



Title	Distributed Byzantine-Resilient Stochastic Optimization With Event-Triggered Communication
Author(s)	Tanaka, Shota; Hayashi, Naoki; Inuiguchi, Masahiro
Citation	Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics. 2026, p. 6494-6499
Version Type	AM
URL	https://hdl.handle.net/11094/104540
rights	©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Distributed Byzantine-Resilient Stochastic Optimization With Event-Triggered Communication

Shota Tanaka, Naoki Hayashi, and Masahiro Inuiguchi

Abstract—We consider Byzantine-resilient distributed optimization with event-triggered communication. The proposed algorithm is designed to handle non-convex optimization problems in a network of agents where some agents may exhibit Byzantine behavior. Each normal agent has an estimate of a critical point of the global cost function and transmits the estimate to neighbors when the difference between the current and previously communicated values exceeds a predefined threshold. Normal agents then update their estimates by averaging the received values from a trusted set that is determined through the Iterative Outlier Scissor (IOS) procedure. By combining the event-triggered communication and the IOS filtering procedure, the proposed approach ensures resilience to Byzantine behavior and guarantees convergence even in adversarial settings.

I. INTRODUCTION

Distributed optimization is a framework in which multiple agents collaboratively solve an optimization problem. In distributed optimization, each agent has access to only a subset of the total information, and the objective is to find an optimal solution that is collectively determined by all agents through iterative communication and computation [1]–[3].

In recent years, the increasing interconnectivity of agents in distributed systems has introduced significant vulnerabilities and a critical threat to the reliability and security of the networks [4]. These vulnerabilities are critical in the presence of Byzantine attacks, where malicious agents are capable of acting arbitrarily to disrupt the system. There have been developed various defensive approaches that utilize the aggregation mechanisms [5], [6]. These robust aggregation techniques incorporate filtering methods that identify and mitigate the influence of outlier information. Examples of such Byzantine-resilient aggregation rules include a geometric median [7] which selects a point minimizing the sum of distances to all received inputs, Krum [8] which identifies the vector with minimal sum of distances to its closest neighbors, and trimmed mean approaches [9]–[12] which exclude extreme values before computing an average.

However, most existing algorithms for Byzantine-resilient methods rely on a time-triggered communication scheme, where agents are required to exchange information with their neighbors at every iteration. The time-triggered scheme can be highly inefficient in resource-constrained environments, particularly in sensor networks where individual sensors have

limited battery capacity and restricted communication capabilities. To address this limitation, event-triggered communication has been investigated as an alternative approach that reduces unnecessary message exchanges while maintaining reliable system performance [13]–[16]. In the event-triggered scheme, agents transmit information to their neighbors only when a specific trigger condition is met. By dynamically adjusting communication frequency based on real-time system conditions, this approach significantly conserves network resources and extends battery life in sensor networks.

This paper proposes a distributed Byzantine-resilient algorithm for non-convex optimization that integrates event-triggered communication. Each normal agent has an estimate of a critical point of the global cost function. The local communication between agents is carried out in an event-triggered manner, where each normal agent transmits its estimate to neighboring agents only when the difference between the current value and the last transmitted value exceeds a predefined threshold. After the local communication, normal agents update their estimates by averaging the received values from a trusted set that is determined through the Iterative Outlier Scissor (IOS) procedure [10]. By combining the event-triggered communication and the IOS filtering procedure, the proposed approach ensures resilience to Byzantine behavior and guarantees convergence even in adversarial settings.

Compared to existing Byzantine-resilient approaches [5]–[12], the proposed method introduces a more adaptive and resource-efficient communication strategy, which extends battery life in sensor networks and lowers bandwidth requirements. Moreover, by limiting unnecessary message exchanges, the event-triggered strategy can reduce the exposure of the system to adversarial manipulations. Therefore, the proposed event-triggered Byzantine-resilient algorithm can enhance security and robustness of distributed optimization in resource-limited multiagent systems.

This paper is organized as follows. Section II addresses the formulation of a stochastic optimization problem with Byzantine attacks. Section III presents a distributed resilient algorithm with event-triggered communication. The analysis of the proposed algorithm is shown in Section IV. The numerical example is presented in Section V. Finally, concluding remarks are given in Section VI.

Notations: The sets of real numbers and nonnegative integers are represented by $\mathbb{R}$ and $\mathbb{N}$, respectively. $\|\cdot\|$ represents an ℓ_2 -norm of a vector or a spectral norm of a matrix. $\mathbf{1}_n$ represents the n -dimensional vector whose elements are all 1. For a set $\mathcal{S}$, $|\mathcal{S}|$ represents the cardinality

This work was partially supported by the joint project of Kyoto University and Toyota Motor Corporation, titled “Advanced Mathematical Science for Mobility Society” and JSPS KAKENHI Grant Number JP21K04121.

S. Tanaka, N. Hayashi, and M. Inuiguchi are with the Graduate School of Engineering Science, The University of Osaka, Toyonaka, Osaka 560-8531, Japan (e-mail: tanaka@inulab.sys.es.osaka-u.ac.jp, {n.hayashi, inuiguti}@sys.es.osaka-u.ac.jp). (Corresponding author: Shota Tanaka)

of $\mathcal{S}$. When referring to the elements of a matrix, the notation $[A]_{ij}$ is used to represent the (i, j) -th element of the matrix A . The projection of $x \in \mathbb{R}^n$ to a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$ is given by $\Pi_{\mathcal{Z}}[x] = \arg \min_{y \in \mathcal{Z}} \|x - y\|$.

