Strategic Topology Switching for Secure Control of Multi-Agent Systems

Yanbing Mao, Emrah Akyol, and Ziang Zhang

***Abstract*—In this technical note, we study secure coordination control for second-order multi-agent systems under a simplified control protocol and a special class of stealthy false data injection attacks, namely the "zero-dynamics" attack (ZDA). The proposed mathematical tool is to strategically switch network topology to detect ZDA. We first characterize a sufficient condition on switching times in reaching the second-order consensus. We then obtain the necessary and sufficient conditions on switching topologies in detecting ZDA. Employing the strategies on switching times and topologies, we then propose an attack detection algorithm based on the Luenberger observer. The advantages of the proposed method are three-fold: (i) in achieving consensus in the absence of ZDA, the control protocol does not need velocity measurements or impose any constraint on the magnitudes of coupling weights; (ii) in tracking system in the absence of ZDA, the Luenberger observer does not impose any constraint on the magnitudes of observer gains or on the number of monitored agents; (iii) in detecting ZDA, the knowledge of attack-starting time and misbehaving agents is not needed.**

***Index Terms*—Multi-agent systems, zero-dynamics attack, strategic topology switching, attack detection.**

I. Introduction

Several current and emerging multi-agent systems, such as connected vehicles, spacecrafts and robots, rely on the second-order dynamics. Substantial research efforts have been devoted to the second-order consensus of multi-agent systems, which can represent many coordination protocols, see e.g., flocking [1], swarming [2], velocity synchronization and regulation of relative distances [3], among many others. Beyond coordination control [4], the system designer has another important objective: security. Security of networked systems is a difficult task as highlighted by several recent incidents, including the distributed denial-of-service (DDOS) attack on Estonian web sites [5], and cyber attacks on smart grids [6]. Particularly, a special class of false data injection attacks namely the "zero-dynamics attack" (ZDA), poses a significant security challenge [7]–[9], since it hides the attack signal in the null-space of state-space representation and hence evades detection via conventional methods. Such stealthy attacks require design principles with both security and control goals in mind.

Recent research focused on variations of ZDA and associated defense strategies. In [10], Kim et al. proposed a discretized ZDA for the sampled-data control systems, where the conventional attack-detection signal is constantly zero at the sampling times, hence the proposed ZDA evades detection. Park et al. [11] designed a robust ZDA for stochastic cyber-physical systems, where the objective of the attacker is to guarantee that attack-detection signal stays below a threshold over a finite horizon. Jafarnejadsani et al. in [8] and Naghnaeian et al. in [12] proposed a multi-rate adaptive controller to detect ZDA in the sampled-data control systems, by removing certain unstable zeros of discrete-time systems. Back et al. in [13] utilized "generalized hold" to render impact of bounded ZDA.

While developing defense strategies for ZDAs in networked systems have recently gained interest [14]–[17], the space of solutions is yet to be thoroughly explored. The common features of the defense strategies include the imposing constraints on the connectivity of network topology and the number of the misbehaving agents (i.e., the agents under attack) [14]–[17], and the knowledge of the system about the attacker such as the number of misbehaving agents and the attack-starting time (set as the initial time) [14]–[16], [18]. One of our primary objectives here is to remove such constraints by utilizing a new defense/detection strategy: strategic topology switching.

This strategy is motivated by the idea of creating a mismatch between the models of attacker and defender, as first explored by Teixeira et al. in [18] for the ZDA problem. Specifically, in [18] the attacker uses the original system dynamics to design ZDA, while the defender strategically changes the system dynamics, however *only once* to detect the attack. In [18], as done in all prior work, the attacker is unnecessarily restricted to start the attack at the initial time, and also it is assumed to be unaware of a simple defense strategy of changing dynamics. An intuitive defense strategy against an informed attacker (who might be aware of one change in the dynamics) would be changing the dynamics not only once but infinitely many times (over an infinite time horizon), which suggests a dynamic topology design approach pursued in this note.

