Preprints of the 23rd IFAC World Congress
Busan, Republic of Korea, August 23-28, 2026

Byzantine-Resilient Leaderless Formation Control in Open Multi-Agent Systems ⋆

Xince Wang * **Xin Gong** *

* *School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China.*

Abstract: This paper presents a fixed-time leaderless formation control framework for open multi-agent systems under Byzantine edge attacks. A coordinate-independent vector MSR estimation layer and a nonlinear control law are integrated to ensure resilient and predictable convergence. The method guarantees stability under topology switching, agent membership variation, and persistent adversarial behaviors.

Keywords: Leaderless formation; Fixed-time convergence; Byzantine resilience; Multi-agent systems; Distributed control; Open networks

1. INTRODUCTION

Formation control of multi-agent systems (MASs) has become an important topic in distributed control and cooperative robotics, with applications ranging from autonomous vehicles and sensor networks to industrial automation (Wang et al. (2023); LeBlanc et al. (2013)). Most existing formation control schemes assume a fixed communication topology and a designated leader that provides a global reference for coordination (Shang (2020)). However, such assumptions are difficult to satisfy in open and adversarial environments, where agents may dynamically join or leave the system and communication channels may be unreliable or even maliciously manipulated.

In practical open systems, such as intelligent transportation networks and industrial Internet platforms, the population of agents often changes over time. Meanwhile, adversarial nodes may launch Byzantine edge attacks (BEA), sending inconsistent or falsified information to different neighbors (Amirian and Shamaghdari (2021)) has been extensively studied. When openness and Byzantine behaviors coexist, conventional leader-based architectures become fragile: once the leader is isolated or corrupted, the entire formation may fail. This motivates the design of a distributed, leaderless, and Byzantine-resilient control framework capable of maintaining formation coherence and ensuring predictable convergence in such dynamic environments (Xiao and Dong (2021)).

Although resilient formation control has been extensively studied, several key challenges remain unresolved (Almuzakki and Jayawardhana (2023); Usevitch and Panagou (2019)). The first lies in modeling and coordination under dynamic membership: when agents join or leave, preserving the relative formation pattern without relying on a global leader is nontrivial. The second challenge arises from Byzantine edge attacks, where compromised agents send distinct data to different neighbors, violating the basic symmetry assumptions used in consensus theory. Finally, it remains difficult to guarantee fixed-time convergence—ensuring all agents reach formation consensus within a uniform time bound that does not depend on initial states—especially when both openness and adversarial behavior are present.

To overcome these difficulties, this paper proposes a fixed-time leaderless formation control framework for open multi-agent systems under Byzantine edge attacks (Zhang et al. (2025)). The framework consists of two cooperative layers that operate in continuous time. The first layer is a coordinate-independent vector Mean-Subsequence-Reduced (MSR) estimator that filters out malicious neighbor information and reconstructs a consistent local reference for each normal agent. Based on these filtered references, the second layer implements a nonlinear control law with adaptive parameters to guarantee fixed-time convergence under general linear dynamics. Together, the two layers form a unified distributed mechanism that ensures both resilience and predictability of system behavior, independent of initial conditions.

The proposed approach extends the classical MSR-based consensus methods to high-dimensional leaderless formation problems and provides explicit convergence guarantees under $(2f+1)$-robust connectivity. It also establishes a theoretical foundation for secure coordination in open and adversarial environments, where system membership, topology, and attack patterns may all evolve over time (Liu et al. (2024); Sun et al. (2024)). The main contributions are listed as follows:

(1) **Byzantine-resilient leaderless formation:** A novel two-layer distributed control framework is proposed

⋆ This work was supported by the National Natural Science Foundation of China under Grant 62403128, the Jiangsu Provincial Natural Science Foundation of China under Grant No. BK20241284, the OpenFoundation of Key Laboratory of Cyberspace Security, Ministry of Education (No. KLCS20240403), the Open Research Project of State KeyLaboratory of Industrial Control Technology, China under Grant No.ICT2025B56, and Start-up Research Fund of Southeast University under Grants RF1028623260.
Corresponding author: Xin Gong.
E-mail: xincewang@seu.edu.cn (X. Wang); xingong@seu.edu.cn (X. Gong);

©2026 the authors. Accepted by IFAC for publication under a Creative Commons Licence CC-BY-NC-ND

to achieve leaderless formation consensus in open multi-agent systems under Byzantine edge attacks, without relying on any global leader or fixed topology.

(2) **Fixed-time nonlinear control law:** A dual-power nonlinear feedback control law is designed to guarantee fixed-time convergence of formation errors, independent of initial conditions, thus improving predictability in transient performance.

(3) **Unified stability under openness and switching:** Rigorous Lyapunov-based analysis establishes uniform boundedness and fixed-time convergence under dynamic membership, switching topologies, and persistent Byzantine attacks—covering realistic open-network conditions.