II. PROBLEM FORMULATION

We consider a multiagent system where the information exchange among n agents is represented by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \dots, n\}$ is the set of agents and $\mathcal{E}$ is the set of the communication links. The graph $\mathcal{G}$ is said to have a spanning tree if there exists a node ℓ such that there is a path from ℓ to j for all nodes $j \in \mathcal{V} \setminus \{\ell\}$. For any given $r \in \mathbb{N}$, a subset of agents $\mathcal{C} \subseteq \mathcal{V}$ is said to be r -reachable if there exists an agent $i \in \mathcal{C}$ such that $|\mathcal{N}_i \setminus \mathcal{C}| \geq r$, where $\mathcal{N}_i = \{j \in \mathcal{V} \mid (j, i) \in \mathcal{E}, j \neq i\}$ is the set of in-neighbor agents of agent i [6]. For $r \in \mathbb{N}$, $\mathcal{G}$ is said to be r -robust if for all pairs of disjoint nonempty subsets $\mathcal{C}_1, \mathcal{C}_2 \subset \mathcal{V}$, at least one of $\mathcal{C}_1$ or $\mathcal{C}_2$ is r -reachable [6]. An agent $j \in \mathcal{V}$ is said to be a Byzantine agent if it is capable of sending arbitrary values to different neighbor agents [6]. In this paper, the set of Byzantine agents is defined by $\mathcal{B} \subset \mathcal{V}$ and $\mathcal{R} = \mathcal{V} \setminus \mathcal{B}$ is the set of normal agents which cooperatively solve a given optimization problem. The number of the normal agents is represented by $R = |\mathcal{R}|$.

Let ξ_i be an independent sample of agent i , where $\xi_i \sim \mathcal{S}_i$ and $\mathcal{S}_i$ is the data set of agent i . Each agent i has a local non-convex cost function $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$ such that $f_i(x) = \mathbb{E}_{\xi_i}[f_i(x; \xi_i)]$, where $f_i(x; \xi_i)$ represents the local loss associated with the random variable ξ_i . A group of normal agents collaboratively solve the following stochastic optimization problem:

$$\text{minimize } f(x) = \frac{1}{R} \sum_{i \in \mathcal{R}} f_i(x). \quad (1)$$

We consider a distributed algorithm to estimate a critical point $x^* \in \mathbb{R}^m$ of the problem (1). Let $x_{i,t} \in \mathbb{R}^m$ be the estimate of the critical point x^* of agent i at iteration t of the algorithm (see Algorithm 1). The random variable $\xi_{i,t}$ is assumed to be independent of the estimates $\{x_{i,\tau}\}_{\tau=0}^t$. We define $\nabla f_i(x; \xi_{i,t})$ as the stochastic gradient of f_i at $x \in \mathbb{R}^m$, where $\xi_{i,t}$ is the random variable generated by agent i at iteration t .

Assumption 1: i) $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$ is a smooth function for all $i \in \mathcal{R}$.

ii) There exists $\check{f} > -\infty$ such that $f_i(x) \geq \check{f}$ for all $i \in \mathcal{R}$ and $x \in \mathbb{R}^m$.

iii) The expected value of the stochastic gradient satisfies $\mathbb{E}_{\xi_{i,t}}[\nabla f_i(x; \xi_{i,t})] = \nabla f_i(x)$ for all $i \in \mathcal{R}$, $t \in \mathbb{N}$, and $x \in \mathbb{R}^m$.

iv) There exists a positive constant σ such that $\mathbb{E}_{\xi_{i,t}}[\|\nabla f_i(x; \xi_{i,t}) - \nabla f_i(x)\|^2] \leq \sigma^2$ for all $i \in \mathcal{R}$, $t \in \mathbb{N}$, and $x \in \mathbb{R}^m$.

v) There exists a positive constant δ such that $\max_{i \in \mathcal{R}} \|\nabla f_i(x) - \nabla f(x)\|^2 \leq \delta^2$ for all $i \in \mathcal{R}$ and $x \in \mathbb{R}^m$.

vi) There exists a positive constant D_f such that $\|\nabla f_i(x; \xi_{i,t})\| \leq D_f$ for all $i \in \mathcal{R}$, $t \in \mathbb{N}$, and $x \in \mathbb{R}^m$.

Assumption 1 plays an important role in guaranteeing the convergence of the distributed optimization algorithm [17]. From the smoothness of the local cost function in Assumption 1, for all $i \in \mathcal{R}$ and $x, y \in \mathbb{R}^m$, there exists a positive constant L such that $\|\nabla f_i(x) - \nabla f_i(y)\| \leq L\|x - y\|$.

For simplicity of notation, in the following discussion, we use the abbreviation $\mathbb{E}$ for $\mathbb{E}_u$, where u is a random variable.

III. EVENT-TRIGGERED DISTRIBUTED RESILIENT ALGORITHM

In this section, we propose a Byzantine-resilient distributed optimization algorithm with event-triggered communication. The proposed algorithm is summarized in Algorithm 1.

At each step, normal agent i computes the intermediate variable $z_{i,t}$, which corresponds to a projected stochastic gradient descent step, where $\mathcal{Z} \subset \mathbb{R}^m$ is a closed and convex set. At the first iteration, the agent initializes a variable $\tilde{z}_{i,0}$ as $z_{i,0}$, which stores the most recent estimate that was transmitted to its neighbors. The variable $\tilde{z}_{i,t} \in \mathbb{R}^m$ represents the estimate that the normal agent i sent to the neighbors at the last trigger time until iteration t :

$$\tilde{z}_{i,t} = \begin{cases} z_{i,t}, & \text{if } t \in \kappa^i, \\ \tilde{z}_{i,t-1}, & \text{otherwise,} \end{cases} \quad (2)$$

where κ^i is the set of all trigger times of agent i . We note that, from the projection step (3), there exists a positive constant D_z such that $\|z_{i,t}\| \leq D_z$ for all $i \in \mathcal{R}$ and $t \in \mathbb{N}$.