Intuitionally, another question pertaining to the dynamic defense strategy and protocol arises: can the second-order consensus protocol be simplified under the dynamic topology switching? We note that the conventional control protocols that achieve second-order consensus needs both the position and velocity measurements [19]. However, for automatous vehicles which are not equipped with velocity sensors to save cost, space and weight, its velocity cannot be precisely measured [20] as its estimation errors directly depends on the errors in 3D bounding box center coordinates. To address this issue, a few sampling-based consensus protocols that require only relative position measurements are proposed in [21]–[23],

Y. Mao is with the Departments of Computer Science and Mechanical Science and Engineering, University of Illinois at Urbana–Champaign, Urbana, IL, 61801 USA (e-mail: ybmao@illinois.edu).

E. Akyol and Z. Zhang are with the Department of Electrical and Computer Engineering, Binghamton University–SUNY, Binghamton, NY 13902 USA (e-mail: {eakyol, zhangzia}@binghamton.edu). The material in this technical note was partially presented at the 57th and 58th IEEE Conferences on Decision and Control, 2018 and 2019, respectively.

Limited circulation. For review only
IEEE-TAC Submission no.: 20-1698.1

which have constraints on coupling weights and sampling period. The answer to the arising question uncovers that the defense of dynamic topology switching aids to remove the need of position sampling and constraints on coupling weights for consensus in the absence of attack.

This note investigates the strategic topology switching for secure coordination control against ZDA. We first explore *when* topology should switch such that system of agents reaches the second-order consensus in the absence of attacks. We then analyze *what* topology to switch to detect ZDA. Our main contributions are summarized as follows.

- Necessary and sufficient conditions for detectability of the proposed ZDA variation under strategic topology switching are characterized.
- A Luenberger observer that tracks the multi-agent system in the absence of ZDA, without imposing any constraints on the number of monitored agents and observer gains, is proposed.
- Based on the characterized strategies on switching times and topologies, and through employing the Luenberger observer, a topology-switching algorithm for ZDA detection is proposed.

II. Preliminaries

A. Notation

We let $\mathbb{Q}$ denote the set of rational numbers. We use $P > (\geq)\ 0$ to denote a positive definite (positive semi-definite) matrix P. We let $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ denote the sets of n-dimensional real vectors and $m \times n$-dimensional real matrices, respectively. The symbol $\mathbb{N}$ represents the set of natural numbers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. We let $\mathbf{I}_n$ and $\mathbf{0}_{n\times n}$ be the $n\times n$-dimensional identity and zero matrices, respectively. $\mathbf{1}_n \in \mathbb{R}^n$ and $\mathbf{0}_n \in \mathbb{R}^n$ denote the vectors with all ones and all zeros, respectively. The superscript '$\top$' stands for matrix transpose. $\mu_P(A)$ denotes the induced P-norm matrix measure of $A \in \mathbb{R}^{n\times n}$, with $P > 0$, i.e., $\mu_P(A) = \frac{1}{2}\max\limits_{i=1,\ldots,n}\left\{\lambda_i\left(P^{1/2}AP^{-1/2} + P^{-1/2}A^\top P^{1/2}\right)\right\}$. $|\mathbb{V}|$ denotes the size of set $\mathbb{V}$. $\mathbb{V}\backslash\mathbb{K}$ denotes the complement set of $\mathbb{K}$ with respect to $\mathbb{V}$. $\mathfrak{S}(r)$ is the r^{th} element of ordered set $\mathfrak{S}$. lcm$(\cdot)$ is the least common multiple among scalars.