Notations: Throughout this paper, $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$ denote the sets of real n-dimensional vectors and $n \times m$ matrices, respectively. For a vector $x = [x_1, \ldots, x_n]^\top$, $\|x\|$ denotes the Euclidean norm, and $\mathrm{sgn}(x)$ is the element-wise sign function. I_n and $\mathbf{1}_n$ denote the $n \times n$ identity matrix and the all-ones vector, respectively. The operator $\mathrm{diag}(\cdot)$ constructs a diagonal matrix, and $\otimes$ denotes the Kronecker product. The communication topology at time t is represented by a directed graph $\mathcal{G}(t) = (V(t), E(t))$ with adjacency matrix $A(t) = [a_{ij}(t)]$ and Laplacian $L(t) = D(t) - A(t)$. The symmetric part of the Laplacian is $Q(t) = \frac{1}{2}(L(t)+L^\top(t))$, and its smallest nonzero eigenvalue λ_Q characterizes the algebraic connectivity of the graph. The neighbor set of agent i is denoted by $\mathcal{N}_i(t)$, and $\dot{f}(t)$ represents the time derivative of $f(t)$. All vectors and matrices are assumed to have compatible dimensions.

2. PRELIMINARIES AND PROBLEM FORMULATION

2.1 Graph Theory and Robustness

Consider a time-varying directed graph $\mathcal{G}(t) = (V(t), E(t))$, where $V(t) = \{1, 2, \ldots, N(t)\}$ denotes the set of agents at time t, and $E(t) \subseteq V(t) \times V(t)$ represents the set of directed communication edges. If $(j, i) \in E(t)$, agent j can transmit information to agent i. The adjacency matrix associated with $\mathcal{G}(t)$ is denoted as $A(t) = [a_{ij}(t)] \in \mathbb{R}^{N(t)\times N(t)}$, where $a_{ij}(t) > 0$ if $(j, i) \in E(t)$ and $a_{ij}(t) = 0$ otherwise. The in-degree of agent i is $d_i(t) = \sum_{j=1}^{N(t)} a_{ij}(t)$, and the corresponding Laplacian matrix is defined as

$$L(t) = D(t) - A(t), \tag{1}$$

where $D(t) = \mathrm{diag}(d_1(t), \ldots, d_{N(t)}(t))$. To handle asymmetric topologies, we introduce the symmetric Laplacian

$$Q(t) = \tfrac{1}{2}\big(L(t) + L^\top(t)\big), \tag{2}$$

which is positive semidefinite for any connected graph. The smallest nonzero eigenvalue λ_Q characterizes the algebraic connectivity and will be crucial in later Lyapunov-based convergence analysis.

Let $H(t)$ denote a projection matrix satisfying $H(t)\mathbf{1} = 0$ and $H(t)H^\top = I$. For any state vector $x = [x_1^\top, \ldots, x_{N(t)}^\top]^\top$, the quantity $\|H(t)x\|$ measures the disagreement among agents and vanishes if and only if consensus is achieved.

Definition 1 ($(2f+1)$-robust graph)**.** *A directed graph $\mathcal{G}(t)$ is said to be $(2f+1)$-robust if, for any two disjoint nonempty subsets $S_1, S_2 \subset V(t)$, at least one of them contains no fewer than $(2f+1)$ nodes that have incoming edges from $V(t) \setminus S_k$.*

This robustness property guarantees that even if up to f neighbors are malicious, sufficient normal information still flows between subsets of agents. It has been shown that $(2f+1)$-robustness is a necessary and sufficient condition for resilient consensus under the f-local Byzantine model.

2.2 Byzantine Edge Attack Model

In adversarial environments, compromised nodes may transmit inconsistent or falsified data to different neighbors. Such behavior is modeled as a *Byzantine Edge Attack (BEA)*. Specifically, a Byzantine node j may send distinct falsified values $y_{ji}(t)$ to each neighbor $i \in \mathcal{N}_j(t)$ at time t. Let $\mathcal{N}_i(t)$ denote the set of neighbors of agent i. It can be partitioned into two disjoint subsets: the normal neighbors $\mathcal{N}_i^n(t)$ and the Byzantine neighbors $\mathcal{N}_i^b(t)$.

Definition 2 (f-local Byzantine model)**.** *At any time t, each normal node i has at most f Byzantine neighbors, i.e.,*

$$|\mathcal{N}_i^b(t)| \le f. \tag{3}$$

This assumption restricts only the local density of adversarial nodes, not their total number, and is widely used in distributed fault-tolerant control. Under an $(2f+1)$-robust graph, every normal agent can safely remove up to f largest and f smallest neighbor values without losing all normal information, which serves as the foundation of the resilient estimation algorithm developed later.

2.3 System Model

Each agent $i \in V(t)$ is governed by the following linear dynamics:

$$\dot{x}_i(t) = Ax_i(t) + Bu_i(t), \tag{4}$$

where $x_i(t) \in \mathbb{R}^n$ and $u_i(t) \in \mathbb{R}^m$ are the state and control input, respectively, while $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ are constant matrices shared by all agents.