If the error between their current estimate $z_{i,t}$ and their last transmitted estimate $\tilde{z}_{i,t-1}$ exceeds a threshold $E_{i,t}$, the normal agent updates $\tilde{z}_{i,t}$ to match the current estimate $z_{i,t}$ and sends this updated value to its neighbors. Similarly, if the trigger condition is satisfied for one of the neighbors, the agent receives the updated estimate from that neighbor. Otherwise, the agent continues using the last received value.

After exchanging information with its neighbors, each normal agent runs the Iterative Outlier Scissor (IOS) procedure [10] in Algorithm 2 to filter out potentially faulty estimates. This filtering ensures that the normal agent only incorporates trustworthy information from its neighbors into its update. At each step of the IOS procedure, the agent first calculates the total weight $w'_i(\mathcal{U}_{i,t}^{(s)})$ in the current trusted set at Step 3.

Here, we consider the weight w_{ij} of the communication link $\{i, j\}$ that determines how much influence each neighbor's estimate has on the overall decision.

Assumption 2: For each agent $i \in \mathcal{V}$, there exists a positive constant $C_w \in (0, 1)$ such that

$$w_{ij} \begin{cases} \geq C_w, & j \in \mathcal{N}_i, \\ = 0, & j \notin \mathcal{N}_i, \end{cases}$$

and $w_{ii} = 1 - \sum_{j \in \mathcal{N}_i} w_{ij} > C_w$. $\forall j \in \mathcal{V}$.

With these weights, the normal agent i computes a weighted average $\tilde{z}_{i,t}^{(s)}$ of all the event-triggered estimates

Algorithm 1 Event-triggered resilient algorithm of normal agent i

Require: Initial estimate $x_{i,0} \in \mathbb{R}^m$.

```

1: while  $t = 0, 1, \dots, T - 1$  do
2:   Compute the variable  $z_{i,t}$  by
      
$$z_{i,t} = \Pi_{\mathcal{Z}}[x_{i,t} - \alpha \nabla f_i(x_{i,t}; \xi_{i,t})]. \quad (3)$$

3:   if  $t = 0$  then
4:     Set  $\tilde{z}_{i,0} = z_{i,0}$ .
5:   end if
6:   if trigger error of agent  $i$  satisfies  $\|\tilde{z}_{i,t-1} - z_{i,t}\| > E_{i,t}$  then
7:     Set  $\tilde{z}_{i,t} = z_{i,t}$ .
8:     Send  $\tilde{z}_{i,t}$  to neighbor agent  $j \in \mathcal{N}_i$ .
9:   else
10:    Set  $\tilde{z}_{i,t} = \tilde{z}_{i,t-1}$ .
11:  end if
12:  if trigger error of neighbor agent  $j \in \mathcal{N}_i$  satisfies  $\|\tilde{z}_{i,t-1} - z_{i,t}\| > E_{j,t}$  then
13:    Receive  $\tilde{z}_{j,t}$  from  $j$ .
14:  end if
15:  Run the IOS procedure in Algorithm 2.
16:  Update the estimate by
      
$$x_{i,t+1} = (1 - r_{i,t}) \sum_{j \in \mathcal{U}_{i,t}^{(q)} \setminus \{i\}} p_{ij,t} \tilde{z}_{j,t} + r_{i,t} z_{i,t}. \quad (4)$$

17:  Set  $t := t + 1$ .
18: end while

```

in the trusted set $\mathcal{U}_{i,t}^{(s)}$ at Step 4 in Algorithm 2. This average serves as a reference point for filtering out the faulty estimates. At the next step, the normal agent identifies the estimate that deviates the most from the computed weighted average $\tilde{z}_{i,t}^{(s)}$ at Step 5. The agent $j^{(s)}$ whose estimate has the largest deviation from the average is flagged as the most suspicious at that iteration. The flagged agent is then removed from the trusted set at Step 6. After q iterations, Algorithm 2 outputs the set $\mathcal{U}_{i,t}^{(q)}$ that contains the neighbors of the normal agent i whose estimates have been considered reliable through the IOS procedure.

Based on the filtered information, at Step 16 in Algorithm 1, the normal agent i updates its own estimate by (4), which computes the weighted average of the agent's intermediate variable and the values received from neighboring agents who are considered trustworthy after the IOS procedure.

Let $e_{i,t}$ be the trigger error of the normal agent i due to the event-triggered communication, that is, $e_{i,t} = \tilde{z}_{i,t} - z_{i,t}$. Steps 6–14 in Algorithm 1 guarantee the boundedness of the trigger error as

$$\|e_{i,t}\| \leq E_t, \quad \forall i \in \mathcal{R}, \forall t \in \mathcal{T} = \{1, 2, \dots, T\}, \quad (5)$$

where $E_t = \max_{i \in \mathcal{R}} E_{i,t}$. In this paper, we assume the square-summability condition on the trigger threshold.

