algorithms in the literature. We show that the proposed algorithm is highly effective in reducing communication rates compared with algorithms proposed in the literature.**

***Index Terms*—Distributed optimization, event-triggered control, networked control systems.**

I. Introduction

IN RECENT years, distributed optimization problems, where multiple compute nodes collectively and distributively solve a global optimization problem, have gained a considerable amount of attention and interest. This is largely attributed to their wide range of potential applications, including but not limited to resource allocation, multirobot control, and machine learning.

In the distributed convex optimization problem, each compute node holds a private local objective function and does not have access to global or centralized information, such as the topology of the network or the overall objective function. Instead, each node can only rely on local information in a subset of the network to solve the problem by communicating with others to obtain a consensual solution. The setup is analogous to the multiagent consensus problem, where the agents aim to reach consensus in their states. The main difference is that distributed optimization problems require optimality besides consensus. Owing to the similarity, a numerous amount of literature on distributed optimization has related to consensus control protocols, such as the proportional consensus controller [1], [2], [3], [4], [5], [6] and proportional-integral (PI) controller [7], [8], [9], [10].

The study on distributed optimization has traditionally focused on discrete-time systems [11], [12]. Nedić and Ozdaglar [4] introduced the subgradient method combining the gradient descent method and consensus control and showed that their algorithm converges to a certain accuracy. Nedić and Olshevsky [13] further developed a subgradient-push algorithm based on the subgradient method and showed that it works under weaker assumptions. Lei et al. [14] proposed a distributed algorithm based on PI control with improved performance compared with the original subgradient method.

There has also been an increasing prevalence in the study of continuous-time distributed optimization. Wang and Elia [7] proposed a continuous-time distributed PI algorithm to tackle the problem that the original subgradient method by Nedić and Ozdaglar [4] is not sufficient in continuous time to reach both consensus and optimality. Lu and Tang [6] proposed a zero-gradient-sum second-order algorithm based on Hessian matrices, which converges to the optimal value exponentially.

The aforementioned algorithms all require continuous communication among agents as they rely on the information of other nodes in real time. This is detrimental in practice because distributed algorithms are often deployed on mobile or remote agents with constraints on communication bandwidth and energy. To overcome this drawback, event-triggered protocols are introduced. State-of-the-art event-triggering models for the consensus problem can be found in [15], [16], [17], [18], [19], [20], and [21], which have been the foundations for event-triggered algorithms in distributed optimization. Du et al. [22] and Yi et al. [23] adopted the dynamic event-triggering law introduced in [17] for the distributed optimization problems with single- and double-integrator agents, respectively. Li et al. [24] proposed an input-feedforward passive event-triggering algorithm. Chen and Ren [25] proposed a zero-gradient-sum event-triggering algorithm with time-varying threshold. Kia et al. [8] implemented an event-triggered distributed continuous-time PI algorithm.

Most event-triggering protocols introduced for consensus and distributed optimization are deterministic. However, it has

Manuscript received 5 October 2022; accepted 31 October 2022. Date of publication 16 December 2022; date of current version 17 August 2023. The work of Kam Fai Elvis Tsang and Karl Henrik Johansson was supported in part by the Knut and Alice Wallenberg Foundation, in part by the Swedish Research Council, and in part by the Swedish Foundation for Strategic Research. The work of Mengyu Huang and Ling Shi was supported by the Hong Kong Research Grants Council General Research Fund under Grant 16206620. Recommended by Associate Editor Bahman Gharesifard. *(Corresponding author: Kam Fai Elvis Tsang.)*

Kam Fai Elvis Tsang and Karl Henrik Johansson are with the Division of Decision and Control Systems, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden (e-mail: kfetsang@kth.se; kallej@kth.se).

Mengyu Huang and Ling Shi are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: mhuangak@ust.hk; eesling@ust.hk).

Digital Object Identifier 10.1109/TCNS.2022.3229769

been shown that stochastic event-triggering protocols are highly effective in reducing the communication rate for networked control systems [26], [27], [28]. By assigning different triggering probabilities to the events based on the local information and deterministic threshold, stochastic event-triggering protocols can prioritize more urgent events to reduce unnecessary communications. It is, therefore, worthwhile to investigate stochastic event-triggered (SET) algorithms for the distributed optimization problem to achieve better communication efficiency.

In this article, we solve the continuous-time distributed convex optimization problem by designing a SET algorithm. The main contributions of this article are summarized as follows.

1) We propose a novel stochastic event-triggering algorithm for the considered distributed optimization problem.
2) We prove that the proposed algorithm can solve the distributed optimization problem asymptotically to any accuracy, without exhibiting Zeno behavior.
3) We propose a systematic design of the stochastic process in the event-triggering protocol to achieve a tradeoff between the asymptotic optimization accuracy and communication rate, without the need for any global parameters.
4) We show that the proposed algorithm is effective in reducing peak communication rate and outperforms algorithms proposed in the literature.

Preliminary versions of this work have been presented at conferences [26], [28]. Major extensions in this article include a generalization of the design parameters and a systematic design for the stochastic event-triggering protocol. These extensions allow greater design flexibility and, therefore, the ability to adjust the performance tradeoff between optimality and communication, whether it is in the transient or steady state. The problem of the distributed support vector machine (SVM) is used as a numerical example to illustrate the effectiveness of the proposed algorithm in a practical data mining application and compared to state-of-the-art algorithms.

The rest of this article is organized as follows. Section II introduces notations and preliminaries. Section III describes the system setup and problem formulation, while Section IV presents the proposed algorithm to solve the problem. Section V provides analysis on the optimality and nonexistence of Zeno behavior. Section VI presents a systematic method for parameter selection and tuning to balance the tradeoff between residual optimization error and communication rate. Section VII illustrates the effectiveness and performance of the proposed algorithm against existing literature. Finally, Section VIII concludes this article.

II. Notations and Preliminaries

A. Linear Algebra

The matrix $I_n \in \mathbb{R}^{n\times n}$ denotes an $n \times n$ identity matrix with size and $\mathbf{1}_n$ denotes a vector with all entries being 1. The operator $\|\cdot\|_p$ is the p-norm for vectors and the induced p-norm for matrices. The norm operator $\|\cdot\|$ without subscripts denotes the 2-norm for vectors and Frobenius norm for matrices. For any matrices M and N with appropriate dimensions, $M \geq N$ means $M - N$ is positive semidefinite. The nth smallest eigenvalue for matrix M is denoted by $\lambda_n(M)$.

B. Algebraic Graph Theory

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected weighted graph, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the set of nodes and $\mathcal{E} \subset \{(i, j) : \forall i, j \in \mathcal{V}\}$ the set of edges. Let $\mathcal{N}_i$ denote the neighbors of node i: $\mathcal{N}_i = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\}$, for $i = 1, \ldots, N$. A graph is strongly connected if every node $v \in \mathcal{V}$ is reachable from any other node. In other words, for any starting and ending nodes $v_1, v_K \in \mathcal{V}$, there exists a path $\{v_1, v_2, \ldots, v_K\}$, where $(v_i, v_{i+1}) \in \mathcal{E}$ for $i = 1, 2, \ldots, K - 1$. The adjacency matrix $A = [a_{ij}] \in \mathbb{R}^{N\times N}$ is commonly used to describe the structure of a weighted graph, where a_{ij} is the weight of the edge $(i, j) \in \mathcal{E}$ and $a_{ij} = 0$ if $(i, j) \notin \mathcal{E}$. The degree matrix $D = \text{diag}\{d_{11}, \ldots, d_{NN}\} \in \mathbb{R}^{N\times N}$ captures the degree of each node such that $d_{ii} = \sum_j a_{ij}$. The Laplacian matrix $L = [L_{ij}]$ is then defined as $L = D - A$, which is positive semidefinite for the undirected and connected graph. In the remainder of this article, the shorthanded notation $\lambda_n(L) = \lambda_n$ will be used to avoid excessively tedious mathematical expressions.

C. Miscellaneous

For any real-valued function $F : \mathbb{R} \mapsto \mathbb{R}$, the notation $F(x^-)$ is the left limit of $\lim_{y\to x} F(y)$ should it exist. The function $W : \mathbb{R} \mapsto \mathbb{R}$ denotes the Lambert W function, which is the inverse function of $f(x) = x \exp(x)$. For any random variable $X \sim \text{Beta}(\alpha, \beta)$ with continuous beta distribution, the underlying distribution of the random variable $Z = (c - a)X + a$, where $0 \leq a < c$, is a reparameterization of the beta distribution, denoted by $\text{Beta}(\alpha, \beta, a, c)$.

III. Problem Formulation

Consider a networked control systems with N compute nodes represented by a weighted, undirected, and connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E}$ represents the set of bidirectional communication channels among nodes. Each node in the system holds a private local objective function $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ that is unknown to any other nodes. The objective is to solve

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) = \sum_{i=1}^{N} f_i(x). \tag{1}$$

Because the local objective functions f_i are private to each node, the optimization problem (1) needs to be solved distributively with cooperation among the nodes. A case in point is distributed classification where each node has a set of private data but wishes to find the globally optimal classifier for all data in the entire network. This example will be used in the numerical case study in Section VII. Note that problem (1) is equivalent to the constrained problem

$$\begin{aligned} \underset{x_1,x_2,\ldots,x_N\in\mathbb{R}^n}{\text{minimize}} \quad & \sum_{i=1}^{N} f_i(x_i) \\ \text{subject to} \quad & x_i = x_j \quad \forall i, j \in \mathcal{V}. \end{aligned} \tag{2}$$

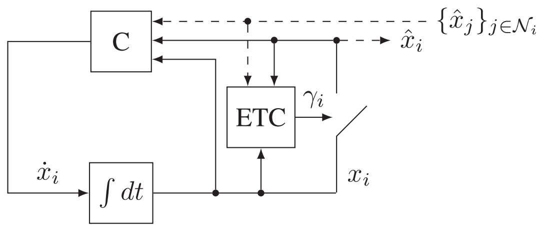


Fig. 1. Block diagram of the generic algorithm for node i.

We assume that the local objective functions $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ for all $i \in \mathcal{V}$ satisfy the following standing assumption.