In open multi-agent systems, the node set $V(t)$ may vary over time due to agent joining or leaving events. Let $\{t_k\}_{k=0}^{\infty}$ denote the sequence of event times such that the membership remains constant in each interval $[t_k, t_{k+1})$. At the switching instant t_k, the membership updates as

$$V(t_k^+) = \big(V(t_k^-) \setminus V_{\mathrm{rmv}}(t_k)\big) \cup V_{\mathrm{add}}(t_k), \tag{5}$$

where $V_{\mathrm{add}}(t_k)$ and $V_{\mathrm{rmv}}(t_k)$ denote the sets of joining and leaving agents, respectively. Each joining agent initializes its state and control input with bounded values to ensure continuity of network dynamics.

The desired formation is described by a set of time-varying reference configurations $\{h_i(t)\}$ that evolve according to

$$\dot{h}_i(t) = Ah_i(t) + Bv_i(t), \tag{6}$$

where $v_i(t)$ serves as an auxiliary feedforward signal ensuring the dynamic feasibility of the formation trajectories. A feasible and widely adopted expression for $v_i(t)$ is given by $v_i(t) = \bar{B}^{\|}\dot{h}_i(t) - \bar{B}^{\|}(A + BK_0)h_i(t)$, where $\bar{B} = [(\bar{B}^{\|})^\top, (\bar{B}^{\perp})^\top]^\top$ is a nonsingular matrix satisfying $\bar{B}^{\|}B = I_m$ and $\bar{B}^{\perp}B = 0$ (Gong et al. (2020)). This formulation compensates internal dynamics and guarantees that the desired reference configurations $\{h_i(t)\}$ are compatible

with the agent model. A sufficient feasibility condition is $\lim_{t\to\infty}\left(\bar{B}^{\perp}Ah_{ij}(t)-\bar{B}^{\perp}\dot{h}_{ij}(t)\right)=0$, which ensures that the time-varying formation configuration can be realized by all normal agents.

The individual formation error is defined as $y_i(t)=x_i(t)-h_i(t)$, which quantifies the deviation of agent i from its nominal formation configuration. Stacking all agents yields the global error vector$y(t)=[y_1^{\top}(t),y_2^{\top}(t),\ldots,y_{N(t)}^{\top}(t)]^{\top}$.

2.4 Problem Definition

We now formalize the leaderless formation consensus problem under Byzantine edge attacks.

Definition 3 (Leaderless Formation Consensus under BEA)**.** *Consider the open multi-agent system described by* (4)*, where each agent evolves under an* $(2f+1)$*-robust communication graph* $\mathcal{G}(t)=(V(t),E(t))$ *and an* f*-local Byzantine edge-attack model. Let* $h_i(t)\in\mathbb{R}^n$ *denote the desired local formation configuration of agent* i*. The agents are said to achieve* leaderless formation consensus *if*

$$\lim_{t\to T_s}\left[(x_i(t)-h_i(t))-(x_j(t)-h_j(t))\right]=0,\forall\, i,j\in V^n(t) \tag{7}$$

for some finite $T_s>0$ *independent of the initial conditions.*

This definition indicates that all normal agents can align their relative formation errors according to their own reference configurations without relying on any global leader or centralized coordination. The goal of this paper is to design a fully distributed control framework that ensures this property under time-varying topology, dynamic membership, and Byzantine edge attacks. Specifically, the proposed framework consists of two layers:

(1) **Byzantine-Resilient Estimation Layer (TL):** A coordinate-independent vector MSR algorithm that filters malicious information and reconstructs consistent local references for each agent.
(2) **Leaderless Fixed-Time Control Layer (CPL):** A nonlinear control law ensuring fixed-time formation consensus for all normal agents under general linear dynamics.

The detailed algorithm design and stability analysis are presented in the following sections.

3. BYZANTINE-RESILIENT ESTIMATION LAYER

In this section, we develop a Byzantine-resilient estimation mechanism to mitigate the effect of edge-level adversarial behaviors. The design is based on a coordinate-independent vector Mean-Subsequence-Reduced (MSR) algorithm, which removes extreme neighbor values in each coordinate dimension and provides each normal agent with a reliable local reference for subsequent control.

3.1 Coordinate-Independent Vector MSR Algorithm

At each time instant t, a normal agent $i\in V^n(t)$ receives the reported states $y_j(t)$ from all its in-neighbors $j\in\mathcal{N}_i(t)\subseteq V(t)$, where some of these messages may be falsified by Byzantine agents. To counteract this, agent i performs a coordinate-wise trimming operation on the received data.

Let $y_j^{(k)}(t)$ denote the k-th coordinate of $y_j(t)$. For each coordinate $k\in\{1,\ldots,n\}$, agent i sorts the collection $\{y_j^{(k)}(t):j\in\mathcal{N}_i(t)\}$ in ascending order, removes the f largest and f smallest entries, and averages the remaining ones. The k-th component of the resilient estimate $\hat{y}_i(t)$ is thus given by

$$\hat{y}_i^{(k)}(t)=\frac{1}{|\mathcal{M}_i^{(k)}(t)|}\sum_{j\in\mathcal{M}_i^{(k)}(t)}y_j^{(k)}(t), \tag{8}$$

where $\mathcal{M}_i^{(k)}(t)=\mathrm{trim}_f\left(\{y_j^{(k)}(t)\}\right)$ denotes the set of neighbor indices that remain after removing the f largest and f smallest coordinate values. The coordinate-wise operation makes the algorithm invariant under coordinate transformations and applicable to vector-valued systems.