The interaction among n agents is modeled by an undirected graph $\mathrm{G} = (\mathbb{V}, \mathbb{E})$, where $\mathbb{V} = \{1, 2, \ldots, n\}$ is the set of agents and $\mathbb{E} \subset \mathbb{V}\times\mathbb{V}$ is the set of edges. The weighted adjacency matrix $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{n\times n}$ is defined as $a_{ij} = a_{ji} > 0$ if $(i, j) \in \mathbb{E}$, and $a_{ij} = a_{ji} = 0$ otherwise. The Laplacian matrix of an undirected graph G is defined as $\mathcal{L} = [l_{ij}] \in \mathbb{R}^{n\times n}$, where $l_{ii} = \sum\limits_{j=1}^{n} a_{ij}$ and $l_{ij} = -a_{ij}$ for $i \neq j$.

B. ZDA Definition

We consider the system: $\dot{z}(t) = Az(t)$, $y(t) = Cz(t)$, where $z(t) \in \mathbb{R}^{\bar{n}}$ and $y(t) \in \mathbb{R}^{\bar{m}}$ denote system state and monitored output, respectively; $A \in \mathbb{R}^{\bar{n}\times\bar{n}}$, $C \in \mathbb{R}^{\bar{m}\times\bar{n}}$. Its corresponding version under attack is described by

$$\dot{\tilde{z}}(t) = A\tilde{z}(t) + Bg(t), \quad \tilde{y}(t) = C\tilde{z}(t) + Dg(t), \tag{1}$$

where $g(t)$ is attack signal vector, $B \in \mathbb{R}^{\bar{n}\times\bar{o}}$ and $D \in \mathbb{R}^{\bar{m}\times\bar{o}}$. We next present the proposed ZDA policy whose attack-starting time can be not the initial time.

Definition 1: [24], [25] The attack signal

$$g(t) = \begin{cases} \underline{g}e^{\eta(t-\rho)}, t \in [\rho, \infty) \\ \mathbf{0}_{\bar{o}}, \qquad t \in [0, \rho) \end{cases} \tag{2}$$

in system (1) is a ZDA, if $\tilde{z}(0) - z(0) \neq \mathbf{0}_{\bar{n}}$, $g(\rho) \neq \mathbf{0}_{\bar{o}}$, $\rho \geq 0$ and $\eta \in \mathbb{C}$ satisfy

$$\tilde{z}(0) - z(0) \in \ker(\mathcal{O}), \text{ if } \rho > 0 \tag{3a}$$

$$\left[\begin{array}{c:c} e^{A\rho}(\tilde{z}(0) - z(0)) & \\ -\underline{g} & \end{array}\right] \in \ker\left(\left[\begin{array}{c:c} \eta\mathbf{1}_{\bar{n}\times\bar{n}} - A & B \\ -C & D \end{array}\right]\right), \tag{3b}$$

where $\mathcal{O} \triangleq \left[\begin{array}{c:c:c:c} C^\top & (CA)^\top & \ldots & \left(CA^{\bar{n}-1}\right)^\top \end{array}\right]^\top$.

Remark 1: To launch ZDA, the attacker must modify initial condition; otherwise, $\tilde{z}(0) - z(0) = \mathbf{0}_{\bar{n}}$, $e^{A\rho}\tilde{z}(0) - z(0) = \mathbf{0}_{\bar{n}}$, which with (3b) implies that $B\underline{g} = \mathbf{0}_{\bar{n}}$ and $D\underline{g} = \mathbf{0}_{\bar{m}}$. Thus, the attack signal (3b) does not have any effect on the system (1). The condition (3a) is to guarantee that false data injected to initial condition does not influence monitored outputs for stealthiness.