Assumption 1: Each of the functions f_i is twice differentiable, strongly convex with convexity parameter m_i, and Lipschitz continuous gradient with Lipschitz constant $\mathcal{L}_i$.

We consider a broadcast-based communication model among the compute nodes, where the nodes broadcast their states to the neighbors when needed. The broadcasting of each node i is based on a binary decision variable $\gamma_i(t)$

$$\gamma_i(t) = \begin{cases} 1, & \text{node } i \text{ broadcasts its state at time } t \\ 0, & \text{otherwise} \end{cases}.$$

We denote the current state of node i as $x_i(t)$ and the last broadcast state as $\hat{x}_i(t)$

$$\hat{x}_i(t) = \begin{cases} x_i(t), & \gamma_i(t) = 1 \\ x_i(\tau_i(t)), & \gamma_i(t) = 0 \end{cases} \tag{3}$$

$$\tau_i(t) = \max\{k < t : \gamma_i(k) = 1\}. \tag{4}$$

Let us define the class of distributed algorithms to be considered. Each node has mainly two components, namely, computation (C) and stochastic event-triggering communication (ETC), based on the information received from the neighbors, as illustrated in Fig. 1. The dashed lines represent information exchange with other nodes, while solid lines represent information flow within the node. We denote such an algorithm a SET algorithm.

Let $x(t) = [x_1(t)^T, x_2(t)^T, \dots, x_N(t)^T]^T$, $\hat{x}(t) = [\hat{x}_1(t)^T, \hat{x}_2(t)^T, \dots, \hat{x}_N(t)^T]^T$, $x^\star = \arg\min_x f(x)$, and

$$\varepsilon(t) = \frac{1}{N}(x(t) - x^\star \otimes \mathbf{1}_N)^T (x(t) - x^\star \otimes \mathbf{1}_N) \tag{5}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \|x_i(t) - x^\star\|^2 \tag{6}$$

be the global optimization error. We adopt the following definitions of optimality.

Definition 1: A SET algorithm solves (1) to an accuracy of ϵ asymptotically if there exists $\epsilon > 0$ such that

$$\lim_{t\to\infty} \varepsilon(t) \le \epsilon.$$

In addition, it solves (1) in expectation to an accuracy of ϵ if there exists $\epsilon > 0$ such that

$$\lim_{t\to\infty} \mathbb{E}\left[\varepsilon(t)\right] \le \epsilon.$$

TABLE I
SUMMARY OF PARAMETERS AND THEIR EFFECTS

Parameter	Convergence Rate	Optimisation Error	Communication Rate
α ↗	↗	–	↗
a ↗	–	↘	↗
κ ↗	–	↗	↘
β ↗	↘	↗	↘
δ ↗	–	↗	↘

The objective of this work is to design the SET algorithm in Fig. 1 to solve problem (1) according to Definition 1 without displaying Zeno behavior.

IV. SET Algorithm

In this section, we propose a specific SET algorithm, which solves the problem (1), as shown in Section V. We start with the computation followed by the ETC.

A. Local Computation

Each compute node in the network computes the state locally as follows:

$$x_i(0) = \arg\min_x f_i(x)$$

$$\dot{x}_i(t) = -\alpha\left(\nabla^2 f_i(x_i(t))\right)^{-1} \sum_{j=1}^{N} L_{ij}\hat{x}_j(t) \tag{7}$$

where $\alpha > 0$ is a constant gain. This is inspired by the zero-gradient-sum algorithm in [6] and [22]. One can also interpret the above computation law as a combination of the standard consensus control [3] and Newton's method for optimization.

B. Event-Triggered Communication

Let $e_i(t) = \hat{x}_i(t) - x_i(t)$ be the local state error. The event-triggered broadcast of each node i is given by

$$\gamma_i(t) = \begin{cases} 1, & \xi_i(t) > \kappa \exp\left(-L_{ii}\rho_i(t)/\delta(t)\right) \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

$$\rho_i(t) = \|e_i(t)\|^2 + \frac{\beta}{L_{ii}}\sum_{j=1}^{N} L_{ij}\|\hat{x}_j(t) - \hat{x}_i(t)\|^2 \tag{9}$$

where $\xi_i(t) \in [a, 1]$, for some $a \in (0, 1)$, are ergodic stationary random processes with identical probability density function f_ξ for all nodes $i \in \mathcal{V}$. The scalars $a \in (0, 1), \kappa \in (1, \infty), \beta \in (0, 1/4)$ are parameters to be designed. The function $\delta(t)$ allows for a fine adjustment of the event-triggering law in the transient. We consider the set of functions where there exists $\delta_{\max}$ and δ_∞ such that $\lim_{t\to\infty}\delta(t) = \delta_\infty$ and $\delta_\infty \le \delta(t) \le \delta_{\max}$.

A summary of parameters and the expected direct effects on the performance is provided in Table I. The cells marked with "–" mean that the corresponding effects are uncertain.

The intuition behind the proposed stochastic event-triggering law (8) is to extend the existing deterministic trigger in the form

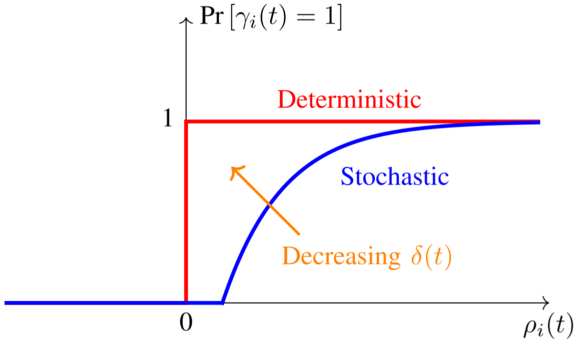


Fig. 2. Intuition of the proposed stochastic event trigger.

of

$$\gamma_i(t) = \begin{cases} 1, & \rho_i(t) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

such as those proposed in [17] and [25], by assigning a probability for each $\rho_i(t) > 0$, instead of triggering whenever the threshold functions $\rho_i(t)$ exceeds 0. In other words, it only triggers with a probability when $\rho_i(t) > 0$, not with certainty. Moreover, the probability should be monotonically nondecreasing in $\rho_i(t)$, such that it prioritizes more urgent cases, which corresponds to a higher value of $\rho_i(t)$.

The effects of $\delta_i(t)$ can also be observed in Fig. 2. Decreasing $\delta(t)$ is equivalent to pushing the probability function toward the deterministic case. This allows system designers to fine-tune the transient behavior and performance. For example, we may want $\delta(t)$ to be large for a period of time to reduce the communication required during convergence. When the system is close to consensus, however, the compute nodes may need to communicate more often in order to achieve higher precision, requiring a smaller $\delta(t)$ as $t \to \infty$.

V. Main Results

In this section, we present the main results of analysis on the optimality and nonexistence of Zeno behavior, which directly impact the feasibility in practical deployment, of the proposed stochastic event-triggering law (8).

A. Preliminary Analysis

We first prove the boundedness of the global optimization error $\varepsilon(t)$ under the proposed algorithm in Section IV.

Lemma 1 (see [26, Lemma 2]): Let $K_N = I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$. For an undirected graph $\mathcal{G}$

$$0 \le \lambda_2 K_N \otimes I_N \le L \otimes I_N.$$

Lemma 2 (see [26, Lemma 5]): For an undirected graph $\mathcal{G}$ and the event-triggering law (8)

$$\hat{x}(t)^T(L \otimes I_n)\hat{x}(t) \ge \zeta_1 x(t)^T(K_N \otimes I_n)x(t) - 2\zeta_2\delta(t)\sum_{i=1}^{N}(\ln\kappa - \ln\xi_i(t))$$

with $L_{\min} = \min_i L_{ii}$ and

$$\zeta_1 = \frac{\lambda_2 L_{\min}}{2L_{\min} + \lambda_n}, \qquad \zeta_2 = \frac{\lambda_n}{L_{\min} + \lambda_n}.$$

The following result states that the optimization error is uniformly upper bounded at all time. It is used later on to show the well-posedness and optimality of the SET algorithm.

Lemma 3: With the SET algorithm (7), (8), the global optimization error is bounded regardless of $\delta(t)$ as follows:

$$\varepsilon(t) \le \frac{2}{N m_{\min}}\left[\left(V_0 - \frac{\omega^\star\delta_{\max}}{\phi^\star}\right)e^{-\phi^\star t} + \frac{\omega^\star\delta_{\max}}{\phi^\star}\right]$$

$$\omega^\star = N\alpha(1-(1-4\beta)\zeta_2)(\ln\kappa - \ln a)$$

$$\phi^\star = \frac{\alpha\zeta_1(1-4\beta)}{\mathcal{L}_{\max}}$$

for any realizations of the stochastic processes $\xi_i(t)$, where $V_0 = \sum_{i=1}^{N}(f_i(x^\star) - \min_x f_i(x))$.

Proof: See Appendix A. ∎

The upper bound in Theorem 3 is likely not tight. However, the bound remains important to show that the proposed algorithm excludes Zeno behavior and the existence of $\mathbb{E}[\varepsilon(t)]$ in later sections.