Remark 1. *Compared with scalar MSR algorithms, the proposed vector version independently trims each coordinate, avoiding cross-dimensional coupling and ensuring robustness even when Byzantine nodes transmit inconsistent vector entries. The algorithm relies only on local neighbor information and is fully distributed.*

3.2 Resilient Consensus Property

The next lemma establishes that, under the $(2f+1)$-robust topology, the estimates of all normal agents asymptotically reach a bounded consensus in finite time.

Lemma 1 (Resilient consensus under vector MSR)**.** *Consider a group of normal agents executing the update rule* (8) *over a* $(2f+1)$*-robust directed graph* $\mathcal{G}(t)$*. Under the* f*-local Byzantine model, there exists a finite time constant* $T_0>0$ *such that for all* $t\geq t_0+T_0$*,*

$$\|\hat{y}_i(t)-\hat{y}_j(t)\|\leq\delta_T,\qquad\forall\, i,j\in V^n(t), \tag{9}$$

where $\delta_T>0$ *is a residual bound determined by communication delay and update resolution. If communication is continuous and synchronous,* $\delta_T=0$*, and the estimates achieve exact consensus.*

Proof. For each coordinate k, define the ordered set of normal agents' data $\mathcal{S}^{(k)}(t)=\{y_i^{(k)}(t):i\in V^n(t)\}$, and let $m^{(k)}(t)=\min\mathcal{S}^{(k)}(t)$ and $M^{(k)}(t)=\max\mathcal{S}^{(k)}(t)$ be its lower and upper envelopes, respectively. Since each normal agent removes up to f extreme values at both ends, and the $(2f+1)$-robust property guarantees that at least one normal neighbor remains after trimming, the convex hull formed by $\mathcal{S}^{(k)}(t)$ is invariant, i.e., $m^{(k)}(t)\leq\hat{y}_i^{(k)}(t)\leq M^{(k)}(t), i\in V^n(t)$. Hence, the extreme values of normal estimates monotonically contract toward each other.

Denote the interval width by $W^{(k)}(t)=M^{(k)}(t)-m^{(k)}(t)$. For the continuous-time evolution of the estimation process, there exists a constant $\rho_s\in(0,1)$ such that after a bounded interval of duration T_0, the width satisfies the contraction condition

$$W^{(k)}(t+T_0)\leq\rho_s\,W^{(k)}(t),\qquad 0<\rho_s<1. \tag{10}$$

The above relation indicates that, at each time horizon of length T_0, the maximal deviation between normal agents in coordinate k shrinks by a constant factor ρ_s. Iterating (10) yields exponential decay of $W^{(k)}(t)$, and therefore for any $\varepsilon>0$ there exists a finite time T_0' such that $W^{(k)}(t)<\varepsilon$ for all $t>T_0'$. Stacking all coordinates leads to (9). □

3.3 Prescribed-Time Extension

To accelerate convergence and ensure that consensus is reached exactly at a predetermined time, a time-scaling function $\rho(t;t_0,T)$ can be introduced. Let $\rho(t;t_0,T)$ be a strictly increasing function satisfying $\rho(t_0) = 0, \lim_{t\to t_0+T}\rho(t;t_0,T) = \infty$. Each normal agent updates its resilient estimate according to

$$\dot{\hat{y}}_i(t) = -\kappa\,\rho(t;t_0,T)\sum_{j\in\mathcal{M}_i(t)}(\hat{y}_i(t)-y_j(t)), \tag{11}$$

where $\kappa>0$ is a control gain. The scaling factor $\rho(t;t_0,T)$ accelerates convergence so that the estimation error vanishes precisely at $t=t_0+T$.

Lemma 2 (Prescribed-time convergence of TL)**.** *Under the update rule* (11)*, the estimation errors of all normal agents satisfy* $\|\hat{y}_i(t)-\hat{y}_j(t)\| \le \beta\exp\Big(-\int_{t_0}^{t}\kappa\rho(\tau;t_0,T)\,d\tau\Big)$*, and therefore* $\lim_{t\to t_0+T}\|\hat{y}_i(t)-\hat{y}_j(t)\|=0$.

