SCIENCE CHINA
Information Sciences

September 2023, Vol. 66 190201:1–190201:16
https://doi.org/10.1007/s11432-022-3702-4

• RESEARCH PAPER •
Special Topic: Control, Optimization, and Learning for Networked Systems

Data-driven cooperative output regulation of multi-agent systems under distributed denial of service attacks

Weinan GAO[1*] & Zhong-Ping JIANG[2*]

[1]*State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang* 110819*, China;*
[2]*Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn* NY 11201*, USA*

Received 31 October 2022/Revised 16 December 2022/Accepted 29 January 2023/Published online 28 August 2023

Abstract This paper addresses an optimal, cooperative output regulation problem for multi-agent systems with distributed denial of service attacks and unknown system dynamics. Unlike existing studies, the proposed solution is essentially a learning-based control strategy such that one can obtain a distributed control policy with internal models through online data and analyze the resilience of closed-loop systems, both without the precise knowledge of system dynamics in the state-space model. The efficiency of the proposed methodology is validated using computer simulations.

Keywords cooperative output regulation, adaptive dynamic programming, learning-based control, multi-agent systems, cybersecurity

Citation Gao W N, Jiang Z-P. Data-driven cooperative output regulation of multi-agent systems under distributed denial of service attacks. Sci China Inf Sci, 2023, 66(9): 190201, https://doi.org/10.1007/s11432-022-3702-4

1 Introduction

Cooperative output regulation is concerned with enabling multiple autonomous systems to track some reference signals and reject some disturbances, where references and disturbances are generated by an autonomous system named the exosystem or the leader [1–5]. The problem itself, therefore, includes the leader-follower consensus problem of multi-agent systems as a special case. Moreover, some communication constraints are usually found in the cooperative output regulation problem, such that some agents cannot directly communicate with the leader, making the problem more challenging.

To address cooperative output regulation problems, distributed control techniques [6, 7] have been introduced and combined with either feedback-feedforward control or the internal model principle, which are popular strategies for solving traditional output regulation problems. When the leader's system dynamics is unavailable for all other agents, a distributed adaptive internal model may be used [8], which comprises an adaptive distributed observer [9] and a distributed internal model. However, most of the present studies on cooperative output regulation problems have neglected the optimality of the closed-loop system with unknown system dynamics.

Adaptive dynamic programming (ADP) is a learning-based adaptive optimal control method that can be applied to learn toward the optimal control policy through online collections with unknown or uncertain system dynamics [10–14]. Therefore, it is a good candidate for addressing the cooperative optimal output regulation problem in a data-driven manner; see [8, 15, 16]. Nevertheless, most learning-based solutions to cooperative (optimal) output regulation problems assume that the system is not under cyberattacks. The purpose of this paper is to investigate when multi-agent systems in a closed loop with learned control policies are sufficiently resilient against malicious cyberattacks.

* Corresponding author (email: gaown@mail.neu.edu.cn, zjiang@nyu.edu)

Denial of service (DoS) attacks are very common in modern control and communication systems, and they usually prevent single-agent systems or multi-agent systems as a whole from sensing or actuating information through network communication, which, for instance, has been studied in [17, 18]. As an enhancement of DoS attacks, distributed DoS (DDoS) attacks, which have been reported since the 1990s, leverage coordinated DoS attacks to disrupt an agent's connectivity in multi-agent systems; see [19, 20]. Unlike DoS attacks, DDoS attacks characterize the duration and frequency of attacks under a specific agent or a communication edge between agents.

To ensure the cybersecurity of control systems with respect to DDoS attacks, the event-triggered mechanism [21] and adaptive distributed resilient observers [22] were developed under the assumption that the system dynamics is known exactly.

In this paper, we aim to solve the cooperative optimal output regulation problems of multi-agent systems under DDoS attacks and unknown system dynamics. The technical difficulties and main contributions are summarized as follows. First, as we considered the onslaught of malicious DDoS attacks, traditional optimal controller design methods may not apply to achieving cooperative output regulation. To address this conundrum, we developed a learning-based control strategy based on ADP such that the optimal control gains can be learned in terms of online data with DDoS attacks. Second, analyzing the resilience of closed-loop systems against DDoS attacks is challenging, particularly when the system dynamics is not known exactly. To address this challenge, we applied the notion of input-to-state stability, Lyapunov stability theory, and comparison functions to propose a novel resilience analysis method for estimating the upper bound of DDoS attack duration that the closed-loop system can endure. Note that this analysis does not rely on the knowledge of any system matrices in the state equations. Third, the assumptions of DDoS attacks in this paper are made with respect to each agent, not the entire network [22] or the communication edges [21, 23]. This approach is more explicit if the designer is interested in observing the performance of specific agents.

The remainder of this paper is organized as follows. In Section 2, we formulate the control problem and recall preliminaries regarding the internal model principle and cooperative optimal output regulation. In Section 3, we include the data-driven controller design and resilience analysis with respect to DDoS attacks. To illustrate the efficiency of the proposed research, simulations and discussions are provided in Section 4. The conclusion is presented in Section 5.

Notations. Throughout this paper, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}_+$ the set of nonnegative real numbers, $\mathbb{Z}_+$ the set of nonnegative integers, and $\mathbb{N}_+$ the set of positive integers. The operator $|\cdot|$ represents the Euclidean norm for vectors and the induced norm for matrices. $\otimes$ indicates the Kronecker product operator, and $\mathrm{vec}(A) = [a_1^{\mathrm{T}}, a_2^{\mathrm{T}}, \ldots, a_m^{\mathrm{T}}]^{\mathrm{T}}$, where $a_i \in \mathbb{R}^n$, for $i = 1, \ldots, m$, are the first through last columns of $A \in \mathbb{R}^{n\times m}$. For an arbitrary column vector $v \in \mathbb{R}^n$, $\mathrm{vecv}(v) = [v_1^2, v_1v_2, \ldots, v_1v_n, v_2^2, v_2v_3, \ldots, v_{n-1}v_n, v_n^2]^{\mathrm{T}} \in \mathbb{R}^{\frac{1}{2}n(n+1)}$. $\mathrm{vecs}(P) = [p_{11}, 2p_{12}, \ldots, 2p_{1m}, p_{22}, 2p_{23}, \ldots, 2p_{m-1,m}, p_{mm}]^{\mathrm{T}} \in \mathbb{R}^{\frac{1}{2}m(m+1)}$ for a symmetric matrix $P \in \mathbb{R}^{m\times m}$, and $\lambda_M(P)$ and $\lambda_m(P)$ denote the maximum and minimum eigenvalue of a real symmetric matrix P, respectively. $P \succ 0$ means that the matrix P is positive definite, i.e., $x^{\mathrm{T}}Px > 0$ for all nonzero vectors $x \in \mathbb{R}^m$. For any piecewise continuous function $u : \mathbb{R}_+ \to \mathbb{R}^m$, $\|u\|$ represents $\sup_{t\geqslant 0} |u(t)|$.

2 Problem formulation and preliminaries

2.1 Problem formulation

Consider a class of multi-agent systems whose dynamics can be described by

$$\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + D_i v(t), \tag{1}$$

$$e_i(t) = C_i x_i(t) + F v(t), \tag{2}$$

$$y_i(t) = C_i x_i(t), \quad i \in \mathcal{N}, \tag{3}$$

where the set $\mathcal{N} = \{1, 2, \ldots, N\}$ with $N \in \mathbb{N}_+$. Signals $y_i(t) \in \mathbb{R}$, $u_i(t) \in \mathbb{R}$, $x_i(t) \in \mathbb{R}^{n_i}$ and $e_i(t) \in \mathbb{R}$ represent the output, the input, the state, and the tracking error of the agent i. The dimensions of system matrices are $A \in \mathbb{R}^{n_i\times n_i}$, $B_i \in \mathbb{R}^{n_i}$, $C_i \in \mathbb{R}^{1\times n_i}$, $D_i \in \mathbb{R}^{n_i\times q}$ and $F \in \mathbb{R}^{1\times q}$, for $i \in \mathcal{N}$. The signal $v(t) \in \mathbb{R}^q$ is the state of the following exosystem:

$$\dot{v}(t) = Sv(t), \tag{4}$$

where $S \in \mathbb{R}^{q\times q}$.

Define an acyclic digraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ with respect to systems (1)–(4), where $\mathcal{V} = \{0, \mathcal{N}\}$ is a set of nodes with the node 0 representing the leader (agent 0) modeled via the exosystem (4) and other nodes being followers characterized by systems (1) and (2). $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is a set of edges. The adjacency matrix is $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{(N+1)\times(N+1)}$ wherein its element $a_{ij} > 0$ if $(j, i) \in \mathcal{E}$ and otherwise $a_{ij} = 0$. The diagonal element of $\mathcal{A}$ satisfies $a_{ii} = 0$ for any $i \in \mathcal{V}$. $\mathcal{N}_i$ defines a set of neighbors of the ith agent, and $\mathcal{N}_i^+ := \mathcal{N}_i/\{0\}$.

In this paper, the following mild assumptions are made on systems (1)–(4), which have been introduced in the existing studies on cooperative output regulation problems; see [1, 24, 25].