B. Guarantee of Minimum Interevent Interval

In this section, we provide analysis on the nonexistence of Zeno behavior in the proposed stochastic event-triggering law for distributed optimization. Let t_k^i be the kth triggering time for agent i, $K_i(t) = \max\{k > 0 : t_k^i \le t\}$, and

$$\tau_i(t) = t^i_{K_i(t)} - t^i_{K_i(t)-1}.$$

Theorem 1: The SET algorithm (7), (8) does not exhibit Zeno behavior if

1) $\delta_\infty > 0$;
2) $\delta(t) \ge \delta_0 e^{-\eta t}$

for some finite $\delta_0 > 0$ and $\eta \in (0, \phi^\star)$. More specifically, there exists a strictly positive lower bound on the interevent interval $\tau_i(t) > \bar{\Delta} > 0$, $\forall i \in \mathcal{V}, t \in [0, \infty)$ with

$$\bar{\Delta} = \begin{cases} \Delta_1, & \delta_\infty > 0 \\ \Delta_2, & \delta(t) \ge \delta_0 e^{-\eta t} \end{cases},$$

$$\Delta_1 = \sqrt{-\frac{\delta_\infty \ln\kappa}{U^2} W\left(-\frac{2}{e^2}\right)\left(W\left(-\frac{2}{e^2}\right) + 2\right)},$$

$$\Delta_2 = \frac{2}{\eta} W\left(\frac{\eta}{\varsigma_1 + \varsigma_2}\sqrt{\frac{\delta_0 \ln\kappa}{L_{\max}}}\right)$$

where $L_{\max} = \max_i L_{ii}$ and

$$\varsigma_1 = \frac{4\sqrt{2}\alpha L_{\max}}{(1-2\sqrt{\beta})m_{\min}^{3/2}}\sqrt{\left|V_0 - \frac{\omega^\star\delta_0}{\phi^\star - \eta}\right|},$$

$$\varsigma_2 = \frac{4\alpha L_{\max}\sqrt{\delta_0}}{(1-2\sqrt{\beta})m_{\min}}\left(\sqrt{\frac{\omega^\star}{\phi^\star - \eta}} + \sqrt{\frac{\ln\kappa - \ln a}{L_{\min}}}\right),$$

$$U = \frac{4\alpha L_{\max}}{m_{\min}(1-2\sqrt{\beta})}\left(\sqrt{\frac{2}{m_{\min}}\max\left(V_0, \frac{\omega^\star\delta_{\max}}{\phi^\star}\right)}\right.$$

$$+\sqrt{\frac{\delta_{\max}(\ln\kappa-\ln a)}{L_{\min}}}\Bigg).$$

Proof: See Appendix B. ∎

Remark 1: Theorem 1 shows the well-posedness of the proposed SET algorithm. The result does not rely on any extra assumptions on the networked system, but only the user-defined function $\delta(t)$, which can be easily satisfied.

Remark 2: If restricted to deterministic settings, the SET algorithm (7), (8) is a reparameterization of some algorithms in literature, such as [22] and [25]. Compared with [25], which is restricted to sampled-data systems, the proposed SET algorithm removes such restriction. In contrast to some literature [8], [22], we provide an explicit lower bound on the interevent interval in addition to proving the nonexistence of Zeno behavior.

C. Optimality

We are now ready to present the theorem on the optimality of the proposed algorithm.

Theorem 2: The SET algorithm (7), (8) solves the distributed optimization problem (1) asymptotically with

$$\lim_{t\to\infty}\varepsilon(t)\le\epsilon_1=\frac{2\mathcal{L}_{\max}\delta_\infty(1+(1-4\beta)\zeta_2)(\ln\kappa-\ln a)}{\zeta_1 m_{\min}(1-4\beta)} \quad (11)$$

$$\lim_{t\to\infty}\mathbb{E}\left[\varepsilon(t)\right]\le\epsilon_2=\frac{2\mathcal{L}_{\max}\delta_\infty(1+(1-4\beta)\zeta_2)(\ln\kappa-\mu_{\ln})}{\zeta_1 m_{\min}(1-4\beta)} \quad (12)$$

for any realizations of the stochastic processes $\xi_i(t)$, where $\mathcal{L}_{\max}=\max_i\mathcal{L}_i$, $m_{\min}=\min_i m_i$, and $\mu_{\ln}=\mathbb{E}[\ln\xi_i(t)]$.

Proof: See Appendix C. ∎

Remark 3: Note that $\mu_{\ln}=\mathbb{E}[\ln\xi_i(t)]\in[\ln a,0]$ is a constant because $\xi_i(t)$ is ergodic and stationary for all $i\in\mathcal{V}$.

Theorem 2 implies that for any $\epsilon>0$, there exists a δ_∞ such that $\lim_{t\to\infty}\varepsilon(t)\le\epsilon$ and $\lim_{t\to\infty}\mathbb{E}[\varepsilon(t)]\le\epsilon$ by setting

$$\delta_\infty\le\frac{\zeta_1 m_{\min}(1-4\beta)\epsilon}{2\mathcal{L}_{\max}(1+(1-4\beta)\zeta_2)(\ln\kappa-\ln a)}$$

which means the algorithm solves problem (1) to any arbitrary accuracy according to Definition 1.

From Theorems 1 and 2, if the steady-state value of $\delta(t)$ is a strictly positive constant, it does not need to be a monotonically decreasing function, hence providing extra freedom in adjusting the transient performance of the algorithm. However, the procedural knowledge for such design is still unknown and is left as a potential direction of future work.

VI. Design of Stochastic Process

In Section IV, the expected influences of the design parameters on the final performance have been discussed. Next, we provide a systematic design for the probability distribution for $\xi_i(t)$ in the proposed algorithm, based on the tradeoff between asymptotic optimization error and interevent interval, as analyzed in Theorems 1 and 2, assuming that the values for the scalar parameters have been decided.

Let μ and σ^2 be the expectation and variance of $\xi_i(t)$ for all i,t. The variance can be expressed as

$$\sigma^2=\theta(1-\mu)(\mu-a) \quad (13)$$

by the Bhatia–Davis inequality [29] for some $\theta\in[0,1]$.

We then propose the following design of $\xi_i(t)$, due to its ability to display a wide range of characteristics:

$$\begin{aligned}\xi_i(t)&\sim\text{Beta}(\alpha_\xi,\beta_\xi,a,1)\\ \alpha_\xi&=\frac{(1-a-\theta)(\mu-a)}{\theta(1-a)}\\ \beta_\xi&=\frac{(1-a-\theta)(1-\mu)}{\theta(1-a)}.\end{aligned} \quad (14)$$

The problem is to design μ and θ in (14), based on the needs of tradeoff between optimization error and communication rate, as formulated in the following design problem:

$$\begin{aligned}\min_{\mu,\theta}\quad & J(\mu,\theta)=\epsilon_e(\mu,\theta)+\bar{\psi}\tau_e^{-1}(\mu,\theta)\\ \text{s.t.}\quad & a\le\mu\le1\\ & 0\le\theta\le1\end{aligned} \quad (15)$$

where $\bar{\psi}\in[0,\infty)$ is the weighting factor, and

$$\epsilon_e(\mu,\theta)=P\left(\ln\kappa-\ln\mu+\frac{\theta(1-\mu)(\mu-a)}{2\mu^2}\right), \quad (16)$$

$$\tau_e(\mu,\theta)=Q\left(\ln\kappa-\ln\mu+\frac{\theta(1-\mu)(\mu-a)}{2\mu^2}\right) \quad (17)$$

$$P=\frac{2\mathcal{L}_{\max}\delta_\infty(1+(1-4\beta)\zeta_2)}{\zeta_1 m_{\min}(1-4\beta)} \quad (18)$$

$$Q=\frac{\sqrt{-W(-2e^{-2})(W(-2e^{-2})+2)\delta_\infty}}{U\sqrt{\ln\kappa-\ln a}}. \quad (19)$$

$\epsilon_e(\mu,\theta)$ and $\tau_e(\mu,\theta)$ from (16) and (17) are proportional to the upper and lower bounds for the asymptotic error and interevent intervals, as outlined in Proposition 1.

Proposition 1: Let $\tau(t)=\min_i\tau_i(t)$. If the lower bound of the stochastic processes $\xi_i(t)$ is strictly positive, i.e., $a>0$

$$\lim_{t\to\infty}\mathbb{E}\left[\varepsilon(t)\right]\le\epsilon_e(\mu,\theta)+\text{HOT}$$

$$\mathbb{E}\left[\tau(t)\right]\ge\tau_e(\mu,\theta)+\text{HOT}\quad\forall t>0$$

where HOT represents higher order terms in Taylor series.

Proof: See Appendix D. ∎

Remark 4: The higher order terms in Proposition 1 are neglected in (15) because they vanish to 0 in factorial order. The exact expression thereof can be found in Appendix D.

Furthermore, we provide the following normalization of the weighting factor, mapping from $\bar{\psi}\in[0,\infty)$ to $\psi\in[0,1]$

$$\bar{\psi}=PQ\left(\ln\kappa+\psi\left(\frac{a}{2\bar{\mu}^2}-\ln\bar{\mu}-\frac{3}{2}\right)\right)^2$$

where $\bar{\mu}=(-(1+a)+\sqrt{a^2+18a+1})/4$. This transformation allows for the use of a normalized weighting factor $\psi\in[0,1]$ and, more importantly, eliminates some global and possibly unknown parameters in the solutions. In the meantime,

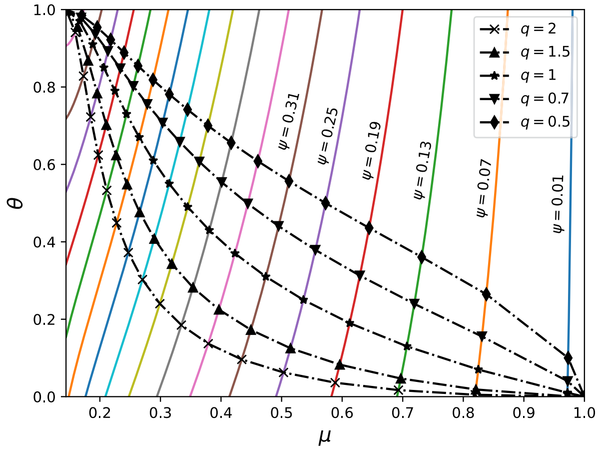


Fig. 3. Indifference curves in the proposed design (20).

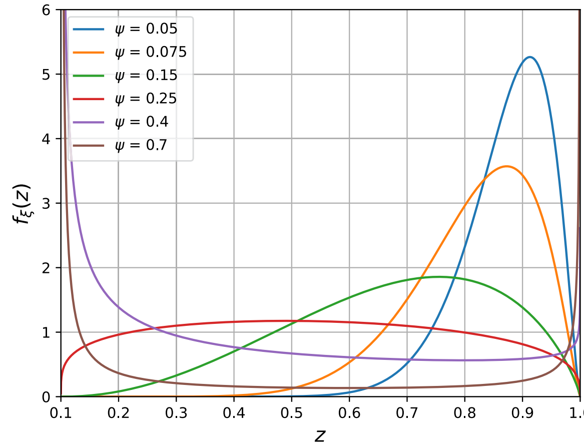


Fig. 4. Beta distributions for various ψ with $a = 0.1$.

it preserves the intuition that an increasing ψ, or equivalently $\bar{\psi}$, has larger emphasis on the communication rate. In addition, $\psi = 0$ yields the same solution as $\bar{\psi} = 0$ and similarly for $\psi = 1$ and $\bar{\psi} \to \infty$. The analysis and verification for this argument, along with the formulation for the normalization, will become clear in the proof of Theorem 3.