III. Problem Formulation

A second-order multi-agent system consists of a population of n agents whose dynamics are governed by:

$$\dot{x}_i(t) = v_i(t), \quad \dot{v}_i(t) = u_i(t), \ i = 1, \ldots, n \tag{4}$$

where $x_i(t) \in \mathbb{R}$ is the position, $v_i(t) \in \mathbb{R}$ is the velocity, and $u_i(t) \in \mathbb{R}$ is the control protocol of agent i. We consider the following simplified control protocol without velocity measurements to achieve the second-order consensus, i.e., $\lim\limits_{t\to\infty}|x_i(t) - x_j(t)| = 0$ and $\lim\limits_{t\to\infty}|v_i(t) - v_j(t)| = 0$, $i, j \in \mathbb{V}$.

$$u_i(t) = \sum_{j=1}^{n} a_{ij}(x_j(t) - x_i(t)), \quad i = 1, 2, \ldots, n. \tag{5}$$

For simplicity, we let the increasingly ordered set $\mathbb{M} \triangleq \{1, 2, \ldots\} \subseteq \mathbb{V}$ denote the set of monitored agents. We refer to an agent under attack as a *misbehaving agent* [15], and let $\mathbb{K} \subseteq \mathbb{V}$ denote the set of misbehaving agents.

A. System in the Absence of ZDA

The system (4) with (5) under switching topologies is

$$\dot{x}_i(t) = v_i(t) \tag{6a}$$

$$\dot{v}_i(t) = \sum_{j=1}^{n} a_{ij}^{\sigma(t)}(x_j(t) - x_i(t)), i = 1, \ldots, n \tag{6b}$$

where we refer to $a_{ij}^{\sigma(t)} \geq 0$ as the coupling weights, and

- $\sigma(t) : [t_0, \infty) \to \mathfrak{S} \triangleq \{1, 2, \ldots, s\}$, is the switching signal of the interaction topology of the network;
- $a_{ij}^{\sigma_k}$ is the entry of the weighted adjacency matrix that describes the activated σ_k topology over the time interval $[t_k, t_{k+1}), k \in \mathbb{N}_0$.

We define the fluctuation terms $\widehat{x}_i(t) \triangleq x_i(t) - \frac{1}{n}\sum\limits_{i=1}^{n} x_i(t)$ and $\widehat{v}_i(t) \triangleq \dot{\widehat{x}}_i(t)$. Then the dynamics (6) is rewritten as

$$\dot{\widehat{x}}_i(t) = \widehat{v}_i(t) \tag{7a}$$

$$\dot{\widehat{v}}_i(t) = \sum_{j=1}^{n} a_{ij}^{\sigma(t)}(\widehat{x}_j(t) - \widehat{x}_i(t)), i = 1, \ldots, n. \tag{7b}$$

B. System in the Presence of ZDA

The system (6) under ZDA is described by

$$\dot{\tilde{x}}_i(t) = \tilde{v}_i(t) \tag{8a}$$

$$\dot{\tilde{v}}_i(t) = \sum_{i=1}^{n} a_{ij}^{\tilde{\sigma}(t)}(\tilde{x}_j(t) - \tilde{x}_i(t)) + \begin{cases} \tilde{g}_i(t), i \in \mathbb{K} \\ 0, \ i \in \mathbb{V}\backslash\mathbb{K} \end{cases} \tag{8b}$$

$$\tilde{y}_j(t) = \tilde{x}_j(t), j \in \mathbb{M} \tag{8c}$$

where $\tilde{y}_j(t)$ is monitoring output and $\tilde{g}_i(t)$ is the ZDA signal in the form of (2):

$$\tilde{g}_i(t) = \begin{cases} \underline{g}_i e^{\eta(t-\rho)}, t \in [\rho, \infty) \\ 0, \qquad \text{otherwise.} \end{cases} \tag{9}$$