Assumption 1. The pair (A_i, B_i) is stabilizable, $\forall i \in \mathcal{N}$.

Assumption 2. $\mathrm{rank}\left[\begin{smallmatrix} A_i - \lambda I & B_i \\ C_i & 0 \end{smallmatrix}\right] = n_i + 1$, $\forall \lambda \in \sigma(S)$, $i \in \mathcal{N}$.

Assumption 3 ([24])**.** The node 0 of graph $\mathcal{G}$ is globally reachable.

Given the systems (1)–(4) and the graph $\mathcal{G}$, the cooperative output regulation problem [1] is solved if one can develop a distributed control policy such that (i) the systems (1) and (2) is asymptotically stable when $v \equiv 0$, and (ii) the tracking error of all the followers asymptotically converges to zero ($\lim_{t\to\infty} e_i(t) = 0, \forall i \in \mathcal{N}$) with respect to any initial conditions $x_i(0)$ and $v(0)$.

2.2 Internal model principle

Select a vector $G \in \mathbb{R}^q$ such that the pair (S, G) is controllable; then the following equation

$$\dot{z}_i(t) = Sz_i(t) + G\hat{e}_i(t), \quad i \in \mathcal{N}$$

formulates a distributed internal model of the multi-agent systems (1)–(4), where $z_i \in \mathbb{R}^q$ and

$$\hat{e}_i(t) = \sum_{j\in\mathcal{N}_i} \frac{a_{ij}(y_i(t) - y_j(t))}{\sum_{j=0}^{N} a_{ij}}, \quad i \in \mathcal{N}, \tag{5}$$

and the reference signal is $y_0(t) = -Fv(t)$. The internal model principle discloses the fact that the output regulation problem can be converted to a stabilization problem for an augmented system. The next lemma shows that this fact holds for the cooperative output regulation of multi-agent systems as well.

Lemma 1. Under Assumptions 1–3, the multi-agent systems (1)–(4) in closed-loop with (5) and a distributed controller in the form of

$$u_i(t) = -K_{xi}x_i(t) - K_{zi}z_i(t), \tag{6}$$
$$\dot{z}_i(t) = Sz_i(t) + G\hat{e}_i(t), \quad i \in \mathcal{N}, \tag{7}$$

achieve cooperative output regulation if the matrix

$$A_{ci} = \begin{bmatrix} A_i - B_iK_{xi} & -B_iK_{zi} \\ GC_i & S \end{bmatrix}, \quad i \in \mathcal{N},$$

is Hurwitz.

Proof. From [26, Lemma 1.27], there exists uniquely a triple (X_i, U_i, Z_i) solving the following equations under Assumptions 1 and 2:

$$X_iS = A_iX_i + B_iU_i + D_i, \tag{8}$$
$$X_iS = (A_i - BK_x)X_i - B_iK_{zi}Z_i + D_i, \tag{9}$$
$$Z_iS = SZ_i, \tag{10}$$
$$0 = C_iX_i + F, \quad i \in \mathcal{N}, \tag{11}$$

which indicates that

$$U_i = -K_{xi}X_i - K_{zi}Z_i.$$

Through the following definitions:

$$\begin{aligned}
&\tilde{x}_i = x_i - X_i v,\ \tilde{z}_i = z_i - Z_i v,\ \tilde{u}_i = u_i - U_i v,\\
&K_i = \begin{bmatrix} K_{xi} & K_{zi} \end{bmatrix},\ \tilde{\xi}_i = \begin{bmatrix} \tilde{x}_i^{\mathrm{T}} & \tilde{z}_i^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}},\ \bar{C}_i = \begin{bmatrix} C_i & 0_{1\times q} \end{bmatrix},\\
&\bar{A}_i = \begin{bmatrix} A_i & 0_{n_i\times q} \\ GC_i & S \end{bmatrix},\ \bar{B}_i = \begin{bmatrix} B_i \\ 0_{q\times 1} \end{bmatrix},
\end{aligned}$$

one can rewrite the closed-loop systems (1)–(7) as

$$\begin{aligned}
\dot{\tilde{x}}_i &= A_i x_i + B_i u_i + D_i v - X_i S v\\
&= A_i x_i + B_i(-K_{xi}x_i - K_{zi}z_i) + D_i v - X_i S v\\
&= (A_i - B_i K_{xi})x_i - B_i K_{zi} z_i + D_i v - X_i S v\\
&= (A_i - B_i K_{xi})x_i - B_i K_{zi} z_i + D_i v\\
&\quad - ((A_i - B_i K_{xi})X_i v - B_i K_{zi} Z_i v + D_i v)\\
&= (A_i - B_i K_{xi})\tilde{x}_i - B_i K_{zi}\tilde{z}_i, \qquad (12)\\
\dot{\tilde{z}}_i &= S z_i + G\hat{e}_i - Z_i S v\\
&= S z_i + \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}(y_i - y_j)}{\sum_{j=0}^{N} a_{ij}} - Z_i S v\\
&= S z_i + GC_i x_i - \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}C_j x_j}{\sum_{j=0}^{N} a_{ij}} - Z_i S v\\
&= S z_i + GC_i x_i - \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}C_j x_j}{\sum_{j=0}^{N} a_{ij}} - S Z_i v\\
&= S\tilde{z}_i + G(C_i\tilde{x}_i - Fv) - \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}(C_j\tilde{x}_j - Fv)}{\sum_{j=0}^{N} a_{ij}}\\
&= S\tilde{z}_i + GC_i\tilde{x}_i - \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}C_j\tilde{x}_j}{\sum_{j=0}^{N} a_{ij}}, \qquad (13)
\end{aligned}$$

where $x_0 = v$, $\tilde{x}_0 = 0$, and $C_0 = -F$.

We can convert (13) into a more compact form,

$$\begin{aligned}
&\dot{\tilde{\xi}}_i = A_{ci}\tilde{\xi}_i - \sum_{j\in\mathcal{N}_i} \frac{a_{ij}}{\sum_{j=0}^{N} a_{ij}} \begin{bmatrix} 0 \\ G\bar{C}_j \end{bmatrix} \tilde{\xi}_j,\\
&e_i = \bar{C}_i\tilde{\xi}_i, \quad i\in\mathcal{N}, \qquad (14)
\end{aligned}$$

where $A_{ci} = \bar{A}_i - \bar{B}_i K_i$.

Based on Assumption 3, we can always label all the followers such that $i < j$ if $(i, j) \in \mathcal{E}$. In this way, one can represent the overall multi-agent systems via

$$\dot{\tilde{\xi}} = A_c\tilde{\xi}, \qquad (15)$$

where

$$\tilde{\xi} = [\tilde{\xi}_1^{\mathrm{T}}, \tilde{\xi}_2^{\mathrm{T}}, \ldots, \tilde{\xi}_N^{\mathrm{T}}]^{\mathrm{T}},$$

and A_c is a block lower-triangular matrix with submatrices A_{ci} on the diagonal, for any $i \in \mathcal{N}$. All nonzero a_{ij} appear in the lower left of the matrix A_c. Because A_c is Hurwitz, we conclude the system (15) is asymptotically stable at the origin, implying that $\lim_{t\to\infty}\tilde{\xi}(t) = 0$ and $\lim_{t\to\infty} e_i(t) = 0$ for any $i \in \mathcal{N}$. The proof is thus completed.

Remark 1. Compared with existing studies [1, 16], the distributed controller (6)–(7) does not depend on any observers of the exostate, which is suitable for the case that only the reference but not the full exostate is accessible by some followers.

2.3 Cooperative optimal output regulation

Define the following vectors and matrices to lump the control inputs and system matrices:

$$\begin{aligned}&\tilde{u} = [\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_N]^{\mathrm{T}},\\ &\bar{A} = \mathrm{blockdiag}(\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_N),\\ &\bar{B} = \mathrm{blockdiag}(\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_N),\\ &Q = \mathrm{blockdiag}(Q_1, Q_2, \ldots, Q_N),\end{aligned}$$

where, for $i \in \mathcal{N}$, $\tilde{u}_i = u_i - U_i v$, $Q_i = Q_i^{\mathrm{T}} \succ 0$.

If all the followers are neighbors of the leader, i.e., $\mathcal{N}_0 = \mathcal{N}$, one can develop a decentralized internal model as follows:

$$\dot{z}_i(t) = Sz_i(t) + Ge_i(t), \quad i \in \mathcal{N}. \tag{16}$$

Multi-agent systems (1)–(4) along with decentralized internal model (16) can be converted to

$$\dot{\tilde{\xi}} = \bar{A}\tilde{\xi} + \bar{B}\tilde{u}. \tag{17}$$

Besides the steady-state tracking performance, the cooperative optimal output regulation problem is formulated as follows by taking the transient performance of the closed-loop system into consideration.