Algorithm 2 Iterative Outlier Scissor (IOS) procedure of normal agent i

Require: Estimates $\{\tilde{z}_{j,t}\}_{j \in \mathcal{N}_i \cup \{i\}}$,
Weights $\{w_{ij}\}_{j \in \mathcal{N}_i \cup \{i\}}$.

```

1: Construct the initial trusted set  $\mathcal{U}_{i,t}^{(0)} = \mathcal{N}_i \cup \{i\}$ .
2: for  $s = 0, 1, \dots, q - 1$  do
3:   Compute the weight  $w_{i,t}^{(s)}$  by  $w'_i(\mathcal{U}_{i,t}^{(s)}) = \sum_{j \in \mathcal{U}_{i,t}^{(s)}} w_{ij}$ .
4:   Compute the weighted average  $\tilde{z}_{i,t}^{(s)}$  by  $\tilde{z}_{i,t}^{(s)} = (1/w'_i(\mathcal{U}_{i,t}^{(s)})) \sum_{j \in \mathcal{U}_{i,t}^{(s)}} w_{ij} \tilde{z}_{j,t}$ .
5:   Choose the removing agent as  $j^{(s)} = \arg \max_{j \in \mathcal{U}_{i,t}^{(s)} \setminus \{i\}} \|\tilde{z}_{j,t} - \tilde{z}_{i,t}^{(s)}\|$ .
6:   Remove agent  $j^{(s)}$  from the trusted set as  $\mathcal{U}_{i,t}^{(s+1)} = \mathcal{U}_{i,t}^{(s)} \setminus \{j^{(s)}\}$ .
7: end for
8: return  $\mathcal{U}_{i,t}^{(q)}$ .

```

Assumption 3: There exists a positive constant E such that $\sum_{t=1}^{\infty} E_t^2 \leq E$.

To guarantee the convergence, we also make the following assumptions:

Assumption 4: $\mathcal{G}$ is $(q+1)$ -robust, where q is the number of iterations performed in the IOS procedure.

Assumption 5: For all $i \in \mathcal{V}$, $t \in \mathcal{T}$, and $s \in \mathcal{S} = \{0, 1, \dots, q-1\}$, $w'_i(\mathcal{B}_{i,t}^{(s)}) < w'_i(\mathcal{R}_{i,t}^{(s)})$ holds, where $\mathcal{R}_{i,t}^{(s)} = \mathcal{R} \cap \mathcal{U}_{i,t}^{(s)}$ and $\mathcal{B}_{i,t}^{(s)} = \mathcal{B} \cap \mathcal{U}_{i,t}^{(s)}$.

Assumption 5 is naturally satisfied if Assumption 2 holds and if the number of Byzantine neighbors is relatively small compared to the number of normal neighbors. Moreover, under Assumption 4, the connectivity of the graph is maintained after applying the IOS procedure, as stated in Lemma 1 [6]:

Lemma 1: Suppose that $\mathcal{G}$ satisfies Assumption 4. Let $\mathcal{G}'$ be a graph obtained by removing q or fewer incoming edges from each node in $\mathcal{G}$. Then, $\mathcal{G}'$ has a spanning tree.

IV. CONVERGENCE ANALYSIS

Let $\mathcal{R}_i$ and $\mathcal{B}_i$ represent the sets of normal and Byzantine neighbors of agent i , that is, $\mathcal{R}_i = \mathcal{N}_i \cap \mathcal{R}$ and $\mathcal{B}_i = \mathcal{N}_i \cap \mathcal{B}$. We define the set $\mathcal{U}_i^{\max}$ as the subset of q neighbors of node i whose sum of the weights is the largest, that is, $\mathcal{U}_i^{\max} = \arg \max_{\mathcal{U} \subseteq \mathcal{N}_i, |\mathcal{U}|=q} \sum_{j \in \mathcal{U}} w'_{ij}$.

We introduce the weight matrix for normal agents $\Omega \in \mathbb{R}^{R \times R}$, where the (i, j) -th element is defined as

$$[\Omega]_{ij} = \begin{cases} w_{ii} + \sum_{\ell \in \mathcal{B}_i} w_{i\ell}, & j = i, i \in \mathcal{R}, \\ w_{ij}, & j \neq i, i, j \in \mathcal{R}. \end{cases} \quad (6)$$

From Assumption 2 and (6), the matrix Ω is row-stochastic, that is, for all $i \in \mathcal{R}$, the following holds:

$$\sum_{j \in \mathcal{R}_i \cup \{i\}} [\Omega]_{ij} = 1, \quad \forall i \in \mathcal{R}. \quad (7)$$

Let $\bar{z}_{i,t}$ be the weighted average of the intermediate variable $z_{i,t}$ such that $\bar{z}_{i,t} = \sum_{j \in \mathcal{R}_i \cup \{i\}} [\Omega]_{ij} z_{j,t}$.

Lemma 2: Suppose that Assumptions 1–5 hold. Suppose also that $B \leq q$ and $w'_i(\mathcal{U}_i^{\max}) < 1/3$. Then, for all $i \in \mathcal{R}$ and $t \in \mathcal{T}$, there exists a constant $\rho > 0$ such that

$$\|x_{i,t+1} - \bar{z}_{i,t}\| \leq \rho \max_{j \in \mathcal{R}_i \cup \{i\}} \|\bar{z}_{j,t} - \bar{z}_{i,t}\| + 2\rho D_z + 2E_t. \quad (8)$$

We define the average estimate of all normal agents as $\bar{x}_t = (1/R) \sum_{i \in \mathcal{R}} x_{i,t}$. We also define the mean squared error as $H_t = (1/R) \sum_{i \in \mathcal{R}} \|x_{i,t} - \bar{x}_t\|^2$.

The next theorem is the main result of this paper, which shows the convergence of the proposed event-triggered Byzantine-resilient algorithm in terms of an upper bound on the average squared gradient norm of the global cost function.