We note that system (8) is equivalent to a switched system:

$$\dot{\tilde{z}}(t) = A_{\tilde{\sigma}(t)}\tilde{z}(t) + g(t), \quad \tilde{y}(t) = C\tilde{z}(t) \tag{10}$$

where we define

$$A_{\tilde{\sigma}(t)} \triangleq \left[\begin{array}{c|c} \mathbf{0}_{n\times n} & \mathbf{I}_n \\ \hline -\mathcal{L}_{\tilde{\sigma}(t)} & \mathbf{0}_{n\times n} \end{array}\right], \tag{11a}$$

$$C \triangleq \left[\mathrm{e}_1 \ \vdots \ \dots \ \vdots \ \mathrm{e}_{|\mathbb{M}|} \ \vdots \ \mathbf{0}_{|\mathbb{M}|\times(2n-|\mathbb{M}|)}\right]^\top, \tag{11b}$$

$$\tilde{z}(t) \triangleq \left[\tilde{x}_1(t) \ \vdots \ \dots \ \vdots \ \tilde{x}_n(t) \ \vdots \ \tilde{v}_1(t) \ \vdots \ \dots \ \vdots \ \tilde{v}_n(t)\right]^\top, \tag{11c}$$

$$g(t) \triangleq \left[\mathbf{0}_n^\top \ \vdots \ a^\top(t)\right]^\top, \tag{11d}$$

$$a_i(t) \triangleq \begin{cases} \tilde{g}_i(t), i \in \mathbb{K} \\ 0, \quad i \in \mathbb{V}\backslash\mathbb{K} \end{cases} \tag{11e}$$

with $\mathrm{e}_i \in \mathbb{R}^n$ denoting the i^{th} vector of the canonical basis. The system (6) is equivalent to a switched system:

$$\dot{z}(t) = A_{\sigma(t)}z(t), \quad y(t) = Cz(t). \tag{12}$$

We make the following assumptions throughout this note.

Assumption 1: The attacker 1) knows currently activated topology at t_k with its dwell time $t_{k+1} - t_k$, 2) is aware of the defense strategy of changing system dynamics, 3) has the memory of the past topology switching sequences.

Assumption 2: The defender 1) designs the topology switching sequences including switching times and topologies, 2) has no knowledge of the attack-starting time; 3) has no knowledge of the number of misbehaving agents.

C. Secure Control Problems

The secure control problems are formally stated as follows.

Problem I: When should the topology switch such that the agents in system (6) reach consensus in the absence of an attack?

Problem II: What is the set of topologies to switch to such that ZDA in system (8) can be detected without the knowledge of misbehaving agents and attack-starting time?

IV. Defense Strategy

The defense strategy consists of a strategy on switching times and a strategy on switching topologies that corresponds to Problems I and II, respectively.

A. Problem I: Strategy on Switching Times

We present a property of system (7) in following lemma.

Lemma 1: [26], [27] Consider the following system

$$\dot{\widehat{x}}(t) = -\mathcal{L}_r \int_0^t \widehat{x}(\tau)\mathrm{d}\tau + \widehat{v}(0), t \geq 0 \tag{13}$$

where $\mathcal{L}_r \in \mathbb{R}^{n\times n}$ is the Laplacian matrix of a connected undirected graph; and $\widehat{x}_i(t) = x_i(t) - \frac{1}{n}\sum_{i=1}^{n} x_i(t)$ and $\widehat{v}_i(t) = \dot{\widehat{x}}_i(t)$. The solution of $\widehat{x}_i(t)$, $i = 1, \dots, n$, is

$$\widehat{x}_i(t) = \sum_{l=2}^{2} q_{li}q_l^\top\left(\widehat{x}(t)\cos(t\sqrt{\lambda_l(\mathcal{L}_r)}) + \frac{\widehat{v}(0)\sin(t\sqrt{\lambda_l(\mathcal{L}_r)})}{\sqrt{\lambda_l(\mathcal{L}_r)}}\right)$$

where $q_l = [q_{l1}, \dots, q_{ln}]^\top \in \mathbb{R}^n$ are the orthogonal vectors associated with eigenvalues $\lambda_l(\mathcal{L}_r)$ ($\lambda_1(\mathcal{L}_r) = 0$), $l = 2, \dots, n$.