Theorem 3: The optimal solution to the design problem (15) is the following indifference curve for $\psi \in (0, 1)$:

$$\frac{\theta(1-\mu)(\mu-a)}{2\mu^2} - \ln \mu = \psi \left(\frac{a}{2\bar{\mu}^2} - \ln \bar{\mu} - \frac{3}{2} \right). \quad (20)$$

Moreover, if $\psi \in \{0, 1\}$, $\mu = 1 + \psi(\bar{\mu} - 1)$ and $\theta = \psi$.

Proof: See Appendix E. ■

Theorem 3 implies that there exist infinitely many choices for $\psi \notin \{0, 1\}$. The following proposition provided a method to obtain a unique solution.

Proposition 2: The selection $\theta = \psi^q$ is valid for all $\psi \in [0, 1]$ and $q > 0$ if $a < 1/3$.

Proof: For $\psi \in \{0, 1\}$, the statement has already been proven in Theorem 3. We will, therefore, only consider $\psi \in (0, 1)$ in this proof. We can rewrite (20) into $\mu = \mathcal{C}(\mu)$, where

$$\mathcal{C}(\mu) = \exp\left(\frac{\psi^q(1-\mu)(\mu-a)}{2\mu^2} - \psi \left(\frac{a}{2\bar{\mu}^2} - \ln \bar{\mu} - \frac{3}{2} \right) \right).$$

Let $\mathcal{C}'(\mu) = \exp(\psi^q(1-\mu)(\mu-a)/2\mu^2)$ with the domain $\mu \in [a, 1]$. It can be verified that $\mathcal{C}'(\mu) \geq 1$ and $\mathcal{C}'(a) = \mathcal{C}'(1) = 1$. The equation $\mu = \mathcal{C}'(\mu)$ has a unique solution at $\mu = 1$ regardless of ψ and q. Then, we have $0 < \mathcal{C}(1) < 1$. If $\mathcal{C}(a) \in (a, 1)$, $\mu = C(\mu)$ must have a solution by the intermediate value theorem as $C(\mu)$ is continuous, i.e.,

$$a < \exp\left(-\psi \left(\frac{a}{2\bar{\mu}^2} - \ln \bar{\mu} - \frac{3}{2} \right) \right) < 1. \quad (21)$$

It is important to note that $\bar{\mu}$ is dependent on a, and an analytical solution to (21) may not exist. Solving this inequality numerically results in a sufficient condition of $a < 1/3$. ■

The selection principle of Proposition 2 is illustrated in Fig. 3, where each colored curve represents an indifference curve for ψ from 0.01 to 0.97 with an interval of 0.06 as labeled. Any choice of $\theta(\psi)$ that crosses all curves, i.e., has a feasible pair of (θ, μ) for all $\psi \in [0, 1]$, is valid.

Remark 5: It should be noted that any distributions satisfying (20) is feasible for implementation. The main reason for choosing beta distribution is to allow for a systematic design process given any (μ, θ), whereas it is difficult to do so with other distributions. The beta distribution is also capable of displaying a wide range of characteristics, as shown in Fig. 4.

A summary of the stochastic process design is as follows, assuming that the scalar parameter a has been chosen.

1) Select the tradeoff parameter $\psi \in [0, 1]$, with larger values for communication reduction.
2) Select $\theta = \psi^q$ for any $q > 0$, e.g., $\theta = \psi$, and find the corresponding μ according to Theorem 3.
3) Obtain the beta distribution parameters α_ξ and β_ξ by (14) and $\xi_i(t) \sim \text{Beta}(\alpha_\xi, \beta_\xi, a, 1)$.

VII. Numerical Simulation

In this section, we illustrate the effectiveness of the proposed event-triggering law (8) with the controller (7) in comparison with the deterministic counterpart (10) along with the state-of-the-art triggering laws in [8], [22], [24], and [25]. In particular, we solve the distributed SVM classification problem, where each node has a set of private data $\mathcal{Z}_i \subset \mathbb{R}^2 \times \{-1, 1\}$ with arbitrary physical units, and the entire network should collectively compute the optimal parameters for the SVM classifier. Each data pair $(z, y) \in \mathcal{Z}_i$ consists of a data point $z \in \mathbb{R}^2$ and a label $y \in \{-1, 1\}$ defining the class to which z belongs.

The targeted SVM classifier can be expressed as $I(z) = w^T\varphi(z) - b$ whose value indicates which class the data point z should belong to. The parameters $w \in \mathbb{R}^3$ and $b \in \mathbb{R}$ are the weight and bias, respectively, to be learned by each node from the training set, while $\varphi : \mathbb{R}^2 \mapsto \mathbb{R}^3$ is a custom nonlinear feature mapping for the data, which is useful when the data are not linearly separable, defined as

$$\varphi\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right) = \begin{bmatrix} z_1 \\ z_2 \\ z_1^2 + z_2 \end{bmatrix}.$$

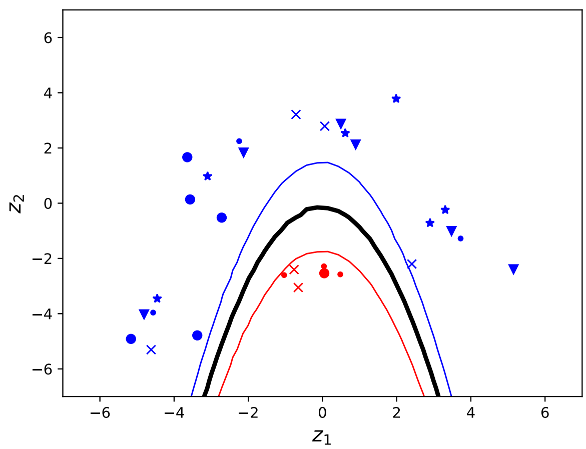


Fig. 5. Decision boundaries of the distributed SVM.

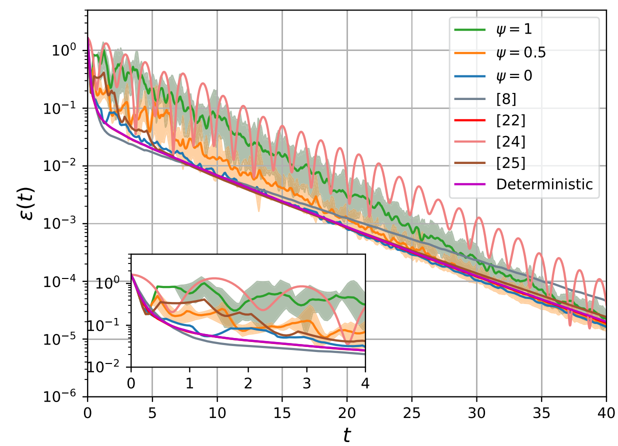


Fig. 6. Optimization errors $\varepsilon(t)$ in semilog scale.

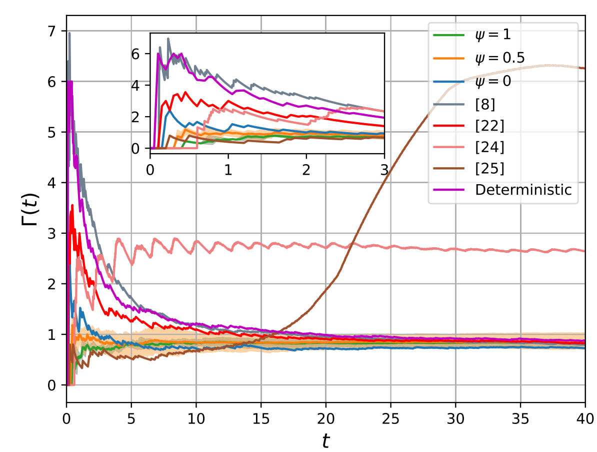


Fig. 7. Average communication rates $\Gamma(t)$.

For $(z, y) \in \mathcal{Z}_i$, $\forall i \in \mathcal{V}$, it is necessary that $y(w^T\varphi(z) - b) > 0$ in order for the resultant SVM to at least correctly classify all data in the training set. More specifically, $I(z) = -1$ and $I(z) = 1$ represent the decision boundaries for classes -1 and 1, respectively. The local objective functions for node i, with the state being $x_i = [w_i^T, b_i]^T \in \mathbb{R}^4$, are as follows:

$$f_i(x) = \sum_{(y,z)\in\mathcal{Z}_i} h(1 - y(w^T\varphi(z) - b)) + \frac{|\mathcal{Z}_i|}{N}\left(\|w\|^2 + b^2\right)$$

where $h(x) = \ln(1 + e^x)$ is the softplus function, an approximation to $\max(0, x)$. Note that the objective function above is slightly modified from the conventional formulation of the SVM, with the inclusion of the softplus function and regularization for the bias b. This is to ensure the validity of assumptions that each local objective function is strongly convex and has a Lipschitz continuous gradient.

We run the simulation with the design parameters $a = 0.05$, $\kappa = 1.05$, and $\beta = 0.1$. The distributions of the random processes $\xi_i(t)$ follow Theorem 3 and Proposition 2 with $q = 0.75$ and $\psi \in \{0, 0.5, 1\}$ for three different cases. In addition, we let $\delta(t) = 2e^{-0.3t} + 10^{-8}$, which satisfies the condition for the nonexistence of Zeno behavior and the assumption for the optimal trigger design. For all stochastic event-triggering laws, the simulation was run for 60 times to compute the empirical mean and max–min range of the metrics.

In addition to the optimization error $\varepsilon(t)$, we define another evaluation metric, namely the average communication rate, as

$$\Gamma(t) = \frac{1}{Nt}\sum_{i=1}^{N}\int_0^t \gamma_i(t)\, d\tau \quad \forall t > 0$$

with $\Gamma(0) = 0$. This definition is quantifying the total number of broadcast in the network per compute node per unit time.