Theorem 1: Suppose that Assumptions 1–5 hold. Suppose also that $B \leq q$ and $w'_i(\mathcal{U}_i^{\max}) < 1/3$. If the step-size satisfies $\alpha \leq 1/(2\sqrt{3}L)$, then we have

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\|\nabla f(\bar{x}_t)\|^2] \\ & \leq \frac{2f(\bar{x}_0)}{\alpha T} + 2\alpha\sigma^2 L \\ & \quad + 3R \left(L^2 + \frac{8(\chi + 9\rho^2)(1 + 3L^2(R+1)\alpha^2)}{\alpha^2} \right) \\ & \quad \times \frac{1}{T} \sum_{t=1}^T \mathbb{E}[H_t] \\ & \quad + \frac{3(32\chi + 297\rho^2 + 16)E}{\alpha^2} \frac{1}{T} + \frac{36\rho^2 D_z^2}{\alpha^2} \\ & \quad + 72(\chi + 9\rho^2)R\delta^2 + 12(8(\chi + 9\rho^2)R + 1)\sigma^2 \\ & \quad + 12(8(\chi + 9\rho^2) + 1)D_f^2, \end{aligned} \quad (9)$$

where $\chi = (1/R)\|\Omega^\top \mathbf{1}_R - \mathbf{1}_R\|^2$.

Proof: From the smoothness of the local cost function, we have

$$\begin{aligned} \mathbb{E}[f(\bar{x}_{t+1})] & \leq f(\bar{x}_t) + \mathbb{E}[\langle \nabla f(\bar{x}_t), \bar{x}_{t+1} - \bar{x}_t \rangle] \\ & \quad + \frac{L}{2} \mathbb{E}[\|\bar{x}_{t+1} - \bar{x}_t\|^2]. \end{aligned} \quad (10)$$

For vectors $x, y \in \mathbb{R}^m$, we have $\langle x, y \rangle = (1/2)\|x+y\|^2 - (1/2)\|x\|^2 - (1/2)\|y\|^2$. Then, the second term of the R.H.S. of (10) is given by

$$\begin{aligned} & \mathbb{E}[\langle \nabla f(\bar{x}_t), \bar{x}_{t+1} - \bar{x}_t \rangle] \\ & = \mathbb{E} \left[\alpha \left\langle \nabla f(\bar{x}_t), \nabla f(\bar{x}_t; \xi_t) - \nabla f(\bar{x}_t) + \frac{1}{\alpha}(\bar{x}_{t+1} - \bar{x}_t) \right\rangle \right] \\ & = \frac{\alpha}{2} \mathbb{E} \left[\left\| \nabla f(\bar{x}_t; \xi_t) + \frac{1}{\alpha}(\bar{x}_{t+1} - \bar{x}_t) \right\|^2 \right] - \frac{\alpha}{2} \mathbb{E}[\|\nabla f(\bar{x}_t)\|^2] \\ & \quad - \frac{\alpha}{2} \mathbb{E} \left[\left\| \nabla f(\bar{x}_t; \xi_t) - \nabla f(\bar{x}_t) + \frac{1}{\alpha}(\bar{x}_{t+1} - \bar{x}_t) \right\|^2 \right], \end{aligned} \quad (11)$$

where $\xi_t = \{\xi_{i,t}\}_{i \in \mathcal{R}}$ and the second equality follows from Assumption 1 iii).

The third term at the R.H.S. of (10) is bounded by

$$\begin{aligned} & \mathbb{E}[\|\bar{x}_{t+1} - \bar{x}_t\|^2] \\ & = \mathbb{E} \left[\left\| \alpha(\nabla f(\bar{x}_t; \xi_t) - \nabla f(\bar{x}_t)) \right. \right. \\ & \quad \left. \left. - \alpha(\nabla f(\bar{x}_t; \xi_t) - \nabla f(\bar{x}_t)) + \alpha \frac{1}{\alpha}(\bar{x}_{t+1} - \bar{x}_t) \right\|^2 \right] \\ & \leq 2\alpha^2\sigma^2 \\ & \quad + 2\alpha^2 \mathbb{E} \left[\left\| \nabla f(\bar{x}_t; \xi_t) - \nabla f(\bar{x}_t) + \frac{1}{\alpha}(\bar{x}_{t+1} - \bar{x}_t) \right\|^2 \right], \end{aligned} \quad (12)$$

where the last inequality follows from the fact that $\|a+b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$ for vectors $a, b \in \mathbb{R}^m$ and Assumption 1 iv).

Since $\alpha \leq 1/(2L)$, by substituting (11) and (12) for (10), we have

$$\begin{aligned} & \mathbb{E}[f(\bar{x}_{t+1})] \\ & \leq f(\bar{x}_t) + \frac{\alpha}{2} \mathbb{E} \left[\left\| \nabla f(\bar{x}_t; \xi_t) + \frac{1}{\alpha}(\bar{x}_{t+1} - \bar{x}_t) \right\|^2 \right] \\ & \quad - \frac{\alpha}{2} (1 - 2\alpha L) \\ & \quad \times \mathbb{E} \left[\left\| \nabla f(\bar{x}_t; \xi_t) - \nabla f(\bar{x}_t) + \frac{1}{\alpha}(\bar{x}_{t+1} - \bar{x}_t) \right\|^2 \right] \\ & \quad - \frac{\alpha}{2} \|\nabla f(\bar{x}_t)\|^2 + \alpha^2\sigma^2 L \\ & \leq f(\bar{x}_t) + \frac{\alpha}{2} \mathbb{E} \left[\left\| \nabla f(\bar{x}_t; \xi_t) + \frac{1}{\alpha}(\bar{x}_{t+1} - \bar{x}_t) \right\|^2 \right] \\ & \quad - \frac{\alpha}{2} \|\nabla f(\bar{x}_t)\|^2 + \alpha^2\sigma^2 L. \end{aligned} \quad (13)$$