Problem 1. The cooperative optimal output regulation problem is solved if one can develop a state-feedback controller in the form of (5)–(7) such that the multi-agent systems (1)–(4) achieve cooperative output regulation. Moreover, the zero-error constrained control $\tilde{u}$ minimizes the cost functional (18) subject to (17).

$$\begin{aligned}&\min_{\tilde{u}} \int_0^{\infty} \left(\tilde{\xi}^{\mathrm{T}} Q \tilde{\xi} + \tilde{u}^{\mathrm{T}} \tilde{u}\right) \mathrm{d}t\\ &\text{subject to (17).}\end{aligned} \tag{18}$$

Based on the linear optimal control theory, the optimal control policy that minimizes the cost (18) subject to (17) is

$$\tilde{u}^* = -K^* \tilde{\xi}. \tag{19}$$

Since $\bar{A}, \bar{B}$ and Q are block diagonal matrices, we can observe that $K^* = \mathrm{blockdiag}(K_1^*, K_2^*, \ldots, K_N^*)$, where the optimal control gain for the ith follower is

$$K_i^* = \bar{B}_i^{\mathrm{T}} P_i^* := \begin{bmatrix} K_{xi}^* & K_{zi}^* \end{bmatrix}. \tag{20}$$

The matrix $P_i^* \succ 0$ is the unique solution to the algebraic Riccati equation as below:

$$\bar{A}_i^{\mathrm{T}} P_i^* + P_i^* \bar{A}_i + Q_i - P_i^* \bar{B}_i \bar{B}_i^{\mathrm{T}} P_i^* = 0. \tag{21}$$

Furthermore, Eq. (19) is equivalent to

$$\begin{aligned}u_i^* &= \tilde{u}_i^* + U_i v\\ &= -K_i^* \tilde{\xi}_i + U_i v\\ &= -K_{xi}^* \tilde{x}_i - K_{zi}^* \tilde{z}_i + U_i v\\ &= -K_{xi}^* x_i - K_{zi}^* z_i,\\ \dot{z}_i(t) &= Sz_i(t) + G\hat{e}_i(t), \quad i \in \mathcal{N}.\end{aligned} \tag{22}$$

As the pair $(\bar{A}_i, \bar{B}_i)$ is stabilizable [26], it is easy to see that the matrix

$$\begin{bmatrix} A_i - B_i K_{xi}^* & -B_i K_{zi}^* \\ GC_i & S \end{bmatrix}$$

is Hurwitz for all $i \in \mathcal{N}$. Based on Lemma 1, the developed controller (22) can be used to solve Problem 1.

3 Main results

In this section, we will first introduce the DDoS attacks that we consider in this paper. Then, we will develop a learning-based control approach to approximate the optimal control policy despite the presence of DDoS attacks. Finally, we will explore the a priori knowledge on the upper bound of DDoS attacks to guarantee that the closed-loop system remains operational. Note that both the learning-based controller design and the exploration do not rely on the knowledge of any system matrices in the state equation (1).

3.1 DDoS attacks

In this paper, we consider the case that DDoS attackers use multiple computers to launch coordinated DoS attacks to prevent targeted agents from exchanging information with their neighbors. In the worst case, all the agents in the network could lose their communications during some intervals, which may affect the cybersecurity of multi-agent systems more significantly compared with traditional DoS attacks.

For any $i \in \mathcal{N}$, an agent i is said to be isolated if this agent cannot receive information from any other agents in the whole network, i.e, $\sum_{j=0}^{N} a_{ij} = 0$. Let $\mathcal{I}_s^i = [h_s^i, h_s^i + \tau_s^i)$ represent the interval of the s-th ($s \in \mathbb{N}_+$) launched DDoS attack such that the agent i is isolated. Time instants h_s^i, $h_s^i + \tau_s^i$, and τ_s^i refer to the start time, the end time, and the duration of the s-th DDoS attack for the agent i.

The following assumptions are made with respect to the DDoS frequency and DDoS duration.

Assumption 4 (DDoS frequency)**.** There exist constants $\eta_i > 1$ and $\tau_D^i > 0$ such that

$$n_i(t_a, t_b) \leqslant \eta_i + \frac{t_b - t_a}{\tau_D^i},\ \forall\ t_b > t_a \geqslant 0,\ i \in \mathcal{N}, \tag{23}$$

where $n_i(t_a, t_b)$ denotes the number of DDoS off/on transitions occurring on the interval $[t_a, t_b]$.

Assumption 5 (DDoS duration)**.** There exist constants $T_i > 1$ and $\kappa_i > 0$ such that

$$\left|\Pi_D^i(t_a, t_b)\right| \leqslant \kappa_i + \frac{t_b - t_a}{T_i},\ \forall\ t_b > t_a \geqslant 0,\ i \in \mathcal{N}, \tag{24}$$

where $\Pi_D^i(t_a, t_b) := (t_a, t_b) \bigcap \bigcup_{s=1}^{\infty} \mathcal{I}_s$ is the time set when the agent i is isolated due to DDoS attacks during $[t_a, t_b]$, and $|\Pi_D^i(t_a, t_b)|$ is the Lebesgue measure of the set $\Pi_D^i(t_a, t_b)$.

Moreover, for any $i \in \mathcal{N}$, we use $\Pi_N^i(t_a, t_b) := [t_a, t_b] \setminus \Pi_D^i(t_a, t_b)$ to denote the set when the agent i can receive information from at least one of its neighbors. It is checkable that, due to the consideration of DDoS attacks, the digraph $\mathcal{G}$ and its adjacency matrix $\mathcal{A}$ of the multi-agent systems (1)–(4) will be time-varying. We define them as $\mathcal{G}(t)$ and $\mathcal{A}(t) := [a_{ij}(t)]$, respectively.

Remark 2. Assumptions 4 and 5 are similar to existing ones of (distributed) DoS attacks that have been made in [21, 22, 27] to characterize the frequencies and duration of attacks. However, since our assumptions quantify the isolation of agents (nodes), different from [21, 22], Assumptions 4 and 5 allow some channels in the network to be attacked at all time.

3.2 Value iteration

Value iteration (VI), a non-model-based ADP approach, has been applied to approximate the optimal control policy even if an admissible control policy is unavailable [28–30]. Given the DDoS attacks considered in this paper, we will develop a VI strategy to learn the optimal values P_i^* and K_i^*.

Essentially, VI updates the value matrix and control gain by

$$\begin{aligned} P_i^{(k+1)} &= \epsilon_k \left(\bar{A}_i^{\mathrm{T}} P_i^{(k)} + P_i^{(k)} \bar{A}_i - P_i^{(k)} \bar{B}_i B_i^{\mathrm{T}} P_i^{(k)} + Q_i\right) + P_i^{(k)}, \\ K_i^{(k)} &= \bar{B}_i^{\mathrm{T}} P_i^{(k)}, \quad i \in \mathcal{N}, \end{aligned} \tag{25}$$

where ϵ_k is the step size satisfying

$$\epsilon_k > 0, \quad \sum_{k=0}^{\infty} \epsilon_k = \infty, \quad \sum_{k=0}^{\infty} \epsilon_k^2 < \infty.$$

It has been shown in [28] that the sequences $\{P_i^{(k)}\}_{k=0}^{\infty}$ and $\{K_i^{(k)}\}_{k=0}^{\infty}$ converge to the optimal values P_i^* and K_i^*. However, Eq. (25) depends on the knowledge of system matrices $\bar{A}_i$ and $\bar{B}_i$. In this paper, we will focus on learning these optimal values in terms of online state and input data.

Note that if the agent i is under DDoS attacks such that it is isolated at t, one can observe that $\sum_{j=0}^{N} a_{ij}(t) = 0$. We see from (5) that $\hat{e}_i(t)$ does not make sense at t since its denominator equals zero, which renders the control policies (6) and (22) depending on the signal $\hat{e}_i(t)$ to be no longer applicable in the presence of DDoS attacks. To avoid this contradiction, we modify the distributed internal model (7) by

$$\dot{z}_i(t) = Sz_i(t) + G\bar{e}_i(t), \quad i \in \mathcal{N}, \tag{26}$$

where

$$\bar{e}_i(t) = \begin{cases} \sum_{j\in\mathcal{N}_i} \frac{a_{ij}(t)(y_i(t)-y_j(t))}{\sum_{j=0}^{N} a_{ij}(t)}, & t \in \Pi_N^i(0,\infty), \\ 0, & t \in \Pi_D^i(0,\infty). \end{cases}$$

The modified internal model (26) can be rewritten as

$$\begin{aligned} \dot{z}_i &= Sz_i + G\bar{e}_i \\ &= Sz_i + \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}(t)(y_i - y_j)}{\sum_{j=0}^{N} a_{ij}(t)}, \\ &= Sz_i + GC_ix_i - \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}C_jx_j}{\sum_{j=0}^{N} a_{ij}(t)}, \quad t \in \Pi_N^i(0,\infty). \end{aligned} \tag{27}$$