Fig. 5 shows the decision boundaries resulting from the proposed event-triggered algorithm at $t = 40$, where blue and red represent $I(z) = -1$ and $I(z) = 1$, respectively. The private data from each node are plotted with a unique marker. The region beyond the blue boundary (away from the black boundary) is supposedly certain that belongs to class -1 and similarly for the red boundary. The black boundary is where the classifier cannot distinguish which class the data point belongs to, i.e., $I(z) = 0$.

The global optimization error $\varepsilon(t)$, or the empirical mean thereof, is plotted in Fig. 6 for each event-triggered algorithm considered. The shades represent the range between maximum and minimum values for those with stochasticity. While the proposed algorithms have slightly lower convergence rate in $t \in [0, 2]$, all algorithms have mostly comparable convergence rate for $t \in [2, 40]$ on average.

The main advantage of the proposed stochastic event-triggering law lies in an effective reduction of communication rate $\Gamma(t)$, particularly the maximum communication rate that determines the required bandwidth for the system, as demonstrated in Fig. 7 and Table II. The percentage reductions in $\max_t \mathbb{E}[\Gamma(t)]$ for the proposed algorithm with $\psi = 1$ compared with each algorithm are shown in the last column of the table. The proposed algorithm with the performance tradeoff design achieved the lowest peak average communication rate for $\psi = 1$, followed by $\psi = 0.5$ and 0, respectively, while the deterministic counterpart showed the highest value of all. The proposed algorithm showed a reduction of up to 87.7% in $\max_t \mathbb{E}[\Gamma(t)]$, meaning that it

TABLE II
COMPARISON OF ALGORITHMS IN $\Gamma(t)$

Algorithm	$\max_t \mathbb{E}[\Gamma(t)]$	% reduction from
$\psi = 1$	0.856	–
$\psi = 0.5$	1.185	-27.8%
$\psi = 0$	2.400	-64.3%
[8]	6.957	-87.7%
[22]	3.556	-75.9%
[24]	2.898	-70.5%
[25]	6.321	-86.5%
Deterministic	6.000	-85.7%

requires significantly lower bandwidth in physical hardware implementation.

VIII. Conclusion

In this article, we considered the distributed optimization problem with the aim to balance between optimization accuracy and communication rate. We proposed a stochastic event-triggering algorithm to significantly reduce the communication rate with the guarantee of arbitrarily small residual optimization error. It is also possible to achieve asymptotic convergence to zero optimization error at the expense of increased communication rate. We also proved that the proposed event-triggering algorithm does not exhibit Zeno behavior. We then derived a systematic design of the stochastic process with a meta-optimization problem based on the guarantee of optimization accuracy and interevent interval.

Potential future directions include the presence of malicious nodes, the extension of the algorithm to more general assumptions and models, and the generalization of the results with various triggering functions. The extension to other models and problems is of particularly interest because this work and [26], [28] focus solely on multiagent consensus and distributed optimization, while other problems involving distributed control or computation, such as Nash equilibrium seeking, could potentially benefit from the proposed stochastic event-triggering algorithm.

Appendix A
Proof of Theorem 3

Consider the Lyapunov candidate

$$V(x(t)) = \sum_{i=1}^{N} \left(f_i(x^\star) - f_i(x_i(t)) - \nabla f_i(x_i(t))^T (x^\star - x_i(t))\right). \tag{22}$$

The shortened notation $V(t)$ will be used for the remainder of this proof. Consider an arbitrary sample path of $\xi_i(t)$, thus $x_i(t), \hat{x}_i(t), \gamma_i(t), V(t), \varepsilon(t)$ by implication. Hence, these variables are no longer stochastic in the following proof. It should be noted that $x_i(t)$ and $\dot{x}_i(t)$ are not necessarily Lipschitz continuous due to the event-triggering law. However, $x_i(t)$ is still continuous and differentiable, while $\dot{x}_i(t)$ is Riemann integrable for any sample paths. Therefore, the time derivative $\dot{V}(t)$ is well defined as follows:

$$\begin{aligned}
&\dot{V}(t) \\
&\leq \sum_{i=1}^{N} (x_i(t) - x^\star)^T \nabla^2 f_i(x_i(t)) \dot{x}_i(t) \\
&= -\alpha \sum_{i=1}^{N} \sum_{j=1}^{N} L_{ij} x_i(t)^T \hat{x}_j(t) \\
&= -\alpha \sum_{i=1}^{N} \sum_{j=1}^{N} L_{ij} \left(\hat{x}_i(t) - e_i(t)\right)^T \hat{x}_j(t) \\
&= \alpha \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} L_{ij} \Bigg(\frac{1}{2} \|\hat{x}_j(t) - \hat{x}_i(t)\|^2 \\
&\quad + e_i(t)^T (\hat{x}_j(t) - \hat{x}_i(t))\Bigg) \\
&\leq \frac{\alpha(\nu-1)}{2\nu} \sum_{i=1}^{N} \sum_{j=1}^{N} L_{ij} \|\hat{x}_j(t) - \hat{x}_i(t)\|^2 \\
&\quad + \frac{\alpha\nu}{2} \sum_{i=1}^{N} L_{ii} \|e_i(t)\|^2 \\
&\leq \frac{\alpha}{2\nu} \left(\beta\nu^2 + \nu - 1\right) \sum_{i=1}^{N} \sum_{j=1}^{N} L_{ij} \|\hat{x}_j(t) - \hat{x}_i(t)\|^2 \\
&\quad + \frac{\alpha\nu}{2} \delta(t) \sum_{i=1}^{N} (\ln\kappa - \ln\xi_i(t)) \\
&\leq -\frac{\alpha(-\beta\nu^2 + \nu - 1)}{\nu} \hat{x}(t)^T (L \otimes I_n) \hat{x}(t) \\
&\quad + \frac{N\alpha\nu}{2} \delta(t) \left(\ln\kappa - \ln a\right).
\end{aligned}$$

To ensure exponential convergence, the coefficient of the first term needs to be strictly positive. To this end, we restrict $\beta = c(\nu-1)/\nu^2 < 1/4$ for some $c \in (0,1)$ and $\nu > 1$. Let $\bar{x}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t)$. From Lemma 2 and the analysis above, we have

$$\begin{aligned}
\dot{V}(t) &\leq -\frac{\alpha\zeta_1(1-c)(\nu-1)}{\nu} x(t)^T (K_N \otimes I_N) x(t) \\
&\quad + \frac{N\alpha}{2\nu} (\nu^2 + 4(1-c)\zeta_2\nu - 4 \\
&\quad \times (1-c)\zeta_2)\delta(t)(\ln\kappa - \ln a) \\
&= \frac{-\alpha\zeta_1(1-c)(\nu-1)}{\nu} \sum_{i=1}^{N} \|x_i(t) - \bar{x}(t)\|^2 \\
&\quad + \frac{N\alpha}{2\nu} (\nu^2 + 4(1-c)\zeta_2\nu \\
&\quad - 4(1-c)\zeta_2)\delta(t)(\ln\kappa - \ln a).
\end{aligned}$$

Following the analysis from [22], inasmuch as $x^\star$ is the global optimal solution, we have $\sum_{i=1}^N f_i(x^\star) = f(x^\star) \leq \sum_{i=1}^N f_i(y)$ for any y. Moreover, we have $\sum_{i=1}^N \nabla f_i(x_i(t)) = 0$, and consequently, $\sum_{i=1}^N \nabla f_i(x_i(t))^T(x^\star - x_i(t)) = \sum_{i=1}^N \nabla f_i(x_i(t))^T(y - x_i(t))$ for any constant vector y. Recalling the choice of the Lyapunov candidate (22) and the analysis above, we have

$$\begin{aligned} &V(x(t)) \\ &\leq \sum_{i=1}^N \left(f_i(\bar{x}(t)) - f_i(x_i(t)) - \nabla f_i(x_i(t))^T(\bar{x}(t) - x_i(t))\right) \\ &\leq \sum_{i=1}^N \frac{\mathcal{L}_i}{2}\|x_i(t) - \bar{x}(t)\|^2 \leq \frac{\mathcal{L}_{\max}}{2}\sum_{i=1}^n \|x_i(t) - \bar{x}(t)\|^2 \end{aligned}$$

where $\mathcal{L}_{\max} = \max_i \mathcal{L}_i$. Therefore

$$\dot{V}(t) \leq -\phi(\nu)V(t) + \omega(\nu)\delta(t) \tag{23}$$

with $\phi(\nu) = 2\alpha\zeta_1(1-c)(\nu-1)/\nu\mathcal{L}_{\max}$ and

$$\omega(\nu) = \frac{N\alpha}{2\nu}(\nu^2 + 4(1-c)\zeta_2\nu - 4(1-c)\zeta_2)(\ln\kappa - \ln a).$$

Solving the differential inequality (23) yields

$$V(t) \leq V(0)e^{-\phi(\nu)t} + \omega(\nu)e^{-\phi(\nu)t}\int_0^t e^{\phi(\nu)\tau}\delta(\tau)\,d\tau \tag{24}$$

$$\leq \left(V(0) - \frac{\omega(\nu)\delta_{\max}}{\phi(\nu)}\right)e^{-\phi(\nu)t} + \frac{\omega(\nu)\delta_{\max}}{\phi(\nu)}. \tag{25}$$

Equation (25) shows that $V(t)$ is bounded for $t \geq 0$. From the definition of $\varepsilon(t)$, we have $\varepsilon(t) \leq 2V(t)/Nm_{\min}$ and

$$\begin{aligned} \varepsilon(t) &\leq \min_{\nu>1}\left\{\frac{2}{Nm_{\min}}\left(V(0) - \frac{\omega(\nu)\delta_{\max}}{\phi(\nu)}\right)e^{-\phi(\nu)t} + \frac{2\delta_{\max}\omega(\nu)}{Nm_{\min}\phi(\nu)}\right\} \\ &\leq \frac{2}{Nm_{\min}}\left[\left(V(0) - \frac{\omega^\star\delta_{\max}}{\phi^\star}\right)e^{-\phi^\star t} + \frac{\omega^\star\delta_{\max}}{\phi^\star}\right] \end{aligned}$$

which concludes the proof.