Proof. Define the Lyapunov function as the total pairwise disagreement energy:

$$V(t)=\frac{1}{4}\sum_{i=1}^{N(t)}\sum_{j=1}^{N(t)}\|\hat{y}_i(t)-\hat{y}_j(t)\|^2, \tag{12}$$

where the summation is taken over all unordered pairs of distinct agents $i,j\in V^n(t)$. This expression is equivalent to $\frac{1}{2}\sum_{1\le i<j\le N(t)}\|\hat{y}_i(t)-\hat{y}_j(t)\|^2$. Differentiating (12) along the dynamics (11) yields $\dot{V}(t)\le -2\kappa\,\rho(t;t_0,T)\,V(t)$. Integrating both sides gives $V(t)\le V(t_0)\exp\Big(-2\kappa\int_{t_0}^{t}\rho(\tau;t_0,T)\,d\tau\Big)$. Since $\rho(t;t_0,T)\to\infty$ as $t\to t_0+T$, it follows that $V(t)\to 0$, and thus all resilient estimates $\hat{y}_i(t)$ coincide exactly at t_0+T. □

The estimation layer (TL) provides each normal agent with a reliable local reference $\hat{y}_i(t)$ by removing up to f malicious inputs per dimension and leveraging $(2f+1)$-robust connectivity. With an optional time-scaling function $\rho(t;t_0,T)$, the TL layer ensures prescribed-time consensus and supplies resilient references for the subsequent control layer. The resulting outputs $\hat{y}_i(t)$ will serve as resilient local references for the control layer in the next section to achieve fixed-time leaderless formation consensus.

4. FIXED-TIME LEADERLESS FORMATION CONTROL LAYER

After the estimation layer (TL) reaches bounded consensus within a finite time T_0, the filtered outputs $\hat{y}_i(t)$ are used to design a distributed control law that guarantees fixed-time leaderless formation consensus of all normal agents.

4.1 Control Law Design

Using the resilient references $\hat{y}_i(t)$ from the TL layer, each normal agent computes a relative formation deviation with respect to its neighbors:

$$\zeta_i(t)=\sum_{j\in\mathcal{N}_i(t)}\big[(\hat{y}_i(t)-h_i(t))-(\hat{y}_j(t)-h_j(t))\big]. \tag{13}$$

The distributed control input is designed as

$$\begin{aligned}u_i(t)=&-\hat{\alpha}_iK_1\zeta_i(t)-\hat{\phi}_iK_2\,\mathrm{sig}^{1-\mu}\big(P\zeta_i(t)\big)\\&-\hat{\varphi}_iK_3\,\mathrm{sig}^{1+\mu}\big(P\zeta_i(t)\big)-v_i(t),\end{aligned} \tag{14}$$

where $K_1,K_2,K_3\in\mathbb{R}^{m\times n}$ are feedback gains and $P>0$ satisfies the Lyapunov equation

$$A^\top P+PA-PBB^\top P+I=0. \tag{15}$$

The exponent $0<\mu<1$ adjusts the nonlinearity of convergence, and $\mathrm{sig}^\gamma(x)$ denotes the element-wise signed power: $\mathrm{sig}^\gamma(x)=[|x_1|^\gamma\mathrm{sgn}(x_1),\dots,|x_n|^\gamma\mathrm{sgn}(x_n)]^\top$. Adaptive parameters evolve as

$$\dot{\hat{\alpha}}_i=\eta_1\|\zeta_i\|, \tag{16a}$$

$$\dot{\hat{\phi}}_i=\eta_2\|P\zeta_i\|^{1-\mu}, \tag{16b}$$

$$\dot{\hat{\varphi}}_i=\eta_3\|P\zeta_i\|^{1+\mu}. \tag{16c}$$

with positive gains $\eta_1,\eta_2,\eta_3>0$.

Remark 2. *The control law* (14) *combines low-order and high-order nonlinear feedback to ensure fixed-time convergence independent of initial conditions. Employing* $\hat{y}_i(t)$ *instead of raw* $y_j(t)$ *makes the scheme resilient to Byzantine distortions, as* $\hat{y}_i(t)$ *has been filtered through the TL layer.*

4.2 Closed-Loop System Representation

Define the stacked formation error $y(t)=[y_1^\top,\dots,y_{N(t)}^\top]^\top$. Applying (14) to the system dynamics gives

$$\begin{aligned}\dot{y}(t)=&(I_N\otimes(A-BK_1))y(t)+\Delta_{\mathrm{TL}}(t)\\&-(L(t)\otimes BK_3)\,\mathrm{sig}^{1+\mu}\big((P\otimes I_N)y(t)\big)\\&-(L(t)\otimes BK_2)\,\mathrm{sig}^{1-\mu}\big((P\otimes I_N)y(t)\big),\end{aligned} \tag{17}$$

where $\Delta_{\mathrm{TL}}(t)$ denotes the residual from imperfect estimation. From Lemma 1, $\|\Delta_{\mathrm{TL}}(t)\|\le\delta_T$ and $\Delta_{\mathrm{TL}}(t)=0$ for all $t\ge t_0+T_0$ under continuous communication.

4.3 Fixed-Time Convergence Analysis

Lemma 3 (Spectral lower bound)**.** *If the communication graph* $\mathcal{G}(t)$ *is* α*-strongly connected, the symmetric Laplacian* $Q(t)=\frac{1}{2}(L(t)+L^\top(t))$ *satisfies*$x^\top Q(t)x\ge\lambda_Q\|x-\bar{x}\mathbf{1}\|^2$*, where* $\bar{x}=\frac{1}{N(t)}\mathbf{1}^\top x$.