By combining (1) and (27), we have

$$\begin{aligned} \dot{\xi}_i(t) &= \bar{A}_i\xi_i(t) + \bar{B}_iu_i(t) + \begin{bmatrix} D_i \\ \alpha_{i0}(t)GF \end{bmatrix} v(t) \\ &\quad - \sum_{j\in\mathcal{N}_i^+} \alpha_{ij}(t) \begin{bmatrix} 0 \\ G\bar{C}_j \end{bmatrix} \xi_j(t) \\ &= \bar{A}_i\xi_i(t) + \bar{B}_iu_i(t) + (\bar{D}_i + \Theta_{i0})(t)v(t) \\ &\quad + \sum_{j\in\mathcal{N}_i^+} \Theta_{ij}(t)\xi_j(t), \quad t \in \Pi_N^i(0,\infty), \end{aligned} \tag{28}$$

where

$$\begin{aligned} &\xi_i = \begin{bmatrix} x_i^{\mathrm{T}} & z_i^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \\ &\alpha_{ij}(t) = \frac{a_{ij}(t)}{\sum_{j=0}^{N} a_{ij}(t)}, \\ &\bar{D}_i = \begin{bmatrix} D_i \\ 0 \end{bmatrix}, \\ &\Theta_{i0}(t) = \begin{bmatrix} 0 \\ \alpha_{i0}(t)GF \end{bmatrix}, \\ &\Theta_{ij}(t) = \alpha_{ij}(t) \begin{bmatrix} 0 \\ G\bar{C}_j \end{bmatrix}, \quad i \in \mathcal{N},\ j \in \mathcal{N}_i,\ t \in \Pi_N^i(0,\infty). \end{aligned}$$

Select a quadratic function $W_i = \xi_i^{\mathrm{T}} P_i^{(k)} \xi_i$. By taking its derivative along the trajectories of system (28), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}(\xi_i^{\mathrm{T}} P_i^{(k)} \xi_i) = \left(\bar{A}_i\xi_i + \bar{B}_iu_i + (\bar{D}_i + \Theta_{i0})v - \sum_{j\in\mathcal{N}_i^+} \Theta_{ij}\xi_j \right)^{\mathrm{T}} P_i^{(k)} \xi_i$$

$$
\begin{aligned}
&+\xi_i^{\mathrm{T}}P_i^{(k)}\left(\bar{A}_i\xi_i+\bar{B}_iu_i+(\bar{D}_i+\Theta_{i0})v-\sum_{j\in\mathcal{N}_i^+}\Theta_{ij}\xi_j\right)\\
&=\xi_i^{\mathrm{T}}\mathcal{H}_i^{(k)}\xi_i+2u_iK_i^{(k)}\xi_i+2v^{\mathrm{T}}\bar{D}_i^{\mathrm{T}}P_i^{(k)}\xi_i+2v^{\mathrm{T}}\Theta_{i0}^{\mathrm{T}}P_i^{(k)}\xi_i\\
&\quad-2\sum_{j\in\mathcal{N}_i^+}\xi_j^{\mathrm{T}}\Theta_{ij}^{\mathrm{T}}P_i^{(k)}\xi_i,\quad i\in\mathcal{N},\ t\in\Pi_N^i(0,\infty),
\end{aligned}\tag{29}
$$

where $\mathcal{H}_i^{(k)}=\bar{A}_i^{\mathrm{T}}P_i^{(k)}+P_i^{(k)}\bar{A}_i$.

Based on Assumption 5, there always exists a sequence $\{t_l^i\}_{l=0}^{\infty}$ such that the agent i is not isolated in all the following intervals $[t_0^i,t_1^i],[t_2^i,t_3^i],[t_4^i,t_5^i],\ldots$. By integrating both sides of (29) during the interval $[t_0^i,t_1^i]$, we have

$$
\begin{aligned}
&\xi_i^{\mathrm{T}}(t_1^i)P_i^{(k)}\xi_i(t_1^i)-\xi_i^{\mathrm{T}}(t_0^i)P_i^{(k)}\xi_i(t_0^i)\\
&\quad=\int_{t_0^i}^{t_1^i}\xi_i^{\mathrm{T}}\mathcal{H}_i^{(k)}\xi_i+2u_iK_i^{(k)}\xi_i+2v^{\mathrm{T}}\bar{D}_i^{\mathrm{T}}P_i^{(k)}\xi_i\mathrm{d}\tau\\
&\qquad+\int_{t_0^i}^{t_1^i}2v^{\mathrm{T}}\Theta_{i0}^{\mathrm{T}}P_i^{(k)}\xi_i-2\sum_{j\in\mathcal{N}_i^+}\xi_j^{\mathrm{T}}\Theta_{ij}^{\mathrm{T}}P_i^{(k)}\xi_i\mathrm{d}\tau\\
&\quad=\left(\int_{t_0^i}^{t_1^i}\mathrm{vecv}(\xi_i(\tau))\mathrm{d}\tau\right)^{\mathrm{T}}\mathrm{vecs}\left(\mathcal{H}_i^{(k)}\right)\\
&\qquad+2\left(\int_{t_0^i}^{t_1^i}\xi_i^{\mathrm{T}}\otimes u_i\mathrm{d}\tau\right)\mathrm{vec}\left(K_i^{(k)}\right)\\
&\qquad+2\left(\int_{t_0^i}^{t_1^i}\xi_i^{\mathrm{T}}\otimes v^{\mathrm{T}}\mathrm{d}\tau\right)\mathrm{vec}\left(\bar{D}_i^{\mathrm{T}}P_i^{(k)}\right)\\
&\qquad+2\left(\int_{t_0^i}^{t_1^i}\xi_i^{\mathrm{T}}\otimes v^{\mathrm{T}}\mathrm{d}\tau\right)\mathrm{vec}\left(\Theta_{i0}^{\mathrm{T}}P_i^{(k)}\right)\\
&\qquad-2\sum_{j\in\mathcal{N}_i^+}\left(\int_{t_0^i}^{t_1^i}\xi_i^{\mathrm{T}}\otimes\xi_j^{\mathrm{T}}\mathrm{d}\tau\right)\mathrm{vec}\left(\Theta_{ij}^{\mathrm{T}}P_i^{(k)}\right).
\end{aligned}\tag{30}
$$

For any two vectors a,b and a sufficiently large number $L^i>0$, define

$$
\begin{aligned}
\delta_a^i&=\left[a\otimes a|_{t_0^i}^{t_1^i},a\otimes a|_{t_2^i}^{t_3^i},\ldots,a\otimes a|_{t_{2L^i}^i}^{t_{2L^i+1}^i}\right]^{\mathrm{T}},\\
\Gamma_{a,b}^i&=\left[\int_{t_0^i}^{t_1^i}a\otimes b\mathrm{d}\tau,\int_{t_2^i}^{t_3^i}a\otimes b\mathrm{d}\tau,\ldots,\int_{t_{2L^i}^i}^{t_{2L^i+1}^i}a\otimes b\mathrm{d}\tau\right]^{\mathrm{T}},\\
\Gamma_a^i&=\left[\int_{t_0^i}^{t_1^i}\mathrm{vecv}(a)\mathrm{d}\tau,\int_{t_2^i}^{t_3^i}\mathrm{vecv}(a)\mathrm{d}\tau,\ldots,\int_{t_{2L^i}^i}^{t_{2L^i+1}^i}\mathrm{vecv}(a)\mathrm{d}\tau\right]^{\mathrm{T}}.
\end{aligned}
$$

Eqs. (29) and (30) imply the following equation:

$$
\Psi_i\begin{bmatrix}\mathrm{vecs}(\mathcal{H}_i^{(k)})\\ \mathrm{vec}(K_i^{(k)})\\ \mathrm{vec}(\bar{D}_i^{\mathrm{T}}P_i^{(k)})\end{bmatrix}=\Phi_i^{(k)},\tag{31}
$$

where

$$
\Psi_i=[\Gamma_{\xi_i}^i,2\Gamma_{\xi_iu_i}^i,2\Gamma_{\xi_iv}^i],
$$

Algorithm 1 VI algorithm

1: Select a $c > 0$. Apply any locally essentially bounded input u_i on $[t_0, t_L]$ s.t. Eq. (32) holds $\forall i \in \mathcal{N}$;
2: $i \leftarrow 1$;
3: **repeat**
4: $k \leftarrow 0$, $p \leftarrow 0$. Pick a $P_i^{(0)} \succ 0$;
5: **loop**
6: Solve $\mathcal{H}_i^{(k)}$ and $K_i^{(k)}$ from (31);
7: $\tilde{P}_i^{(k+1)} \leftarrow P_i^{(k)} + \epsilon_k(\mathcal{H}_i^{(k)} + Q_i - (K_i^{(k)})^{\mathrm{T}} K_i^{(k)})$;
8: **if** $\tilde{P}_i^{(k+1)} \notin \mathcal{B}_p$ **then**
9: $P_i^{(k+1)} \leftarrow P_i^{(0)}$, $p \leftarrow p+1$;
10: **else if** $|P_i^{(k)} - P_i^{(k-1)}|/\epsilon_k < c$ **then**
11: **return** $(P_i^{(k)}, K_i^{(k)})$;
12: **else** $P_i^{(k+1)} \leftarrow \tilde{P}_i^{(k+1)}$;
13: **end if**
14: $k \leftarrow k+1$;
15: **end loop**
16: $i \leftarrow i+1$;
17: **until** $i = N+1$.

$$\Phi_i^{(k)} = \delta_{\xi_i}^i \mathrm{vec}\left(P_i^{(k)}\right) - 2\Gamma_{\xi_i v}^i \mathrm{vec}\left(\Theta_{i0}^{\mathrm{T}} P_i^{(k)}\right) + 2\sum_{j\in\mathcal{N}_i^+} \Gamma_{\xi_i\xi_j}^i \left(\Theta_{ij}^{\mathrm{T}} P_i^{(k)}\right).$$

The uniqueness of the solution to (31) is guaranteed under some rank conditions as shown in Lemma 2 below. We omit the proof as it is similar to the proofs in [28, 31].