Appendix B
Proof of Theorem 1

The outline of the proof is as follows. For any node $i \in \mathcal{V}$ in $t \in [t_k^i, t_{k+1}^i)$, an upper bound, E_i^+, is first derived for $\|e_i(t)\|^2$. In addition, the proposed algorithm (8) ensures a lower bound E_i^- if the node i is triggered at t_{k+1}^i. Then, we showed that $t_{k+1}^i - t_k^i$ has a strictly positive lower bound from $E_i^+ > E_i^-$, which shows the nonexistence of Zeno behavior.

From the definition of the computation law (7), one can derive a useful bound for $\|u_i(t)\|$ as follows:

$$\begin{aligned} \|u_i(t)\| &\leq \frac{2\alpha}{m_i}\left\|\sum_{j=1}^N a_{ij}(\hat{x}_j(t) - \hat{x}_i(t))\right\| \\ &\leq \frac{2\alpha L_{ii}}{m_i}\max_{i,j\in\mathcal{V}}\|\hat{x}_j(t) - \hat{x}_i(t)\|. \end{aligned} \tag{26}$$

By the triangular inequality

$$\begin{aligned} \|\hat{x}_j(t) - \hat{x}_i(t)\| &= \|x_j(t) - x_i(t) + e_j(t) - e_i(t)\| \\ &\leq \|x_j(t) - x_i(t)\| + \|e_j(t) - e_i(t)\| \\ &\leq \|x_j(t) - x_i(t)\| + 2\max_{i\in\mathcal{V}}\|e_i(t)\|. \end{aligned} \tag{27}$$

From (8) and (9), the SET algorithm guarantees that

$$\|e_i(t)\|^2 \leq -\frac{\beta}{L_{ii}}\sum_{j=1}^N L_{ij}\|\hat{x}_j(t) - \hat{x}_i(t)\|^2 + \frac{\delta(t)}{L_{ii}}(\ln\kappa - \ln a) \tag{28}$$

$$\leq \beta\max_{i,j\in\mathcal{V}}\|\hat{x}_j(t) - \hat{x}_i(t)\|^2 + \frac{\delta(t)}{L_{\min}}(\ln\kappa - \ln a). \tag{29}$$

The remainder of the proof is divided into two cases, as described in Theorem 1.

Case 1: $\delta_\infty > 0$. From (29), have

$$\|e_i(t)\| \leq \sqrt{\beta}\max_{i,j\in\mathcal{V}}\|\hat{x}_j(t) - \hat{x}_i(t)\| + \sqrt{\frac{\delta_{\max}(\ln\kappa - \ln a)}{L_{\min}}}. \tag{30}$$

Combining (30) with (27), we obtain

$$\begin{aligned} &(1 - 2\sqrt{\beta})\max_{i,j\in\mathcal{V}}\|\hat{x}_j(t) - \hat{x}_i(t)\| \\ &\leq \max_{i,j\in\mathcal{V}}\|x_j(t) - x_i(t)\| + \sqrt{\frac{4\delta_{\max}(\ln\kappa - \ln a)}{L_{\min}}}. \end{aligned} \tag{31}$$

Notice that

$$\begin{aligned} \|x_i(t) - x_j(t)\| &\leq \|x_i(t) - x^\star\| + \|x_j(t) - x^\star\| \\ &\leq 2\sqrt{N\varepsilon(t)}. \end{aligned} \tag{32}$$

Then, substituting (32) back into (31) yields

$$\begin{aligned} \|u_i(t)\| &\leq \frac{4\alpha L_{\max}}{m_{\min}(1 - 2\sqrt{\beta})} \times \left(\sqrt{N\varepsilon(t)} + \sqrt{\frac{\delta_{\max}(\ln\kappa - \ln a)}{L_{\min}}}\right) \\ &\leq \frac{4\alpha L_{\max}}{m_{\min}(1 - 2\sqrt{\beta})}\left(\sqrt{\frac{2}{m_{\min}}\max\left(V_0, \frac{\omega^\star\delta_{\max}}{\phi^\star}\right)} + \sqrt{\frac{\delta_{\max}(\ln\kappa - \ln a)}{L_{\min}}}\right) = U. \end{aligned}$$

For $t \in [t_k^i, t_{k+1}^i)$ and any arbitrary constant $\varsigma > 0$, we have

$$\begin{aligned} \frac{d}{dt}\|e_i(t)\|^2 &= 2(\hat{x}_i(t) - x_i(t))^T\left(\dot{\hat{x}}_i(t) - \dot{x}_i(t)\right) \\ &= -2e_i(t)u_i(t) \\ &\leq \varsigma\|e_i(t)\|^2 + \frac{1}{\varsigma}U^2. \end{aligned}$$

Solving the above ordinary differential inequality results in

$$\|e_i(t)\|^2 \le \left\|e_i(t_k^i)\right\|^2 e^{\varsigma(t-t_k^i)} + \frac{U^2}{\varsigma} e^{\varsigma(t-t_k^i)} \int_{t_k^i}^{t} e^{\varsigma(t_k^i-\tau)}\, d\tau = \left(\frac{U}{\varsigma}\right)^2 \left(e^{\varsigma(t-t_k^i)} - 1\right). \quad (33)$$

In order for the node i to trigger broadcasting at t_{k+1}^i, a necessary condition is

$$\left\|e_i(t_{k+1}^{i-})\right\|^2 > -\frac{\beta}{L_{ii}} \sum_{i=1}^{N} L_{ij} \left\|\hat{x}_j(t_{k+1}^{i-}) - \hat{x}_i(t_{k+1}^{i-})\right\|^2 + \frac{\delta(t_{k+1}^{i-})}{L_{ii}} \left(\ln\kappa - \ln\xi_i(t_{k+1}^{i-})\right). \quad (34)$$

Recall that $\|e_i(t)\|^2$ has an upper bound (33) when the node is not at triggering instance. Combining (33) and (34) and solving the inequality thereof leads to

$$t_{k+1}^i - t_k^i > \frac{1}{\varsigma} \ln\left(1 + \frac{\varsigma^2 \delta_\infty (\ln\kappa - \ln\xi_i(t_{k+1}^{i-}))}{U^2}\right). \quad (35)$$

Since the above inequality holds for any $\varsigma > 0$ and $k > 0$, we can choose the maximum of the right-hand side over ς, i.e.,

$$\tau_i(t) \ge \max_{\varsigma>0} \frac{1}{\varsigma} \ln\left(1 + \frac{\varsigma^2 \delta_\infty \ln\kappa}{U^2}\right) = \sqrt{-\frac{\delta_\infty \ln\kappa}{U^2} W\left(-\frac{2}{e^2}\right)\left(W\left(-\frac{2}{e^2}\right)+2\right)} > 0. \quad (36)$$

Case 2a: $\delta(t) = \delta_0 e^{-\eta t}$ for some $\delta_0 > 0$ and $\eta \in [0, \phi)$. From the definition of $e_i(t)$, one can find that $\dot{e}_i(t) = -\dot{x}_i(t) = -u_i(t)$; therefore, $\|e_i(t)\| \le \int_{t_k^i}^t \|u_i(t)\|\, dt$. Following (29), we have

$$\|e_i(t)\| \le \sqrt{\beta} \max_{i,j\in\mathcal{V}} \|\hat{x}_j(t) - \hat{x}_i(t)\| + \sqrt{\frac{\delta_0(\ln\kappa - \ln a)e^{-\eta t}}{L_{\min}}}. \quad (37)$$

Similar to case 1, we have

$$(1 - 2\sqrt{\beta}) \max_{i,j\in\mathcal{V}} \|\hat{x}_j(t) - \hat{x}_i(t)\| \le \max_{i,j\in\mathcal{V}} \|x_j(t) - x_i(t)\| + \sqrt{\frac{4\delta_0(\ln\kappa - \ln a)}{L_{\min}}} e^{-\eta t/2}. \quad (38)$$

Putting (24), (26)–(32), and (38) together, for $t \in [t_k^i, t_{k+1}^i)$, we have

$$\|u_i(t)\| \le \varsigma_1 e^{-\phi^\star t/2} + \varsigma_2 e^{-\eta t/2} \le \varsigma_1 e^{-\phi^\star t_k^i/2} + \varsigma_2 e^{-\eta t_k^i/2}$$

since $\varsigma_1, \varsigma_2, \phi, \eta > 0$. One can then obtain

$$\|e_i(t)\| \le \left(\varsigma_1 e^{-\phi^\star t_k^i/2} + \varsigma_2 e^{-\eta t_k^i/2}\right)(t - t_k^i).$$

A necessary condition for triggering is

$$\left\|e_i(t_{k+1}^{i-})\right\| > \sqrt{\frac{(\delta_0 \ln\kappa)e^{-\eta t_{k+1}^i}}{L_{\max}}} = \varsigma_3\, e^{-\eta t_{k+1}^i/2}$$

$$t_{k+1}^i - t_k^i > \frac{\varsigma_3 \exp(-\eta t_{k+1}^i/2)}{\varsigma_1 \exp(-\phi^\star t_k^i/2) + \varsigma_2 \exp(-\eta t_k^i/2)} = \frac{\varsigma_3 \exp(-\eta(t_{k+1}^i - t_k^i)/2)}{\varsigma_1 \exp(-(\phi^\star - \eta)t_k^i/2) + \varsigma_2} \ge \frac{\varsigma_3}{\varsigma_1 + \varsigma_2} \exp\left(-\frac{\eta(t_{k+1}^i - t_k^i)}{2}\right).$$

where $\varsigma_3 = \sqrt{\delta_0 \ln\kappa / L_{\max}}$. Solving the last inequality yields

$$\tau_i(t) \ge t_{k+1}^i - t_k^i > \frac{2}{\eta} W\left(\frac{\eta\varsigma_3}{2(\varsigma_1 + \varsigma_2)}\right) > 0.$$

Case 2b: $\delta(t) > \delta_0 e^{-\eta t}$ for some $\delta_0 > 0$ and $\eta \in [0, \phi)$. Since a larger $\delta(t)$ implies a higher threshold to trigger, provided identical states for all other variables, the lower bound found in case 2a is also applicable in this case.