Proof. The result follows from standard spectral graph arguments. For an α-strongly connected digraph, its undirected counterpart contains an α-edge-connected spanning subgraph, yielding a uniform lower bound $\lambda_Q=\frac{4\alpha}{N(t)\,\mathrm{diam}(\mathcal{G}(t))}>0$. □

Theorem 1 (Fixed-time leaderless formation consensus)**.** *Consider the open multi-agent system* (4) *subject to the control law* (14)*. Under* $(2f+1)$*-robust connectivity and the standing assumptions on switching and events stated in Section 5, all normal agents achieve leaderless formation consensus in fixed time* $T_s>0$ *independent of initial conditions. Formally, there exists a finite time* $T_s>0$ *such that*

$$\lim_{t\to T_s^-}\max_{\substack{i,j\in V^n(t)\\ i\ne j}}\|(x_i(t)-h_i(t))-(x_j(t)-h_j(t))\|=0, \tag{18}$$

where $V^n(t)$ *denotes the set of normal (non-Byzantine) agents present in the network at time* t.

Proof. Define the Lyapunov candidate function

$$V(t) = y^\top (Q(t) \otimes P) y + \sum_{i \in V^n(t)} \frac{(\hat{\alpha}_i - \alpha)^2}{2\eta_1}, \tag{19}$$

where $\alpha > 0$ is a constant and $P > 0$ satisfies the algebraic Riccati equation $A^\top P + PA - PBB^\top P + I = 0$. The first term quantifies the collective formation energy among normal agents, while the second term penalizes adaptive gain deviations.

Differentiating (19) along the trajectories of (16)–(17) yields $\dot{V} \leq -c_1 V^{\frac{2-\mu}{2}} - c_2 V^{\frac{2+\mu}{2}} + \|\Delta_{\mathrm{TL}}(t)\|^2$, where $c_1, c_2 > 0$ depend on P, B, K_2, K_3, and λ_Q. The perturbation term $\Delta_{\mathrm{TL}}(t)$ arises from residual estimation errors in the TL layer. For $t \geq t_0 + T_0$, the TL layer has converged, so $\Delta_{\mathrm{TL}}(t) = 0$. Therefore, for $t \geq t_0 + T_0$,

$$\dot{V} \leq -c_1 V^{\frac{2-\mu}{2}} - c_2 V^{\frac{2+\mu}{2}}. \tag{20}$$

Applying the standard fixed-time stability lemma (e.g., Polyakov, 2012), we obtain

$$T_s \leq \frac{2(c_1 + c_2)}{\sigma \mu c_1 c_2}, \tag{21}$$

where $\sigma > 0$ satisfies $\dot{V} \leq -\sigma(c_1 V^p + c_2 V^q)$ with $0 < p < 1 < q$. The upper bound (21) is independent of initial conditions, ensuring fixed-time convergence.

Since $V(t)$ is positive definite with respect to the collective error vector $y(t)$, the convergence of $V(t) \to 0$ implies $\|(x_i - h_i) - (x_j - h_j)\| \to 0, \forall\, i, j \in V^n(t),\ i \neq j$, which verifies (18). The total convergence time of the two-layer system is thus $T_{\mathrm{total}} = T_0 + T_s$. □

Remark 3. *Bound* (21) *provides an explicit upper limit for convergence. Unlike finite-time schemes, T_s is independent of the initial $V(t_0)$, ensuring predictable transient performance. Since the TL layer completes within T_0, the global formation achieves consensus by $T_{total} = T_0 + T_s$.*

The control law (14) achieves Byzantine resilience and fixed-time convergence in a fully distributed, leaderless manner, where each agent relies only on local filtered data. It ensures robustness to falsified information, convergence within the preset time bound T_s, and stability under dynamic membership. The next section discusses overall system stability under topology switching and dynamic membership.

5. STABILITY ANALYSIS AND DISCUSSIONS

In this section, we analyze the stability of the closed-loop system overall after the TL layer has reached its bounded consensus at time $t = t_0 + T_0$. We show that the proposed two-layer framework preserves fixed-time convergence and uniform boundedness under switching topologies, open membership variations, and persistent Byzantine edge attacks.

5.1 Stability under Switching Topologies

Let the communication graph $\mathcal{G}(t)$ switch among a finite set $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_M\}$. Define the switching signal $\sigma(t) : [t_0, \infty) \to \{1, \ldots, M\}$ such that $\mathcal{G}(t) = \mathcal{G}_{\sigma(t)}$.