The system under switching topology (7) can be viewed as a switched linear system, whose equilibrium point is $(\widehat{x}^*, \widehat{v}^*) = (\mathbf{0}_n, \mathbf{0}_n)$. Let $\sigma(t) = r \in \mathfrak{S}$ for $t \in [t_k, t_{k+1})$, $k \in \mathbb{N}_0$, the subsystem of (7) can be rewritten as $\dot{\widehat{x}}(t) = -\mathcal{L}_r\int_{t_k}^{t}\widehat{x}(\tau)\mathrm{d}\tau + \widehat{v}(t_k), t \in [t_k, t_{k+1})$. Then, Lemma 1 implies that the system (7) under each topology is oscillating. Hence, Problem I simplifies to designing a stabilizing switching rule for a switched system without stable subsystems. We first recall a prior result on this subject.

Lemma 2: [28] Consider a switched linear system

$$\dot{z}(t) = A_{\sigma(t)}z(t), \tag{14}$$

where $z(t) \in \mathbb{R}^m$, $A_{\sigma(t)} \in \mathbb{R}^{m\times m}$ and $\sigma(t) \in \mathfrak{S}$. Given scalars $\alpha > 0$, $1 > \beta > 0$, $\widehat{\tau}_{\max} \geq \widehat{\tau}_{\min} > 0$ and $\kappa \in \mathbb{N}$, if there exists a set of matrices $P_{r,q} > 0, q = 0, 1, \dots, \kappa$, $r \in \mathfrak{S}$, such that $\forall q = 0, 1, \dots, \kappa - 1$, $\forall r, s \in \mathfrak{S}$, such that

$$A_r^\top P_{r,q} + P_{r,q}A_r + \Psi_r^q - \alpha P_{r,q} < 0, \tag{15}$$

$$A_r^\top P_{r,q+1} + P_{r,q+1}A_r + \Psi_r^q - \alpha P_{r,q+1} < 0, \tag{16}$$

$$A_r^\top P_{r,\kappa} + P_{r,\kappa}A_r - \alpha P_{r,\kappa} < 0, \tag{17}$$

$$P_{s,0} - \beta P_{r,\kappa} \leq 0, s \neq r \tag{18}$$

$$\ln\beta + \alpha\widehat{\tau}_{\max} < 0, \tag{19}$$

where $\Psi_r^q = \frac{\kappa(P_{r,q+1} - P_{r,q})}{\widehat{\tau}_{\min}}$, then the system (14) is globally uniformly asymptotically stable under any switching signal $\sigma(t)$ satisfying

$$\widehat{\tau}_{\min} \leq t_{k+1} - t_k \leq \widehat{\tau}_{\max}, \forall k \in \mathbb{N}_0. \tag{20}$$

We note however that, in its current form, Lemma 2 cannot be applied to our system (7), as revealed in [27]. The solution in Lemma 1 implies that under the condition:

$$\forall r \in \mathfrak{S} : \sqrt{\frac{\lambda_i(\mathcal{L}_r)}{\lambda_j(\mathcal{L}_r)}} \in \mathbb{Q}, \ \text{for } \forall i, j = 2, \dots, n \tag{21}$$

the state of system (7) under fixed topology has a period T_r:

$$T_r = \mathrm{lcm}\left(\frac{2\pi}{\sqrt{\lambda_i(\mathcal{L}_r)}}; i = 2, ..., n\right), \tag{22}$$

such that

$$\begin{cases} \tilde{v}(t) = -\tilde{v}\left(t + \frac{T_r}{2}\right), \\ \tilde{x}(t) = -\tilde{x}\left(t + \frac{T_r}{2}\right), \sigma(t) = r \in \mathfrak{S} \text{ for } t \in [t_k, t_{k+1}). \end{cases} \tag{23}$$

The period $T_{\sigma(t_k)}$ can be used to make Lemma 2 applicable to the multi-agent system (7) to derive a strategy on the switching times, as stated in the following theorem, whose proof is presented in Appendix A.