Appendix C
Proof of Theorem 2

Let $h(t) = e^{-\phi(\nu)t}$, $g(t) = \omega(\nu)\delta(t)$ with domain $t \in [0, \infty)$ and $h(t) = g(t) = 0$ for $t < 0$, $H(s), G(s)$ be the Laplace transform of $h(t)$ and $g(t)$, respectively. It is a well-known result that $H(s) = \frac{1}{s+\phi(\nu)}$. Moreover

$$\lim_{t\to\infty} e^{-\phi(\nu)t} \int_0^t e^{\phi(\nu)\tau} g(\tau)\, d\tau = \lim_{t\to\infty} \int_{-t}^t h(\tau - t) g(\tau)\, d\tau = \lim_{t\to\infty} h(t) * g(t) \quad (39)$$

where $*$ represents convolution. By the final value theorem, we have

$$\lim_{t\to\infty} h(t) * g(t) = \lim_{s\to 0} \frac{sG(s)}{s + \phi(\nu)} = \frac{1}{\phi(\nu)} \lim_{t\to\infty} g(t). \quad (40)$$

Recall that for any realizations of $\xi_i(t)$, hence $V(t)$

$$V(t) \le V(0)e^{-\phi(\nu)t} + \omega(\nu)e^{-\phi(\nu)t} \int_0^t e^{\phi(\nu)\tau} \delta(\tau)\, d\tau = V(0)e^{-\phi(\nu)t} + \int_0^t h(\tau - t) g(\tau)\, d\tau. \quad (41)$$

Substituting the result from (40) into (41) leads to

$$\lim_{t\to\infty} V(t) \le \lim_{t\to\infty} h(t) * g(t) = \frac{1}{\phi(\nu)} \lim_{t\to\infty} g(t) = \frac{\omega(\nu)\delta_\infty}{\phi(\nu)}.$$

From the definition of the Lyapunov candidate, we also have the inequality

$$V(t) \ge \sum_{i=1}^N \frac{m_i}{2} \|x_i - x^\star\|^2 \ge \frac{m_{\min}}{2} \sum_{i=1}^N \|x_i - x^\star\|^2. \quad (42)$$

Following (24) from the proof of Theorem 3 and (42), we have

$$\lim_{t\to\infty} \varepsilon(t) \le \lim_{t\to\infty} \frac{2}{N m_{\min}} V(t) \quad (43)$$

$$\le \min_{\nu>1} \frac{2\omega(\nu)\delta_\infty}{N\phi(\nu)m_{\min}} \quad (44)$$

$$= \frac{2\mathcal{L}_{\max}\delta_\infty(1 + (1-4\beta)\zeta_2)(\ln\kappa - \ln a)}{\zeta_1 m_{\min}(1 - 4\beta)}. \quad (45)$$

The equality is obtained by the straightforward minimization of the rational function in (45) with respect to ν. Inasmuch as the Lyapunov function is bounded for all t, its expectation exists and

$$\frac{d}{dt}\mathbb{E}\left[V(t)\right] = \mathbb{E}\left[\frac{d}{dt}V(t)\right].$$

Then, one can follow a similar procedure as in (22), (23), and (39)–(45) to obtain

$$\lim_{t\to\infty}\mathbb{E}\left[\varepsilon(t)\right] \le \frac{2\mathcal{L}_{\max}\delta_\infty(1+(1-4\beta)\zeta_2)(\ln\kappa - \mathbb{E}\left[\ln\xi_i(t)\right])}{\zeta_1 m_{\min}(1-4\beta)}.$$

Appendix D
Proof of Proposition 1

By Taylor series expansion around $\mu = \mathbb{E}[\xi_i(t)]$, the quantity $\mu_{\ln} = \mathbb{E}[\ln\xi_i(t)]$ can be written as

$$\mu_{\ln} = \ln\mu - \frac{\sigma^2}{2\mu^2} + \underbrace{\sum_{j=3}^{\infty}\frac{(-1)^{j-1}\mathbb{E}\left[(\xi_i(t)-\mu)^j\right]}{j!\mu^j}}_{\text{Higher Order Terms (HOT)}}.$$

Let μ_k and μ_k' be the kth central moment and moment for Beta(α_ξ, β_ξ), respectively; then, the HOT can be written as

$$\begin{aligned}\text{HOT} &= \sum_{j=3}^{\infty}\frac{(-1)^{j-1}(1-a)^j\mu_j}{j!\mu^j}\\ &= \sum_{j=3}^{\infty}\sum_{k=0}^{j}\frac{(-1)^{k+1}(1-a)^j C_k^j\mu_k'}{j!\mu^k}\\ &= \sum_{j=3}^{\infty}\sum_{k=0}^{j}\frac{(-1)^{k+1}(1-a)^j}{(j-k)!k!}\prod_{r=1}^{k}\frac{\alpha_\xi + r}{\alpha_\xi+\beta_\xi+r}\end{aligned}$$

which vanishes in factorial order. Combining with Theorem 2, we obtain

$$\begin{aligned}\epsilon_2 &\le \frac{2\mathcal{L}_{\max}\delta_\infty(1+(1-4\beta)\zeta_2)}{\zeta_1 m_{\min}(1-4\beta)}\left(\ln\kappa - \ln\mu + \frac{\sigma^2}{2\mu^2} + \text{HOT}\right)\\ &= P\left(\ln\kappa - \ln\mu + \frac{\theta(1-\mu)(\mu-a)}{2\mu^2} + \text{HOT}\right)\\ &= \epsilon_e(\mu,\theta) + \text{HOT}.\end{aligned}$$

Let $\mathcal{W} = -W(-2e^{-2})(W(-2e^{-2})+2)$. Following (35) in the proof in Theorem 1, for any $i\in\mathcal{V}$, $k\in\mathbb{N}$, we have

$$\begin{aligned}\mathbb{E}\left[t_{k+1}^i - t_k^i\right] &\ge \mathbb{E}\left[\sqrt{\frac{\mathcal{W}\delta_\infty(\ln\kappa - \ln\xi_i(t_{k+1}^-))}{U^2}}\right]\\ &\ge \mathbb{E}\left[\frac{\sqrt{\mathcal{W}\delta_\infty}(\ln\kappa - \ln\xi_i(t_{k+1}^-))}{U\sqrt{\ln\kappa - \ln a}}\right]\\ &\ge Q\left(\ln\kappa - \ln\mu + \frac{\theta(1-\mu)(\mu-a)}{2\mu^2} + \text{HOT}\right)\\ &= \tau_e(\mu,\theta) + \text{HOT}.\end{aligned}$$

From the definition of $\tau(t)$, we have $\tau(t) \ge \min_{i,k}\{t_{k+1}^i - t_k^i\}$; hence, $\mathbb{E}[\tau(t)] \ge \tau_e(\mu,\theta) + \text{HOT}$, which concludes the proof.

Appendix E
Proof of Theorem 3

The following proof consists of two major steps: unveiling the hidden convexity in the objective function $J(\mu,\theta)$ and remapping the weighting factor from $\bar{\psi}$ to ψ by exploiting the constraints to eliminate the unknown quantities P and Q.

Let $y(\mu,\theta) = \ln\kappa - \ln\mu + \theta(1-\mu)(\mu-a)/(2\mu^2)$. While $J(\mu,\theta)$ is nonconvex in μ, it is convex in y for $y>0$. We start by setting the derivative $\partial J/\partial y = 0$

$$\frac{\partial J}{\partial y} = P - \frac{\bar{\psi}}{Qy^2} = 0 \implies y^\star = \sqrt{\frac{\bar{\psi}}{PQ}}$$

where $y^\star$ is the optimal solution to $\min_y J(y)$. In other words, any pair of (μ,θ) satisfying

$$\ln\kappa - \ln\mu + \frac{\theta(1-\mu)(\mu-a)}{2\mu^2} = \sqrt{\frac{\bar{\psi}}{PQ}} \tag{46}$$

is an optimal solution to (15) if and only if the constraints are also satisfied. Now, we investigate the feasibility of the indifference curve. We first rewrite (46) into

$$\theta = \frac{2\mu^2}{(1-\mu)(\mu-a)}(y^\star - \ln\kappa + \ln\mu).$$

If θ is to be feasible, it is sufficient and necessary that

$$-\ln\mu \le y^\star - \ln\kappa \le \frac{(1-\mu)(\mu-a)}{2\mu^2} - \ln\mu, \quad \mu\in[a,1].$$

That is, $\exists\mu\in[a,1]$ such that the solution (46) is feasible if

$$\min_{\mu\in[a,1]} -\ln\mu \le y^\star - \ln\kappa \le \max_{\mu\in[a,1]}\frac{(1-\mu)(\mu-a)}{2\mu^2} - \ln\mu.$$

Since $-\ln\mu$ is monotone in $\mu\in[a,1]$, its minimum can be found by comparing the boundary value, which is at $\mu = 1 \Rightarrow -\ln\mu = 0$. For the rightmost side of the above inequality, first, we verify the existence of a unique critical point in $\mu\in[a,1]$

$$\frac{d}{d\mu}\left(\frac{(1-\mu)(\mu-a)}{2\mu^2} - \ln\mu\right) = -\frac{a(\mu-2)+\mu(2\mu+1)}{2\mu^3} = 0$$

$$\mu = \frac{-(1+a)+\sqrt{a^2+18a+1}}{4} = \bar{\mu}.$$

Then

$$\frac{d^2}{d\mu^2}\left(\frac{(1-\mu)(\mu-a)}{2\mu^2} - \ln\mu\right) = \frac{\mu^2+(1+a)\mu-3a}{\mu^4}$$

which is strictly negative at $\mu=\bar{\mu}$, implying concavity, hence a local maximum. To determine the global maximum, we compare with the endpoint $\mu = a$ ($\mu = 1$ is neglected as it yields 0 for both terms). With simple algebraic manipulations

$$\begin{aligned}\frac{(1-\bar{\mu})(\bar{\mu}-a)}{2\bar{\mu}^2} - \ln\bar{\mu} &= \frac{-\bar{\mu}^2+(1+a)\bar{\mu}-a}{2\bar{\mu}^2} - \ln\bar{\mu}\\ &= \frac{a}{2\bar{\mu}^2} - \ln\bar{\mu} - \frac{3}{2} \ge -\ln a\end{aligned}$$

for $0 < a < 1$. Therefore, the feasibility condition becomes

$$0 \leq y^\star - \ln\kappa \leq \frac{a}{2\bar{\mu}^2} - \ln\bar{\mu} - \frac{3}{2}$$

$$\ln\kappa \leq \sqrt{\frac{\bar{\psi}}{PQ}} \leq \frac{a}{2\bar{\mu}^2} - \ln\bar{\mu} + \ln\kappa - \frac{3}{2}$$

or we can restrict the range by the transformation on $\bar{\psi}$

$$\sqrt{\frac{\bar{\psi}}{PQ}} = (1-\psi)\ln\kappa + \psi\left(\frac{a}{2\bar{\mu}^2} - \ln\bar{\mu} + \ln\kappa - \frac{3}{2}\right)$$

$$\bar{\psi} = PQ\left(\ln\kappa + \psi\left(\frac{a}{2\bar{\mu}^2} - \ln\bar{\mu} - \frac{3}{2}\right)\right)^2. \quad (47)$$