Assumption 1 (Average dwell-time condition)**.** *There exist constants $\tau_d > 0$ and $N_0 \geq 1$ such that, for any $t > t_0$, the number of topology switches $N_\sigma(t, t_0)$ satisfies*

$$N_\sigma(t, t_0) \leq N_0 + \frac{t - t_0}{\tau_d}. \tag{22}$$

Here, τ_d represents the average dwell time*—the minimum duration for which each topology must persist before switching occurs. To guarantee monotonic energy dissipation of the Lyapunov function $V(t)$ during switching, τ_d is further required to satisfy the following dwell-time constraint:*

$$\tau_d > \frac{\ln(1/\gamma_s)}{\kappa_s}, \tag{23}$$

where $\gamma_s \in (0, 1)$ bounds the instantaneous energy jump at each switching instant, and $\kappa_s > 0$ denotes the Lyapunov decay rate associated with each fixed topology.

This assumption prevents arbitrarily fast switching and ensures that the dissipative effect of each topology interval dominates the transient energy increase caused by switching events. Since the TL layer has converged by $t = t_0 + T_0$, all agents share consistent references $\hat{y}_i(t)$, and the CPL-layer dynamics evolve according to the Lyapunov inequality (20).

Theorem 2 (Uniform stability under switching)**.** *Consider the closed-loop system* (17) *under the control law* (14) *and Assumption 1. If the average dwell time τ_d satisfies the constraint* (23)*, then for all $t \geq t_0 + T_0$, the Lyapunov function $V(t)$ satisfies the uniform exponential bound$V(t) \leq \Omega\, e^{-\kappa_s (t - t_0 - T_0)} + \varepsilon_s$, where $\kappa_s > 0$ depends on τ_d, and $\varepsilon_s > 0$ is an exponentially vanishing term induced by topology switching.*

Proof. During each fixed-topology interval $[t_k, t_{k+1})$, the Lyapunov derivative satisfies $\dot{V} \leq -c_1 V^p - c_2 V^q$ with $0 < p < 1 < q$. At switching instants, the Laplacian $L_{\sigma(t)}$ changes, producing jumps $\Delta V_k = V(t_k^+) - V(t_k^-)$. Because $Q(t)$ is uniformly bounded, $|\Delta V_k| \leq \gamma_s V(t_k^-)$ for some $\gamma_s \in (0, 1)$. Applying the average dwell-time property yields$V(t) \leq e^{-\kappa_s (t - t_0 - T_0)} V(t_0 + T_0) + \frac{\gamma_s N_0}{1 - e^{-\kappa_s \tau_d}}$. Under the constraint (23), the cumulative decay term dominates, and the Lyapunov function decreases monotonically, leading to the stated exponential bound with $\Omega = V(t_0 + T_0)$ and $\varepsilon_s = \mathcal{O}(e^{-\kappa_s \tau_d})$. □

Remark 4. *Theorem 2 confirms that fixed-time convergence persists under sufficiently slow switching. The dwell-time constraint* (23) *is both a practical design rule and a necessary condition for the validity of the theorem, ensuring that the decay of $V(t)$ dominates the transient energy jumps caused by switching.*

5.2 Robustness to Agent Joining and Leaving

We next consider open-system behavior where agents may join or leave dynamically. Let t_k denote an event time when membership changes occur.

Assumption 2 (Bounded membership variation)**.** *At each event t_k, the number of joining and leaving agents satisfies $|V_{\mathrm{add}}(t_k)| + |V_{\mathrm{rmv}}(t_k)| \leq N_{\max}$, and joining agents initialize with bounded energy $\|y_i(t_k^+)\| \leq y_{\max}$.*

Assumption 3 (Minimum inter-event interval)**.** *There exists a minimum inter-event time $\tau_e > 0$ such that any two consecutive membership update instants t_k and t_{k+1} satisfy $t_{k+1} - t_k \geq \tau_e$. To ensure that the Lyapunov dissipation dominates the transient energy injection caused by membership changes, τ_e must further satisfy*

$$\tau_e > \frac{\ln(1/\gamma_e)}{\kappa_s}, \tag{24}$$

where $\gamma_e \in (0,1)$ bounds the instantaneous jump of $V(t)$ at an event, and $\kappa_s > 0$ is the decay rate of $V(t)$ between consecutive events.

Lemma 4 (Energy monotonicity at membership events)**.** *Under Assumptions 2–3, the Lyapunov function $V(t)$ satisfies $V(t_k^+) \leq V(t_k^-) + \chi N_{\max} y_{\max}^2$, where $\chi > 0$ depends on the largest eigenvalue of P.*

Theorem 3 (Uniform ultimate boundedness in open systems)**.** *Under the control law* (14)*, Assumptions 1–3, and the $(2f+1)$-robust topology, the closed-loop system is uniformly ultimately bounded (UUB), provided that the inter-event interval satisfies the constraint* (24)*. Specifically, there exist constants $c_V > 0$ and $r_V > 0$ such that $\limsup_{t\to\infty} \|y_i(t) - y_j(t)\| \leq r_V$, where $V(t) \leq c_V$ and $\forall t \geq t_0$.*

Proof. Between switching events, $V(t)$ decreases due to (20). At each membership change t_k, $V(t)$ increases by at most $\chi N_{\max} y_{\max}^2$ (Lemma 4). Because both the number and magnitude of these events are bounded and the inter-event interval satisfies (24), the total energy increment is finite: $\sum_{k=0}^{\infty}\big(V(t_k^+) - V(t_k^-)\big) < \infty$. Hence $V(t)$ remains uniformly bounded. Using $\|y_i - y_j\|^2 \leq 2V/(\lambda_Q \lambda_{\min}(P))$, the inter-agent formation errors are also bounded by $r_V = \sqrt{2c_V/(\lambda_Q \lambda_{\min}(P))}$. □

5.3 Consensus Preservation under Persistent Attacks

Even if a subset of Byzantine agents remain persistently malicious, the TL layer ensures resilience by filtering up to f extreme neighbor values.