The second term of the R.H.S. of (13) is given by

$$\begin{aligned} & \left\| \nabla f(\bar{x}_t; \xi_t) + \frac{1}{\alpha}(\bar{x}_{t+1} - \bar{x}_t) \right\|^2 \\ & = \left\| \nabla f(\bar{x}_t; \xi_t) + \frac{1}{\alpha R} \sum_{i \in \mathcal{R}} (x_{i,t+1} - \bar{x}_t) \right\|^2 \\ & \leq 3S_1 + 3S_2 + 3S_3, \end{aligned} \quad (14)$$

where $S_1 = \|\nabla f(\bar{x}_t; \xi_t) - (1/R) \sum_{i \in \mathcal{R}} \nabla f_i(x_{i,t}; \xi_{i,t})\|^2$, $S_2 = \|(1/(\alpha R)) \sum_{i \in \mathcal{R}} (\bar{z}_{i,t} - \bar{x}_t + (\alpha/R) \sum_{\ell \in \mathcal{R}} \nabla f_\ell(x_{\ell,t}; \xi_{\ell,t}))\|^2$, and $S_3 = \|(1/(\alpha R)) \sum_{i \in \mathcal{R}} (x_{i,t+1} - \bar{z}_{i,t})\|^2$. We note that $\mathbb{E}[\|x\|^2] = \|\mathbb{E}[x]\|^2 + \mathbb{E}[\|x - \mathbb{E}[x]\|^2]$ holds for any probability variable $x \in \mathbb{R}^m$. Thus, for S_1 , we have

$$\begin{aligned} & \mathbb{E}[\|\nabla f_i(\bar{x}_t; \xi_{i,t}) - \nabla f_i(x_{i,t}; \xi_{i,t})\|^2] \\ & = \|\nabla f_i(\bar{x}_t) - \nabla f_i(x_{i,t})\|^2 \\ & \quad + \mathbb{E}[\|(\nabla f_i(\bar{x}_t; \xi_{i,t}) - \nabla f_i(x_{i,t}; \xi_{i,t})) \\ & \quad - (\nabla f_i(\bar{x}_t) - \nabla f_i(x_{i,t}))\|^2] \\ & \leq L^2 \|\bar{x}_t - x_{i,t}\|^2 + 4\sigma^2, \end{aligned} \quad (15)$$

where the first equality and the last inequality follow from Assumption 1 iii) and iv), respectively. Thus, S_1 is bounded

as follows:

$$\mathbb{E}[S_1] \leq L^2 R \frac{1}{R} \sum_{i \in \mathcal{R}} \|x_{i,t} - \bar{x}_t\|^2 + 4\sigma^2. \quad (16)$$

For the term S_2 , we have

$$\begin{aligned} & \frac{1}{\alpha R} \sum_{i \in \mathcal{R}} \left(\bar{z}_{i,t} - \bar{x}_t + \frac{\alpha}{R} \sum_{\ell \in \mathcal{R}} \nabla f_\ell(x_{\ell,t}; \xi_{\ell,t}) \right) \\ &= \frac{1}{\alpha R} \sum_{i \in \mathcal{R}} (\bar{z}_{i,t} - \bar{z}_t) + \frac{1}{\alpha R} \sum_{\ell \in \mathcal{R}} \psi_{\ell,t}, \end{aligned} \quad (17)$$

where $\bar{z}_t = (1/R) \sum_{i \in \mathcal{R}} z_{i,t}$ and $\psi_{i,t} = z_{i,t} - (x_{i,t} - \alpha \nabla f_i(x_{i,t}; \xi_{i,t}))$. Since $\sum_{j \in \mathcal{U}_{i,t}^{(q)}} (w_{ij} / \sum_{\ell \in \mathcal{U}_{i,t}^{(q)}} w_{i\ell}) = 1$ and $z_{i,t} \in \mathcal{Z}$ for all $i \in \mathcal{R}$ and $t \in \mathbb{N}$, we have $v_{i,t}^{(q)} \in \mathcal{Z}$ for all $i \in \mathcal{R}$ and $t \in \mathbb{N}$. Then, from the property of the projection in [18], we have $\|z_{i,t+1} - v_{i,t}^{(q)}\|^2 \leq \|\varepsilon_{i,t}^{(q)} - \alpha \nabla f_i(x_{i,t+1}; \xi_{i,t+1})\|^2 - \|\psi_{i,t+1}\|^2$. This yields

$$\|\psi_{i,t+1}\| \leq E_t + D_f \alpha. \quad (18)$$

From (17) and (18), we have

$$\begin{aligned} S_2 &\leq 2 \left\| \frac{1}{\alpha R} \sum_{i \in \mathcal{R}} (\bar{z}_{i,t} - \bar{z}_t) \right\|^2 + 2 \left\| \frac{1}{\alpha R} \sum_{\ell \in \mathcal{R}} \psi_{\ell,t} \right\|^2 \\ &= 2 \left\| \frac{1}{\alpha R} \mathbf{1}_R^\top \left(\Omega Z_t - \frac{1}{R} \mathbf{1}_R \mathbf{1}_R^\top Z_t \right) \right\|^2 \\ &\quad + 2 \frac{1}{\alpha^2 R^2} R \sum_{\ell \in \mathcal{R}} \|\psi_{\ell,t}\|^2 \\ &= \frac{2}{\alpha^2 R^2} \left\| (\mathbf{1}_R^\top \Omega - \mathbf{1}_R^\top) \left(Z_t - \frac{1}{R} \mathbf{1}_R \mathbf{1}_R^\top Z_t \right) \right\|^2 \\ &\quad + \frac{2}{\alpha^2 R} \sum_{\ell \in \mathcal{R}} \|\psi_{\ell,t}\|^2 \\ &\leq \frac{2\chi R}{\alpha^2} \frac{1}{R} \max_{i \in \mathcal{R}} \|z_{i,t} - \bar{z}_t\|^2 + \frac{4}{\alpha^2} (E_t^2 + D_f^2 \alpha^2), \end{aligned} \quad (19)$$

where $Z_t \in \mathbb{R}^{R \times m}$ is the matrix whose rows are the vectors $\{z_{i,t}^\top\}_{i \in \mathcal{R}}$.