Lemma 2. If there exists a $L \in \mathbb{Z}_+$ such that for all $L^i > L$, $i \in \mathcal{N}$,

$$\mathrm{rank}([\Gamma_{\xi_i,\xi_i}^i, \Gamma_{\xi_i,u_i}^i, \Gamma_{\xi_i,v}^i]) = \frac{(n+q)(n+3q+3)}{2}, \tag{32}$$

then the matrix Ψ_i has full column rank.

By defining a collection of bounded sets $\{\mathcal{B}_p\}_{p=0}^{\infty}$ as follows:

$$\mathcal{B}_p \subset \mathcal{B}_{p+1}, \quad p \in \mathbb{Z}_+, \quad \lim_{p\to\infty} \mathcal{B}_p = \mathcal{P}_n.$$

The VI algorithm is proposed in Algorithm 1. From [28, Theorem 5.1], under the rank condition (32), one can show that $\lim_{k\to\infty} P_i^{(k)} = P_i^*$ and $\lim_{k\to\infty} K_i^{(k)} = P_i^*$ where sequences $\{P_i^{(k)}\}_{k=0}^{\infty}$ and $\{K_i^{(k)}\}_{k=0}^{\infty}$ are obtained from Algorithm 1, for any $i \in \mathcal{N}$.

Remark 3. By applying Algorithm 1, one can learn towards the control policy (19), which is optimal with respect to the cost (18) if $\mathcal{N}_0 = \mathcal{N}$ and the multi-agent systems are attack-free. The learned control policy will be suboptimal under communication constraints and DDoS attacks described in Assumptions 3–5.

3.3 Resilience analysis under DDoS attacks

Considering the effect of DDoS attacks, the control input and internal model applied to the process can be expressed as

$$u_i(t) = -K_i^* \xi_i(t), \tag{33}$$
$$\dot{z}_i(t) = S z_i(t) + G\bar{e}_i(t). \tag{34}$$

By observing the system (1) in closed-loop with the controller (33) and internal model (34), Eq. (12) holds. Moreover, when the agent i is not isolated, we have

$$\begin{aligned}\dot{\tilde{z}}_i &= S z_i + G\hat{e}_i - Z_i S v \\ &= S z_i + \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}(y_i - y_j)}{\sum_{j=0}^N a_{ij}} - Z_i S v\end{aligned}$$

$$
\begin{aligned}
&= Sz_i + GC_ix_i - \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}C_jx_j}{\sum_{j=0}^{N} a_{ij}} - Z_iSv\\
&= Sz_i + GC_ix_i - \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}C_jx_j}{\sum_{j=0}^{N} a_{ij}} - SZ_iv\\
&= S\tilde{z}_i + GC_i\tilde{x}_i - \sum_{j\in\mathcal{N}_i} G\frac{a_{ij}C_j\tilde{x}_j}{\sum_{j=0}^{N} a_{ij}}, \quad t\in\Pi_N^i(0,\infty).
\end{aligned} \tag{35}
$$

When the agent i is under a DDoS attack such that it is isolated, we have

$$
\begin{aligned}
\dot{\tilde{z}}_i &= Sz_i - Z_iSv\\
&= Sz_i - SZ_iv\\
&= S\tilde{z}_i, \quad t\in\Pi_D^i(0,\infty).
\end{aligned} \tag{36}
$$

One can rewrite (12), (35)–(36) in a more compact form:

$$
\begin{aligned}
\dot{\tilde{\xi}}_i(t) &= A_{ci}\tilde{\xi}_i(t) + \bar{G}\epsilon_i,\\
e_i(t) &= \bar{C}_i\tilde{\xi}_i(t), \quad i\in\mathcal{N},
\end{aligned} \tag{37}
$$

where

$$
\begin{aligned}
\epsilon_i(t) &= \begin{cases} \epsilon_{Ni}, & t\in\Pi_N^i(0,\infty),\\ -\bar{C}_i\tilde{\xi}_i, & t\in\Pi_D^i(0,\infty),\end{cases}\\
\epsilon_{Ni}(t) &= \begin{cases} -\sum_{j\in\mathcal{N}_i}\frac{a_{ij}(t)(\bar{C}_j\tilde{\xi}_j)}{\sum_{j=0}^{N}a_{ij}(t)}, & t\in\Pi_N^i(0,\infty),\\ 0, & t\in\Pi_D^i(0,\infty),\end{cases}\\
\bar{G} &= \begin{bmatrix} 0_{n\times 1}\\ G\end{bmatrix}.
\end{aligned} \tag{38}
$$

Before analyzing the resilience, we give the following Lemma to explore the output regulation property of multi-agent systems if each agent is input-to-state stable (ISS) [32].

Lemma 3. Under Assumption 3, if, for any $i\in\mathcal{N}$, the system (37) is ISS regarding ϵ_{Ni} defined in (38) as an input, then the multi-agent systems (1)–(4) with (33)–(34) achieve cooperative output regulation.

Proof. Based on Assumption 3, we label all the followers such that $j\notin\mathcal{N}_i$ for any integer $j\geqslant i$. We will prove this lemma by induction.

(1) When $i=1$, we have $\epsilon_{N1}\equiv 0$. If the system (37) is input-to-state stable regarding ϵ_{N1}, then there exist a function σ_1 of class $\mathcal{KL}$ such that

$$
|\tilde{\xi}_1(t)| \leqslant \sigma_1(|\tilde{\xi}_1(0)|, t). \tag{39}
$$

(2) When $i=j>1$, suppose we have

$$
|\zeta_j(t)| \leqslant \sigma_2(|\zeta_j(0)|, t), \tag{40}
$$

where σ_2 is a function of class $\mathcal{KL}$, and

$$
\zeta_j = \left[\tilde{\xi}_1^{\mathrm{T}}, \ldots, \tilde{\xi}_j^{\mathrm{T}}\right]^{\mathrm{T}}.
$$

The input-to-state stability of the system (37) for $i=j+1$ implies that

$$
|\tilde{\xi}_{j+1}(t)| \leqslant \sigma_3(|\tilde{\xi}_{j+1}(0)|, t) + \gamma_1(\|\epsilon_{N,j+1}\|),
$$

where σ_3 is a function of class $\mathcal{KL}$ and γ_1 is a function of class $\mathcal{K}$. From (38), there always exists a function γ_2 of class $\mathcal{K}$ such that $\|\epsilon_{N,j+1}\|\leqslant\gamma_2(\|\zeta_j\|)$. Thus, we have

$$
|\tilde{\xi}_{j+1}(t)| \leqslant \sigma_3(|\tilde{\xi}_{j+1}(0)|, t) + \gamma_3(\|\zeta_j\|), \tag{41}
$$

where $\gamma_3 = \gamma_1 \circ \gamma_2$.

From [33, Lemma 4.7], we have the following inequality:

$$|\zeta_{j+1}(t)| \leqslant \sigma_4(|\zeta_{j+1}(0)|, t),$$

where σ_4 is a function of class $\mathcal{KL}$ defined as follows:

$$\begin{aligned}\sigma_4(r,t) = \sigma_3\left(\sigma_3\left(r, \frac{t}{2}\right) + \gamma_3(\sigma_2(r,0)), \frac{t}{2}\right)\\ + \gamma_3\left(\sigma_2\left(r, \frac{t}{2}\right)\right) + \sigma_2(r,t).\end{aligned}$$

Therefore, we conclude that the multi-agent system (37) with all $i \in \mathcal{N}$ is asymptotically stable at the origin. And the tracking error $e_i(t)$ asymptotically converges to zero as t goes to infinity. This is sufficient to show that the cooperative output regulation of multi-agent systems (1)–(4) with (33)–(34) has achieved. The proof is thus completed.

We develop the following theorem to seek a bound of DDoS duration parameter T_i to ensure the cooperative output regulation under DDoS attacks.