Theorem 4 (Consensus preservation under persistent BEA)**.** *Assume the f-local Byzantine model holds for all t, the TL layer operates continuously, and the effective attack-update interval τ_a satisfies*

$$\tau_a > \frac{\ln(1/\gamma_a)}{\kappa_s}, \tag{25}$$

where $\gamma_a \in (0,1)$ bounds the transient energy perturbation caused by adversarial updates. Then, despite persistent Byzantine agents, the formation errors of all normal agents satisfy $\limsup_{t\to\infty} \|y_i(t) - y_j(t)\| \leq \epsilon_{\text{BEA}}$, where $\epsilon_{\text{BEA}} = \mathcal{O}(f/N_{\min})$ depends on the ratio of malicious to normal neighbors.

Proof. Each normal agent loses at most f neighbor contributions. Under $(2f+1)$-robust connectivity, at least one normal neighbor remains, ensuring the effective Laplacian $L_r(t)$ is connected. Thus, $\dot{V}(t) \leq -c_1 V^{\frac{2-\mu}{2}} - c_2 V^{\frac{2+\mu}{2}} + \omega_f$, where $\omega_f = \mathcal{O}(f/N_{\min})$ captures the steady-state perturbation induced by Byzantine edges. If the attack-update interval satisfies (25), the dissipation rate dominates the injection term, and $\|y_i - y_j\| \leq \epsilon_{\text{BEA}}$ for all normal agents. □

The results show that the proposed two-layer framework ensures fixed-time stability and bounded performance under switching topologies, dynamic membership, and persistent Byzantine attacks.

6. CONCLUSION

This paper proposes a unified framework for fixed-time leaderless formation control of open multi-agent systems under Byzantine edge attacks. By combining a coordinate-independent vector MSR estimator with a nonlinear fixed-time controller, the framework ensures robust and predictable convergence without requiring a global leader or fixed topology. The two-layer design enables secure and adaptive coordination under switching topologies, dynamic membership, and adversarial disturbances. Future work will consider nonlinear and heterogeneous dynamics, as well as resource-efficient communication schemes for large-scale systems.

REFERENCES

Almuzakki, M.Z. and Jayawardhana, B. (2023). Cooperative nearest-neighbor control of multi-agent systems: Consensus and formation control problems. *IEEE Control Syst. Lett.*, 7, 1873–1878.

Amirian, N. and Shamaghdari, S. (2021). Distributed resilient flocking control of multi-agent systems through event/self-triggered communication. *IET Control Theory Appl.*, 15(4), 559–569.

Gong, X., Cui, Y., Shen, J., Feng, Z., and Huang, T. (2020). Necessary and sufficient conditions of formation-containment control of high-order multiagent systems with observer-type protocols. *IEEE Trans. Cybern.*, 52(7), 7002–7016.

LeBlanc, H.J., Zhang, H., Koutsoukos, X., and Sundaram, S. (2013). Resilient asymptotic consensus in robust networks. *IEEE J. Sel. Areas Commun.*, 31(4), 766–781.

Liu, Y., Xie, X., Chadli, M., and Sun, J. (2024). Leaderless consensus control of fractional-order nonlinear multiagent systems with measurement sensitivity and actuator attacks. *IEEE Trans. Control Netw. Syst.*, 11(4), 2252–2262.

Shang, Y. (2020). Resilient consensus in multi-agent systems with state constraints. *Automatica*, 122, 109288.

Sun, H.T., Chen, X., Zhang, Z., Ge, X., and Peng, C. (2024). Data-driven event-triggered sliding mode secure control for autonomous vehicles under actuator attacks. *IEEE Trans. Cybern.*

Usevitch, J. and Panagou, D. (2019). Resilient leader-follower consensus to arbitrary reference values in time-varying graphs. *IEEE Trans. Autom. Control*, 65(4), 1755–1762.

Wang, J., Deng, X., Guo, J., and Zeng, Z. (2023). Resilient consensus control for multi-agent systems: A comparative survey. *Sensors*, 23(6), 2904.

Xiao, S. and Dong, J. (2021). Distributed fault-tolerant containment control for nonlinear multi-agent systems under directed network topology via hierarchical approach. *IEEE/CAA J. Autom. Sin.*, 8(4), 806–816.

Zhang, D., Chen, H., Lu, Q., Deng, C., and Feng, G. (2025). Finite-time cooperative output regulation of heterogeneous nonlinear multi-agent systems under switching dos attacks. *Automatica*, 173, 112062.