Limited circulation. For review only
IEEE-TAC Submission no.: 20-1698.1

Theorem 1: Consider the second-order multi-agent system (6). For the given topology set $\mathfrak{S}$ satisfying (21), the period $T_{\sigma(t_k)}$ computed by (22), scalars $1>\beta>0$, $\alpha>0$ and $\kappa\in\mathbb{N}$, if the switching times satisfy

$$t_{k+1}-t_k=\widehat{\tau}_{\max}+\mathrm{m}\frac{T_{\sigma(t_k)}}{2},\quad k\in\mathbb{N}_0,\mathrm{m}\in\mathbb{N} \tag{24}$$

where

$$0<\widehat{\tau}_{\max}<\frac{-\ln\beta}{\alpha}, \tag{25}$$

$$0<\widehat{\tau}_{\max}+\mathrm{m}\frac{T_{\sigma(t_k)}}{2}-\left(\beta^{-\frac{1}{\kappa}}-1\right)\frac{\kappa}{\alpha-\xi}, \tag{26}$$

$$\xi<\alpha, \tag{27}$$

$$\xi=\max_{r\in\mathfrak{S},i=1,\ldots,n}\left\{1-\lambda_i\left(\mathcal{L}_r\right),-1+\lambda_i\left(\mathcal{L}_r\right)\right\}, \tag{28}$$

then the second-order consensus is achieved.

B. Problem II: Strategy on Switching Topologies

We recall a few definitions for the presentation of strategy.

Definition 2: [15] Consider the systems (10) and (12). The attack signal $g(t)$ in the system (10) is said to be undetectable if $\tilde{y}(t)=y(t)$ for any $t\geq 0$.

Definition 3 (Components of Graph [29]): The *components* of a graph are its maximal connected subgraphs.

Definition 4: The difference graph $\mathrm{G}_{\text{diff}}^{rs}=(\mathbb{V}_{\text{diff}}^{rs},\mathbb{E}_{\text{diff}}^{rs})$ of two graphs G_r and G_s is generated as

$$\mathbb{V}_{\text{diff}}^{rs}=\mathbb{V}_r\cup\mathbb{V}_s,\quad (i,j)\in\mathbb{E}_{\text{diff}}^{rs}\text{ if }a_{ij}^r-a_{ij}^s\neq 0, \tag{29}$$

where $\mathbb{V}_r$ and a_{ij}^r are the set of vertices (agents) and the entry of weighted adjacency matrix of the graph G_r, respectively.

We now define the union of difference graphs of every two switching graphs in (29) as: $\mathrm{G}_{\text{diff}}\triangleq\left(\bigcup_{r,s\in\mathfrak{S}}\mathbb{V}_{\text{diff}}^{rs},\bigcup_{r,s\in\mathfrak{S}}\mathbb{E}_{\text{diff}}^{rs}\right)$. We use $\mathbb{C}_i(\mathrm{G}_{\text{diff}})$ to denote the set of agents in i^{th} component of union difference graph G_{diff}. Obviously, $\mathbb{V}=\mathbb{C}_1\bigcup\mathbb{C}_2\bigcup\ldots\bigcup\mathbb{C}_{\mathrm{d}}$, and $\mathbb{C}_{\mathrm{p}}\bigcap\mathbb{C}_{\mathrm{q}}=\emptyset$ if $\mathrm{p}\neq\mathrm{q}$, where d is the number of total components in graph G_{diff}.

In the following theorem, whose proof is in Appendix B, we characterize the condition on topology set for detectability.