Theorem 1. Under Assumptions 1–5, the multi-agent systems (1)–(4) in closed-loop with the controller (33) and internal model (34) under DDoS attacks achieve cooperative output regulation if the DDoS duration criterion T_i satisfies

$$T_i > 1 + \frac{\lambda_M(P_i^*)\max\left\{0, 2|P_i^*\bar{G}\bar{C}_i| - \lambda_m(Q_i + (K_i^*)^{\mathrm{T}}K_i^*)\right\}}{\lambda_m(P_i^*)\lambda_m(Q_i + (K_i^*)^{\mathrm{T}}K_i^*)} := T_i^*, \quad \forall i \in \mathcal{N}. \tag{42}$$

Proof. By taking $V_i = \tilde{\xi}_i^{\mathrm{T}} P_i^* \tilde{\xi}_i$ as a Lyapunov function, along the closed-loop system (37), we have that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V_i &= \tilde{\xi}_i^{\mathrm{T}}(A_{ci}^{\mathrm{T}}P_i^* + P_i^*A_{ci})\tilde{\xi}_i + 2\tilde{\xi}_i^{\mathrm{T}}P_i^*\bar{G}\epsilon_i\\ &= -\tilde{\xi}_i^{\mathrm{T}}(Q_i + (K_i^*)^{\mathrm{T}}K_i^*)\tilde{\xi}_i + 2\tilde{\xi}_i^{\mathrm{T}}P_i^*\bar{G}\epsilon_i.\end{aligned} \tag{43}$$

Considering the internal $t \in [h_s^i + \tau_s^i, h_{s+1}^i)$ where the agent i is not isolated and $\epsilon_i(t) = \epsilon_{Ni}(t)$. By Young's inequality the fact that $\lambda_m(P_i^*)|\tilde{\xi}_i|^2 \leqslant V_i \leqslant \lambda_M(P_i^*)|\tilde{\xi}_i|^2$, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V_i &\leqslant -\tilde{\xi}_i^{\mathrm{T}}(Q_i + (K_i^*)^{\mathrm{T}}K_i^*)\tilde{\xi}_i + \frac{1}{\delta_i}\tilde{\xi}_i^{\mathrm{T}}P_i^*\bar{G}\bar{G}^{\mathrm{T}}P_i^*\tilde{\xi}_i + \delta_i\epsilon_{Ni}^{\mathrm{T}}\epsilon_{Ni}\\ &\leqslant -\lambda_m(Q_i + (K_i^*)^{\mathrm{T}}K_i^*)|\tilde{\xi}_i|^2 + \frac{1}{\delta_i}|P_i^*\bar{G}|^2|\tilde{\xi}_i|^2 + \delta_i|\epsilon_{Ni}|^2\\ &\leqslant -\frac{\lambda_m(Q_i + (K_i^*)^{\mathrm{T}}K_i^*) - \frac{1}{\delta_i}|P_i^*\bar{G}|^2}{\lambda_M(P_i^*)}V_i + \delta_i|\epsilon_{Ni}|^2\end{aligned} \tag{44}$$

for any $\delta_i > \frac{|P_i^*\bar{G}|^2}{\lambda_m(Q_i + (K_i^*)^{\mathrm{T}}K_i^*)}$.

One can further have

$$V_i(\tilde{\xi}_i(t)) \leqslant \mathrm{e}^{-w_{i1}(t - h_s^i - \tau_s^i)}V_i(\tilde{\xi}_i(h_s^i + \tau_s^i)) + \frac{\delta_i}{w_{i1}}|\epsilon_{Ni}|^2, \tag{45}$$

where

$$w_{i1} = \frac{\lambda_m(Q_i + (K_i^*)^{\mathrm{T}}K_i^*) - \frac{1}{\delta_i}|P_i^*\bar{G}|^2}{\lambda_M(P_i^*)}.$$

During the interval $[h_s^i, h_s^i + \tau_s^i)$ that the agent i is isolated, we derive from (43) that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V_i &= -\tilde{\xi}_i^{\mathrm{T}}(Q_i + (K_i^*)^{\mathrm{T}}K_i^*)\tilde{\xi}_i + 2\tilde{\xi}_i^{\mathrm{T}}P_i^*\bar{G}\bar{C}_i\tilde{\xi}_i\\ &\leqslant \tilde{\xi}_i^{\mathrm{T}}\left(2|P_i^*\bar{G}\bar{C}_i| - \lambda_m(Q_i + (K_i^*)^{\mathrm{T}}K_i^*)\right)\tilde{\xi}_i\end{aligned}$$

$$\leqslant w_{i2}V_i, \tag{46}$$

where

$$w_{i2}=\frac{\max\left\{0,2|P_i^*\bar{G}\bar{C}_i|-\lambda_m(Q_i+(K_i^*)^{\mathrm{T}}K_i^*)\right\}}{\lambda_m(P_i^*)}.$$

Then, for any $t\in[h_s^i,h_s^i+\tau_s^i)$, we have

$$V_i(\tilde{\xi}_i(t))\leqslant \mathrm{e}^{w_{i2}(t-h_s^i)}V_i(\tilde{\xi}_i(h_s^i)). \tag{47}$$

By [27, Lemma 3], for all $t\geqslant 0$, the Lyapunov function satisfies

$$\begin{aligned}V_i(\tilde{\xi}_i(t))\leqslant{}&\mathrm{e}^{-w_{i1}|\Pi_N^i(0,t)|}\mathrm{e}^{w_{i2}|\Pi_D^i(0,t)|}V_i(\tilde{\xi}_i(0))\\&+\frac{\delta_i}{w_{i1}}\left[1+2\sum_{s\in\mathbb{N}_+;h_s\leqslant t}\mathrm{e}^{-w_{i1}|\Pi_N^i(h_s^i+\tau_s^i,t)|}\mathrm{e}^{w_{i2}|\Pi_D^i(h_s^i,t)|}\right]\|\epsilon_{Ni}\|^2.\end{aligned} \tag{48}$$

Based on Assumption 5, we have $|\Pi_D^i(h_s^i,t)|\leqslant\kappa_i+\frac{t-h_s^i}{T_i},\forall t\geqslant h_s^i$. We further have $|\Pi_N^i(h_s^i+\tau_s^i,t)|=t-h_s^i-|\Pi_D^i(h_s^i,t)|,\forall t\geqslant h_s^i$. Therefore, one can observe that

$$\sum_{s\in\mathbb{N}_+;h_s\leqslant t}\mathrm{e}^{-w_{i1}|\Pi_N^i(h_s^i+\tau_s^i,t)|}\mathrm{e}^{w_{i2}|\Pi_D^i(h_s^i,t)|}\leqslant\mathrm{e}^{w_{i3}\kappa_i}\sum_{s\in\mathbb{N}_+;h_s^i\leqslant t}\mathrm{e}^{-\beta(t-h_s^i)}, \tag{49}$$

where

$$\begin{aligned}w_{i3}&=w_{i1}+w_{i2}\\&=\frac{\lambda_m(Q_i+(K_i^*)^{\mathrm{T}}K_i^*-\frac{1}{\delta_i}|P_i^*\bar{G}|^2)\lambda_m(P_i^*)}{\lambda_M(P^*)\lambda_m(P^*)}\\&\quad+\frac{\lambda_M(P^*)|2|P_i^*\bar{G}\bar{C}_i|-\lambda_m(Q_i+(K_i^*)^{\mathrm{T}}K_i^*)|}{\lambda_M(P^*)\lambda_m(P^*)},\\\beta&=w_{i1}-\frac{w_{i3}}{T_i}.\end{aligned}$$

Note that T_i^* defined in (42) is exactly $T_i^*=w_{i3}/w_{i1}$. By [27, Lemma 4] and Assumption 4, we have

$$\sum_{s\in\mathbb{N}_+;h_s^i\leqslant t}\mathrm{e}^{-\beta(t-h_s^i)}\leqslant\frac{\mathrm{e}^{-\beta\tau_D^i}\eta_i}{1-\mathrm{e}^{-\beta\tau_D^i}}.$$

Finally, we have the Lyapunov function V_i along the trajectory of the closed-loop system satisfying

$$\begin{aligned}V_i(\tilde{\xi}_i(t))\leqslant{}&\mathrm{e}^{\kappa_i w_{i3}-\beta t}V_i(\tilde{\xi}_i(0))\\&+\gamma_3\left(1+2\mathrm{e}^{\kappa_i w_{i3}}\frac{\mathrm{e}^{\beta\tau_D^i}\eta_i}{1-\mathrm{e}^{-\beta\tau_D^i}}\right)\|\epsilon_{Ni}\|^2.\end{aligned} \tag{50}$$

which immediately indicates that the existence of a function σ_5 of class $\mathcal{KL}$, and a function $\gamma_4>0$ of class $\mathcal{K}$ such that

$$|\tilde{\xi}_i(t)|\leqslant\sigma_5(|\tilde{\xi}_i(0)|,t)+\gamma_4(\|\epsilon_{Ni}\|),\quad i\in\mathcal{N}.$$

From (50), we see that a sufficient condition to ensure the system (37) to be ISS is letting the DDoS duration criterion T_i satisfy the inequality (42) such that $\beta>0$. Moreover, based on Lemma 3, we conclude that multi-agent systems (1)–(4) in closed-loop with the controller (33) and internal model (34) under DDoS attacks achieve cooperative output regulation. The proof is thus completed.

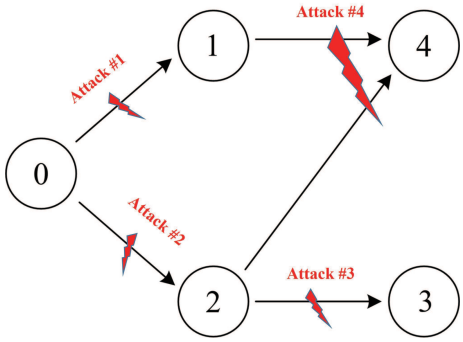

Figure 1 (Color online) Communication topology and DDoS attacks.