From the contraction property in (8), S_3 is bounded as follows:

$$\begin{aligned} S_3 &= \left\| \frac{1}{\alpha R} \sum_{i \in \mathcal{R}} (x_{i,t+1} - \bar{z}_{i,t}) \right\|^2 \\ &\leq \frac{1}{\alpha^2 R} \sum_{i \in \mathcal{R}} \left(\rho \max_{j \in \mathcal{R}_i \cup \{i\}} \|\bar{z}_{j,t} - \bar{z}_{i,t}\| + 2\rho D_z + 2E_t \right)^2 \\ &\leq \frac{3\rho^2}{\alpha^2 R} \sum_{i \in \mathcal{R}} \max_{j \in \mathcal{R}_i \cup \{i\}} \|\bar{z}_{j,t} - \bar{z}_{i,t}\|^2 + \frac{12\rho^2 D_z^2}{\alpha^2} + \frac{12}{\alpha^2} E_t^2. \end{aligned} \quad (20)$$

For any normal agent $i \in \mathcal{R}$, we have

$$\begin{aligned} \|\bar{z}_{i,t} - \bar{z}_t\| &\leq \sum_{j \in \mathcal{R}_i \cup \{i\}} [\Omega]_{ij} \|z_{j,t} - \bar{z}_t\| \\ &\leq \max_{\ell \in \mathcal{R}} \|z_{\ell,t} - \bar{z}_t\|, \end{aligned} \quad (21)$$

where the last equality follows from (7). Therefore, for $i \in \mathcal{R}$, we have

$$\begin{aligned} & \sum_{i \in \mathcal{R}} \max_{j \in \mathcal{R}_i \cup \{i\}} \|\bar{z}_{j,t} - \bar{z}_{i,t}\|^2 \\ &= \sum_{i \in \mathcal{R}} \max_{j \in \mathcal{R}_i \cup \{i\}} \|(\bar{z}_{j,t} - z_{j,t}) + (z_{j,t} - \bar{z}_t) - (\bar{z}_t - \bar{z}_{i,t})\|^2 \\ &\leq 6R \max_{i \in \mathcal{R}} \|z_{i,t} - \bar{z}_t\|^2 + 3RE_t^2. \end{aligned} \quad (22)$$

From (20) and (22), we have

$$S_3 \leq \frac{18\rho^2}{\alpha^2} \max_{i \in \mathcal{R}} \|z_{i,t} - \bar{z}_t\|^2 + \frac{12\rho^2 D_z^2}{\alpha^2} + \frac{3(3\rho^2 + 4)}{\alpha^2} E_t^2. \quad (23)$$

By substituting (16), (19), and (23) for (14), we have

$$\begin{aligned} & \mathbb{E} \left[\left\| \nabla f(\bar{x}_t; \xi_t) + \frac{1}{\alpha} (\bar{x}_{t+1} - \bar{x}_t) \right\|^2 \right] \\ &\leq 3L^2 R \frac{1}{R} \sum_{i \in \mathcal{R}} \|x_{i,t} - \bar{x}_t\|^2 + 12(\sigma^2 + D_f^2) \\ &\quad + \frac{3(6\chi + 9\rho^2)R}{\alpha^2} \mathbb{E} \left[\frac{1}{R} \max_{i \in \mathcal{R}} \|z_{i,t} - \bar{z}_t\|^2 \right] \\ &\quad + \frac{3(9\rho^2 + 16)}{\alpha^2} E_t^2 + \frac{36\rho^2 D_z^2}{\alpha^2}. \end{aligned} \quad (24)$$

By following the similar argument of Lemma 2 in [10], we obtain

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{R} \max_{i \in \mathcal{R}} \|z_{i,t} - \bar{z}_t\|^2 \right] \\ &\leq 4(1 + 3L^2(R+1)\alpha^2)H_t + 4\alpha^2(3\delta^2 + 4\sigma^2) \\ &\quad + \frac{16}{R} (E_t^2 + D_f^2 \alpha^2). \end{aligned} \quad (25)$$

From (13), (24), and (25), we have (9). $\blacksquare$

V. NUMERICAL EXAMPLE

In this section, we evaluate the effectiveness of the proposed event-triggered Byzantine-resilient distributed optimization algorithm (hereinafter referred to as ET-IOS). We consider a classification task by softmax regression using the MNIST dataset [19]. We consider a complete graph with 10 agents, where 2 agents behave as Byzantine agents. The data is partitioned among agents in a non-i.i.d. manner, where each agent holds data belonging to only one specific class. As the Byzantine attack strategy, we employ the Gaussian attack [4], where Byzantine agents send values that are close to the local average of normal agents' transmitted values but are intentionally corrupted by adding Gaussian noise with a large variance.

We evaluate the impact of the trigger threshold E on the learning accuracy and communication cost of the proposed ET-IOS algorithm. We conduct experiments by varying the threshold $E \in \{0, 0.005, 0.01, 0.012, 0.015, 0.02\}$.