Substituting (47) into (46) yields the intended result (20). It is trivial to graphically validate that $\epsilon_e(\mu,\theta)$ increases with ψ with the corresponding optimal solution, and conversely, $\tau_e^{-1}(\mu,\theta)$ decreases. This implies that ψ serves the identical weighting purpose as $\bar{\psi}$ but simply normalized. Therefore, (47) is an appropriate transformation.

One can also verify that $\psi \leq 0 \Rightarrow \partial J/\partial y \geq 0$, hence $J(\mu,\theta)$ being monotonically nondecreasing. On the other hand, we have $\partial J/\partial y \leq 0$ for $\psi \geq 1$. Under the constraints in (15), the optimal solutions when $\psi \notin (0,1)$ are then

$$(\mu,\theta) = \begin{cases} \arg\min_{\mu,\theta} y(\mu,\theta) = (1,0), & \psi \leq 0 \\ \arg\max_{\mu,\theta} y(\mu,\theta) = (\bar{\mu},1), & \psi \geq 1 \end{cases}$$

or, equivalently, $\mu = 1 - \psi + \psi\bar{\mu}$ and $\theta = \psi$ for $\psi \in \{0,1\}$.

References

[1] R. Olfati-Saber and R. M. Murray, "Consensus protocols for networks of dynamic agents," in *Proc. Amer. Control Conf.*, 2003, pp. 951–956.

[2] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.

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

[4] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[5] D. V. Dimarogonas and K. H. Johansson, "Event-triggered control for multi-agent systems," in *Proc. IEEE Conf. Decis. Control Held Jointly Chinese Control Conf.*, 2009, pp. 7131–7136.

[6] J. Lu and C. Y. Tang, "Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case," *IEEE Trans. Autom. Control*, vol. 57, no. 9, pp. 2348–2354, Sep. 2012.

[7] J. Wang and N. Elia, "Control approach to distributed optimization," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, 2010, pp. 557–561.

[8] S. S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, 2015.

[9] D. Shi, T. Chen, and L. Shi, "On set-valued Kalman filtering and its application to event-based state estimation," *IEEE Trans. Autom. Control*, vol. 60, no. 5, pp. 1275–1290, May 2015.

[10] A. Sundararajan, B. Van Scoy, and L. Lessard, "A canonical form for first-order distributed optimization algorithms," in *Proc. Amer. Control Conf.*, 2019, pp. 4075–4080.

[11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[12] T. Yang et al., "A survey of distributed optimization," *Annu. Rev. Control*, vol. 47, pp. 278–305, 2019.

[13] 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.

[14] J. Lei, H. F. Chen, and H. T. Fang, "Primal–dual algorithm for distributed constrained optimization," *Syst. Control Lett.*, vol. 96, pp. 110–117, 2016.

[15] G. S. Seyboth, D. V. Dimarogonas, and K. H. Johansson, "Event-based broadcasting for multi-agent average consensus," *Automatica*, vol. 49, no. 1, pp. 245–252, 2013.

[16] L. Ding, Q. L. Han, X. Ge, and X. M. Zhang, "An overview of recent advances in event-triggered consensus of multiagent systems," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1110–1123, Apr. 2018.

[17] X. Yi, K. Liu, D. V. Dimarogonas, and K. H. Johansson, "Distributed dynamic event-triggered control for multi-agent systems," in *Proc. IEEE Conf. Decis. Control*, 2017, pp. 6683–6698.

[18] W. H. Chen, J. Yang, L. Guo, and S. Li, "Disturbance-observer-based control and related methods—An overview," *IEEE Trans. Ind. Electron.*, vol. 63, no. 2, pp. 1083–1095, Feb. 2016.

[19] Z. Wu, Z. Li, Z. Ding, and Z. Li, "Distributed continuous-time optimization with scalable adaptive event-based mechanisms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 9, pp. 3252–3257, Sep. 2020.

[20] X. Shi, R. Zheng, Z. Lin, and G. Yan, "Consensus of first-order multi-agent systems under event-triggered communication," in *Proc. Chin. Control Decis. Conf.*, 2018, pp. 4679–4683.

[21] W. Xu, D. W. Ho, J. Zhong, and B. Chen, "Event/self-triggered control for leader-following consensus over unreliable network with DoS attacks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 3137–3149, Oct. 2019.

[22] W. Du, X. Yi, J. George, K. H. Johansson, and T. Yang, "Distributed optimization with dynamic event-triggered mechanisms," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 969–974.

[23] X. Yi, L. Yao, T. Yang, J. George, and K. H. Johansson, "Distributed optimization for second-order multi-agent systems with dynamic event-triggered communication," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 3397–3402.

[24] M. Li, L. Su, and T. Liu, "Distributed optimization with event-triggered communication via input feedforward passivity," *IEEE Control Syst. Lett.*, vol. 5, no. 1, pp. 283–288, Jan. 2021.

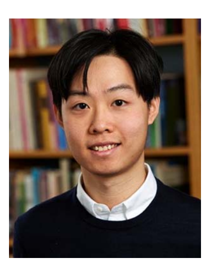
[25] W. Chen and W. Ren, "Event-triggered zero-gradient-sum distributed consensus optimization over directed networks," *Automatica*, vol. 65, pp. 90–97, 2016.

[26] K. F. E. Tsang, J. Wu, and L. Shi, "Zeno-free stochastic distributed event-triggered consensus control for multi-agent systems," in *Proc. Amer. Control Conf.*, 2019, pp. 778–783.

[27] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Trans. Autom. Control*, vol. 60, no. 10, pp. 2661–2675, Oct. 2015.

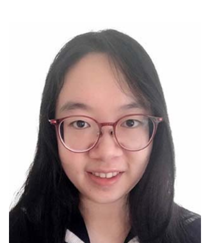
[28] K. F. E. Tsang, J. Wu, and L. Shi, "Distributed optimisation with stochastic event-triggered multi-agent control algorithm," in *Proc. IEEE Conf. Decis. Control*, 2020, pp. 6222–6227.

[29] R. Bhatia and C. Davis, "A better bound on the variance," *Amer. Math. Monthly*, vol. 107, no. 4, pp. 353–357, 2000.



Kam Fai Elvis Tsang received the B.Eng. degree in mechanical engineering and electronic engineering and the M.Phil. degree in electronic engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2018 and 2020, respectively. He is currently working toward the Ph.D. degree in electrical engineering with the KTH Royal Institute of Technology, Stockholm, Sweden.

His research interests include multiagent systems, distributed optimization, and machine learning.



Mengyu Huang received the B.Eng. degree in information engineering from Southeast University, Nanjing, China, in 2018. She is currently working toward the Ph.D. degree in electronic and computer engineering with the Hong Kong University of Science and Technology, Hong Kong.

Her research interests include reinforcement learning, machine learning, and cyber-physical systems security.

Ling Shi (Senior Member, IEEE) received the B.S. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2002, and the Ph.D. degree in control and dynamical systems from the California Institute of Technology, Pasadena, CA, USA, in 2008.

He is currently a Professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, where he is also the Associate Director of the Robotics Institute. His research interests include cyber-physical system security, networked control systems, sensor scheduling, event-based state estimation, and exoskeleton robots.

Dr. Shi was an Editorial Board Member for the European Control Conference from 2013 to 2016. He was a Subject Editor for *International Journal of Robust and Nonlinear Control* from 2015 to 2017. Since 2016, he has been an Associate Editor for IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS. Since 2017, he has also been an Associate Editor for IEEE CONTROL SYSTEMS LETTERS. He was an Associate Editor for a special issue on Secure Control of Cyber Physical Systems in IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS from 2015 to 2017. He was the General Chair of the 23rd International Symposium on Mathematical Theory of Networks and Systems in 2018. He was a Member of the Young Scientists Class 2020 of the World Economic Forum.

Karl Henrik Johansson (Fellow, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree in engineering (automatic control) from Lund University, Lund, Sweden, in 1992 and 1997, respectively.

He is currently the Director of Digital Futures and a Professor with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions with the University of California, Berkeley, CA, USA; California Institute of Technology, Pasadena, CA; Nanyang Technological University, Singapore; Institute of Advanced Studies, Hong Kong University of Science and Technology, Hong Kong; and Norwegian University of Science and Technology, Trondheim, Norway. His research interests include networked control systems, cyber-physical systems, and applications in transportation, energy, and automation.

Dr. Johansson is a Member of the Swedish Research Council's Scientific Council for Natural Sciences and Engineering Sciences. He has served on the IEEE Control Systems Society Board of Governors and the International Federation of Automatic Control (IFAC) Executive Board. He is a Vice-President of the European Control Association Council. He has received several best paper awards and other distinctions from the IEEE, the IFAC, and the Association for Computing Machinery. He has been awarded Distinguished Professor with the Swedish Research Council and Wallenberg Scholar with the Knut and Alice Wallenberg Foundation. He has received the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from the IFAC. He is Fellow of the Royal Swedish Academy of Engineering Sciences and a Distinguished Lecturer of the IEEE Control Systems Society.