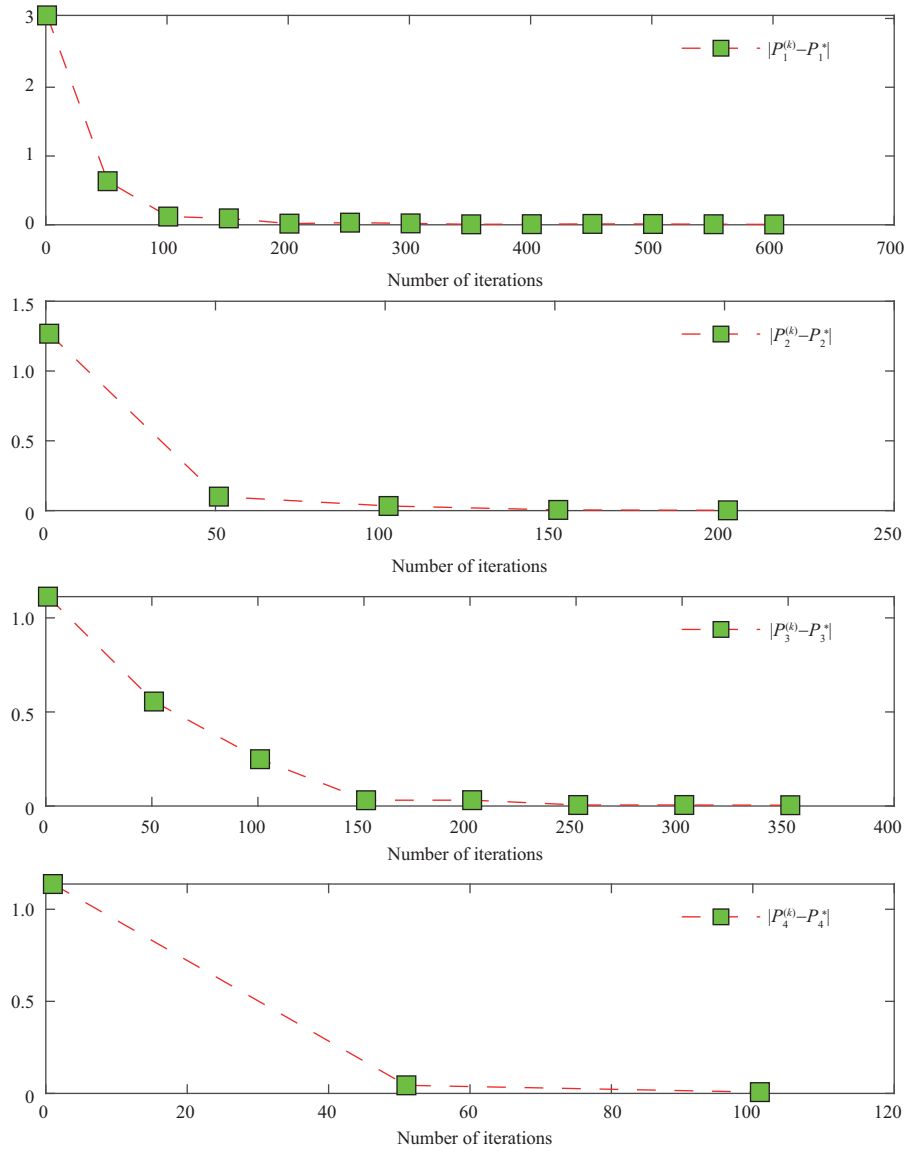

Figure 2 (Color online) Comparison P_i^* and $P_i^{(k)}$ learned at the iteration k via Algorithm 1 for agents $i = 1, 2, 3, 4$.

4 Simulation and discussions

To evaluate the efficiency of the proposed learning Algorithm 1 and the resilience of the closed-loop system, we consider the example of a 5-agent system where agent 0 is the leader modeled by the exosystem (4),

Table 1 Comparison of $K_i^{(k)}$ and K_i^* when the convergence is achieved

Control gain	Value
K_1^*	[1.4180, 1.7567, −0.7008, 0.0941]
$K_1^{(605)}$	[1.4167, 1.7558, −0.6996, 0.0947]
K_2^*	[1.3313, 1.2575, −0.6141, 0.3505]
$K_2^{(221)}$	[1.3317, 1.2574, −0.6140, 0.3515]
K_3^*	[1.2589, 1.0437, −0.5418, 0.4544]
$K_3^{(391)}$	[1.2580, 1.0433, −0.5407, 0.4542]
K_4^*	[1.2083, 0.9242, −0.4911, 0.5087]
$K_4^{(149)}$	[1.2076, 0.9239, −0.4902, 0.5087]

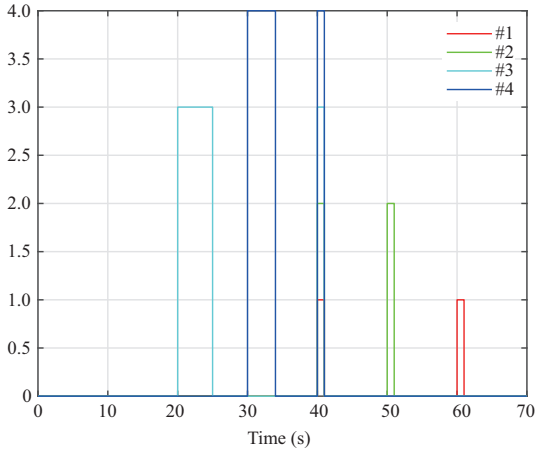

Figure 3 (Color online) The DDoS attack profile for agents $i = 1, 2, 3, 4$. The DDoS attack signal $\text{Attack}_i(t) = i$ if the agent i is isolated, and $\text{Attack}_i(t) = 0$ otherwise.

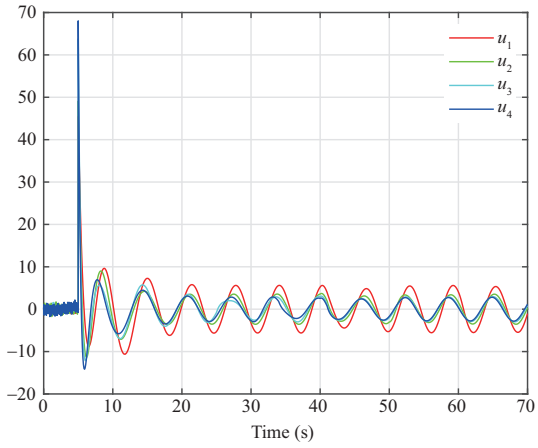

Figure 4 (Color online) The trajectories of control inputs $u_i(t)$ for agents $i = 1, 2, 3, 4$.

the remaining agents $i = 1, 2, 3, 4$ are followers modeled by (1)–(3), and the system matrices are

$$A_i = \begin{bmatrix} 0 & 1 \\ 0.01 \times i & 0 \end{bmatrix}, \quad B_i = \begin{bmatrix} 0 \\ i \end{bmatrix}, \quad D_i = \begin{bmatrix} 0 & 0 \\ 0 & 0.5 \times i \end{bmatrix},$$
$$C_i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\mathrm{T}}, \quad F = \begin{bmatrix} -1 \\ 0 \end{bmatrix}^{\mathrm{T}}, \quad E = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

We choose the vector $G = [0\ 1]^{\mathrm{T}}$, and it is apparent that the pair (E, G) is controllable.

Figure 1 depicts the communication topology of the 5-agent system. Note that its communications are subject to DDoS attacks, which include four distributed attacks, each of which can isolate an agent. For instance, if the system is under attack #1, then agent 1 is isolated. If attack #4 is active, then agent 4 becomes isolated. The system possibly may also be under multiple attacks simultaneously. In the worst case, where all attacks are active, all the agents in the system are isolated. First, we choose the initial control policy as $K_i^{(0)} = [0\ 0\ 0\ 0]$, which is not an admissible control policy for any $i = 1, 2, 3, 4$. Since the proposed algorithm is a VI algorithm, we can start from an arbitrary control policy, not necessarily admissible. During the time interval $t \in [0, 3]$ s, the control input $u_i(t)$ is selected using a combination of sinusoidal signals with angular frequencies ranging from 0.5 to 1000 Hz such that the rank condition (32) holds for $i = 1, 2, 3, 4$. After applying Algorithm 1, we approach the optimal control gains K_i^* and optimal values P_i^* by successive approximations. For comparison, we show the difference between $P_i^{(k)}$ and P_i^* at iteration k in Figure 2. We also illustrate $K_i^{(k)}$ and K_i^* in Table 1. One can check from Figure 2 and Table 1 that convergence can be achieved, although each agent has different dynamics.