Fig. 1 shows the test accuracy evolution for each threshold value. $E = 0$ corresponds to the case with time-triggered method [10], where communication occurs at every iteration. As observed from the figure, even when the threshold E is

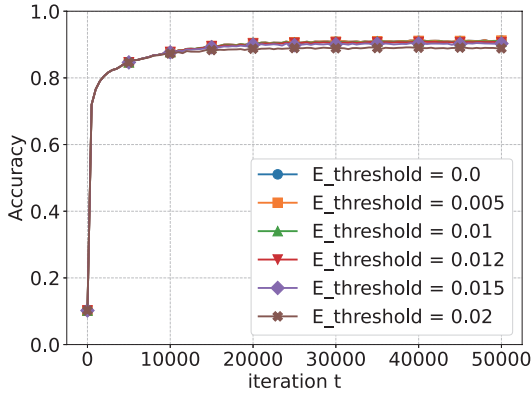


Fig. 1. Test accuracy evolution for ET-IOS with different trigger thresholds.

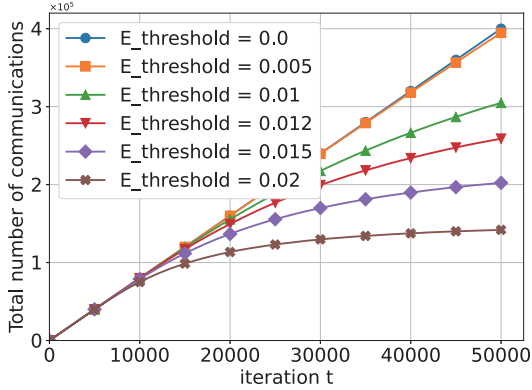


Fig. 2. Cumulative communication events for ET-IOS with different trigger thresholds.

increased from 0 to 0.02, the final test accuracy remains largely unaffected, maintaining a high level (approximately 91%).

Fig. 2 shows the evolution of cumulative communication events for different thresholds. As expected, a larger threshold E leads to less frequent communication, resulting in a slower increase in the cumulative communication count. The ratio of total communication events compared to the case with time-triggering ($E = 0$) can be reduced to 98.7% by setting the threshold $E = 0.005$, 76.2% for $E = 0.01$, and 35.5% for $E = 0.02$.

The results of Figs. 1 and 2 show that the proposed ET-IOS algorithm can achieve substantial communication cost reduction by appropriately setting the trigger threshold while maintaining high accuracy.

VI. CONCLUSION

This paper investigated a distributed Byzantine-resilient algorithm for non-convex optimization that incorporates event-triggered communication. By introducing an adaptive communication strategy, the proposed approach significantly reduces unnecessary message exchanges, thereby enhancing the efficiency and resilience of large-scale distributed optimization. We showed that the proposed algorithm maintains convergence properties even in the presence of Byzantine agents. Future work will focus on extending the proposed

framework to more general settings, such as time-varying networks and heterogeneous agent capabilities.

REFERENCES

- [1] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.
- [2] Y. Zhang, M. Qiu, and H. Gao, "Communication-efficient stochastic gradient descent ascent with momentum algorithms," *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 4602–4610, 2023.
- [3] D. Ishikawa, N. Hayashi, and S. Takai, "Distributed event-triggered stochastic gradient-tracking for nonconvex optimization," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E107-A, no. 5, pp. 762–769, 2024.
- [4] M. Fang, X. Cao, J. Jia, and N. Gong, "Local model poisoning attacks to Byzantine-robust federated learning," *Proceedings of the 29th USENIX security symposium*, no. 1605–1622, 2020.
- [5] Z. Yang and W. U. Bajwa, "ByRDIE: Byzantine-resilient distributed coordinate descent for decentralized learning," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 4, pp. 611–627, 2019.
- [6] S. Sundaram and B. Ghahesifard, "Distributed optimization under adversarial nodes," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1063–1076, 2019.
- [7] Y. Chen, L. Su, and J. Xu, "Distributed statistical machine learning in adversarial settings: Byzantine gradient descent," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 1, no. 2, pp. 1–25, 2017.
- [8] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," *Advances in neural information processing systems*, vol. 30, 2017.
- [9] L. Su and N. H. Vaidya, "Byzantine-resilient multiagent optimization," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2227–2233, 2020.
- [10] Z. Wu, T. Chen, and Q. Ling, "Byzantine-resilient decentralized stochastic optimization with robust aggregation rules," 2023. [Online]. Available: <https://arxiv.org/abs/2206.04568>
- [11] K. Kuwarananchaoen, L. Xin, and S. Sundaram, "Scalable distributed optimization of multi-dimensional functions despite byzantine adversaries," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 10, pp. 360–375, 2024.
- [12] M. Yemini, A. Nedić, A. J. Goldsmith, and S. Gil, "Resilient distributed optimization for multi-agent cyberphysical systems," *IEEE Transactions on Automatic Control*, pp. 1–16, 2025.
- [13] Y. Kajiyama, N. Hayashi, and S. Takai, "Distributed subgradient method with edge-based event-triggered communication," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 2248–2255, 2018.
- [14] X. Cao and T. Başar, "Decentralized online convex optimization with event-triggered communications," *IEEE Transactions on Signal Processing*, vol. 69, pp. 284–299, 2021.
- [15] N. Hayashi and K. Sakurama, "Communication-aware distributed rebalancing for cooperative car-sharing service," *IET Control Theory & Applications*, vol. 17, no. 7, pp. 850–867, 2023.
- [16] Z. Dong, Y. Jin, S. Mao, W. Ren, W. Du, and Y. Tang, "Distributed optimization with asynchronous computation and event-triggered communication," *IEEE Transactions on Automatic Control*, vol. 70, no. 2, pp. 1084–1099, 2025.
- [17] R. Xin, U. A. Khan, and S. Kar, "An improved convergence analysis for decentralized online stochastic non-convex optimization," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1842–1858, 2021.
- [18] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.
- [19] L. Deng, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.