We further compute the upper bound of the DDoS duration criterion through (42), obtaining $T_1^* = 230$, $T_2^* = 149$, $T_3^* = 128$, and $T_4^* = 125$. Note that the system can endure stronger DDoS attacks than the computed upper bound. As a test, we applied the DDoS attack profile depicted in Figure 3, which

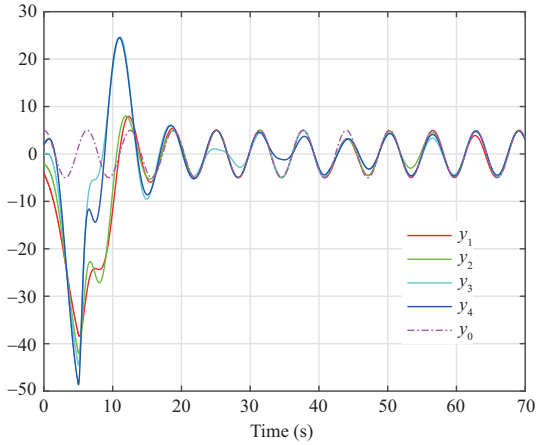

Figure 5 (Color online) The trajectories of outputs $y_i(t)$ for agents $i = 1, 2, 3, 4$ and the reference $y_0(t)$ generated by the agent 0.

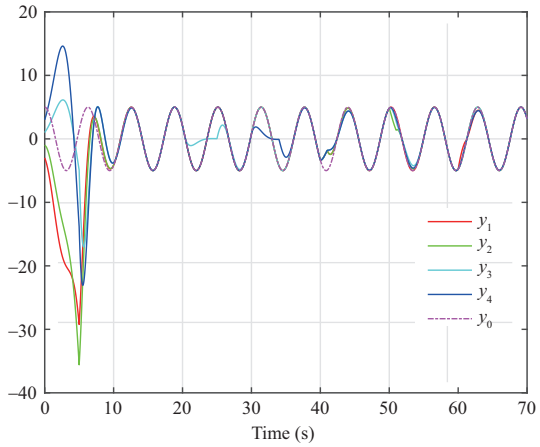

Figure 6 (Color online) The trajectories of outputs $y_i(t)$ for agents $i = 1, 2, 3, 4$ and the reference $y_0(t)$ generated by the agent 0 using feedback-feedforward control method in [1].

includes different DDoS attacks isolating one or multiple agent(s). The inputs and outputs of the multi-agent system are shown in Figures 4 and 5, respectively. As depicted in Figure 3, agent 3 is under DDoS attacks such that it is isolated when $t \in [20, 25]$ s, and agent 4 is isolated when $t \in [30, 35]$ s. These attacks cause the outputs $y_2(t)$ and $y_4(t)$ of agents 2 and 4 to temporarily deviate from the reference $y_0(t)$; see Figure 5. To make DDoS attacks more challenging, let all agents be isolated at $t \in [40, 41]$ s. One can observe from Figure 5 that although the transient performance of multi-agent systems is affected by DDoS attacks, all the agents can asymptotically track the desired reference signal.

As a comparison, we assumed that the system dynamics was exactly known and developed a controller with a distributed observer using the method proposed in [1]. Figure 6 depicts the simulation result. One can see that the DDoS attacks have a stronger effect on the transient performance therein compared with Figure 5.

5 Conclusion

This paper bridged the gap between ADP, the internal model principle, and distributed control theory to solve the data-driven cooperative output regulation problems of linear multi-agent systems. New results in the resilience to DDoS attacks and the robustness to unknown system dynamics are obtained. Our future work will be directed at solving the data-driven cooperative output regulation problems of nonlinear multi-agent systems.

Acknowledgements The work was supported in part by Science and Technology Major Project 2020 of Liaoning Province (Grant No. 2020JH1/10100008), National Natural Science Foundation of China (Grant Nos. 61991404, 61991400), 111 Project 2.0 (Grant No. B08015), and U.S. National Science Foundation (Grant No. CNS-2227153).

References

1 Su Y F, Huang J. Cooperative output regulation of linear multi-agent systems. IEEE Trans Automat Contr, 2012, 57: 1062–1066
2 Liu W, Huang J. Cooperative global robust output regulation for a class of nonlinear multi-agent systems with switching network. IEEE Trans Automat Contr, 2015, 60: 1963–1968
3 Li X, Soh Y C, Xie L, et al. Cooperative output regulation of heterogeneous linear multi-agent networks via H_∞ performance allocation. IEEE Trans Automat Contr, 2019, 64: 683–696
4 Zhang D, Deng C, Feng G. Resilient cooperative output regulation for nonlinear multi-agent systems under DoS attacks. IEEE Trans Automat Contr, 2022. doi: 10.1109/TAC.2022.3184388
5 Yan Y, Chen Z. Cooperative output regulation of linear discrete-time time-delay multi-agent systems by adaptive distributed observers. Neurocomputing, 2019, 331: 33–39
6 Ren W, Beard R. Distributed Consensus in Multi-vehicle Cooperative Control. London: Springer-Verlag 2008
7 Yu Z, Huang D, Jiang H, et al. Distributed consensus for multiagent systems via directed spanning tree based adaptive control. SIAM J Control Optim, 2018, 56: 2189–2217
8 Gao W, Mynuddin M, Wunsch D C, et al. Reinforcement learning-based cooperative optimal output regulation via distributed adaptive internal model. IEEE Trans Neural Netw Learn Syst, 2022, 33: 5229–5240
9 Cai H, Lewis F L, Hu G, et al. The adaptive distributed observer approach to the cooperative output regulation of linear multi-agent systems. Automatica, 2017, 75: 299–305

10 Jiang Z P, Bian T, Gao W. Learning-based control: a tutorial and some recent results. FNT Syst Control, 2020, 8: 176–284
11 Murray J J, Cox C J, Lendaris G G, et al. Adaptive dynamic programming. IEEE Trans Syst Man Cybern C, 2002, 32: 140–153
12 Kiumarsi B, Vamvoudakis K G, Modares H, et al. Optimal and autonomous control using reinforcement learning: a survey. IEEE Trans Neural Netw Learn Syst, 2018, 29: 2042–2062
13 Liu D, Xue S, Zhao B, et al. Adaptive dynamic programming for control: a survey and recent advances. IEEE Trans Syst Man Cybern Syst, 2021, 51: 142–160
14 Gao W, Jiang Z P. Learning-based adaptive optimal output regulation of linear and nonlinear systems: an overview. Control Theor Technol, 2022, 20: 1–19
15 Li J, Modares H, Chai T, et al. Off-policy reinforcement learning for synchronization in multiagent graphical games. IEEE Trans Neural Netw Learn Syst, 2017, 28: 2434–2445
16 Gao W, Jiang Z P, Lewis F L, et al. Leader-to-formation stability of multiagent systems: an adaptive optimal control approach. IEEE Trans Automat Contr, 2018, 63: 3581–3587
17 Ye Z, Zhang D, Wu Z G, et al. A3C-based intelligent event-triggering control of networked nonlinear unmanned marine vehicles subject to hybrid attacks. IEEE Trans Intell Transp Syst, 2022, 23: 12921–12934
18 Deng C, Zhang D, Feng G. Resilient practical cooperative output regulation for MASs with unknown switching exosystem dynamics under DoS attacks. Automatica, 2022, 139: 110172
19 Specht S M, Lee R B. Distributed denial of service: taxonomies of attacks, tools and countermeasure. In: Proceedings of the International Conference on Parallel and Distributed Computing Systems, Tokyo, 2004. 543–550
20 Zargar S T, Joshi J, Tipper D. A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. IEEE Commun Surv Tutorials, 2013, 15: 2046–2069
21 Xu W, Hu G, Ho D W C, et al. Distributed secure cooperative control under denial-of-service attacks from multiple adversaries. IEEE Trans Cybern, 2020, 50: 3458–3467
22 Deng C, Wen C. MAS-based distributed resilient control for a class of cyber-physical systems with communication delays under DoS attacks. IEEE Trans Cybern, 2021, 51: 2347–2358
23 Chen J, Zhang H, Yin G. Distributed dynamic event-triggered secure model predictive control of vehicle platoon against DoS attacks. IEEE Trans Veh Technol, 2023, 72: 2863–2877
24 Wang X, Hong Y, Huang J, et al. A distributed control approach to a robust output regulation problem for multi-agent linear systems. IEEE Trans Automat Contr, 2010, 55: 2891–2895
25 Hu W, Liu L. Cooperative output regulation of heterogeneous linear multi-agent systems by event-triggered control. IEEE Trans Cybern, 2017, 47: 105–116
26 Huang J. Nonlinear Output Regulation: Theory and Applications. Philadelphia: SIAM, 2004
27 de Persis C, Tesi P. Input-to-state stabilizing control under denial-of-service. IEEE Trans Automat Contr, 2015, 60: 2930–2944
28 Bian T, Jiang Z P. Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design. Automatica, 2016, 71: 348–360
29 Al-Tamimi A, Lewis F L, Abu-Khalaf M. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof. IEEE Trans Syst Man Cybern B, 2008, 38: 943–949
30 Bian T, Jiang Z P. Reinforcement learning and adaptive optimal control for continuous-time nonlinear systems: a value iteration approach. IEEE Trans Neural Netw Learn Syst, 2022, 33: 2781–2790
31 Jiang Y, Jiang Z P. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. Automatica, 2012, 48: 2699–2704
32 Sontag E D. Input to state stability: basic concepts and results. In: Nonlinear and Optimal Control Theory. Berlin: Springer-Verlag, 2007. 163–220
33 Khalil H K. Nonlinear Systems. 3rd ed. Upper Saddle River: Prentice Hall PTR, 2002