Theorem 2: Consider the multi-agent systems (6) and (8). Under time-dependent topology switching, i.e., $\sigma(t)=\tilde{\sigma}(t)$ for any $t\geq t_0$, ZDA is detectable without knowledge of the number of misbehaving agents or the attack-starting time (i.e., $\tilde{y}(t)=y(t)$ for any $t\geq t_0$ does not hold), if and only if

$$\mathbb{C}_i(\mathrm{G}_{\text{diff}})\cap\mathbb{M}\neq\emptyset,\forall i=1,\ldots\mathrm{d}. \tag{30}$$

V. Attack Detection Algorithm

We present a Luenberger observer [30] for the system (8) for the attack detection.

$$\dot{\underline{x}}_i(t)=\underline{v}_i(t) \tag{31a}$$

$$\dot{\underline{v}}_i(t)=\sum_{i=1}^{n}a_{ij}^{\sigma(t)}(\underline{x}_j(t)-\underline{x}_i(t))-\begin{cases}\psi_i r_i(t)+\theta_i\dot{r}_i(t),i\in\mathbb{M}\\0,i\in\mathbb{V}\backslash\mathbb{M}\end{cases} \tag{31b}$$

$$r_i(t)=\underline{x}_i(t)-\tilde{y}_i(t),i\in\mathbb{V}\backslash\mathbb{M} \tag{31c}$$

where $\tilde{y}_i(t)$ is the output of monitored agent i in system (8), $r_i(t)$ is the attack-detection signal, ψ_i and θ_i are the observer gains designed by defender.

We define the tracking errors as $e_x(t)\triangleq\underline{x}(t)-\tilde{x}(t)$ and $e_v(t)\triangleq\underline{v}(t)-\tilde{v}(t)$. A dynamics of tracking errors with attack-detection signal is obtained from (31) and (8):

$$\dot{e}_x(t)=e_v(t), \tag{32a}$$

$$\dot{e}_v(t)=-(\mathcal{L}_{\sigma(t)}+\Phi)e_x(t)-\Theta e_v(t)-a(t), \tag{32b}$$

$$r(t)=Ce_x(t), \tag{32c}$$

where $a(t)$ is defined in (11e), $r(t)\triangleq[r_1(t),\ldots,r_{|\mathbb{M}|}(t)]^\top$,

$$\Phi\triangleq\mathrm{diag}\{\psi_1,\ldots,\psi_{|\mathbb{M}|},0,\ldots,0\}\in\mathbb{R}^{n\times n}, \tag{33}$$

$$\Theta\triangleq\mathrm{diag}\{\theta_1,\ldots,\theta_{|\mathbb{M}|},0,\ldots,0\}\in\mathbb{R}^{n\times n}. \tag{34}$$

A. Luenberger Observer under Fixed Topology

In this section, we investigate the stability of tracking error dynamics (32) under a fixed topology. We let $\sigma(t)=s\in\mathfrak{S}$ denote the system (32) under fixed s^{th} topology. We denote the system matrix for (32) as

$$\mathcal{A}_s\triangleq\left[\begin{array}{c|c}\mathbf{0}_{n\times n}&\mathbf{I}_n\\\hline-\mathcal{L}_s-\Phi&-\Theta\end{array}\right]. \tag{35}$$

Theorem 2 implies that if the union difference graph is connected, using only one monitored agent's output is sufficient to detect ZDA. The following theorem, whose proof is available in Appendix C, characterizes the conditions under which only one monitored agent's output is sufficient for the observer (31) to asymptotically track the system (8) in the absence of ZDA.

Theorem 3: Consider the matrix $\mathcal{A}_s$ given by (35), where $\mathcal{L}_s$ is the Laplacian matrix of a connected undirected graph, the gain matrices Φ and Θ defined in (33) and (34) satisfy

$$\mathbf{0}_{n\times n}\neq\Phi\geq 0,\qquad\mathbf{0}_{n\times n}\neq\Theta\geq 0. \tag{36}$$

$\mathcal{A}_s$ is Hurwitz for any $|\mathbb{M}|\geq 1$, if and only if $\mathcal{L}_s$ has distinct eigenvalues.

B. Observer Stability under Switching Topologies

We next recall a known lemma can guarantee the stability of observer (31) under switching topologies.

Lemma 3: [31] Consider the switched systems: $\dot{x}(t)=\mathcal{A}_{\sigma(t)}x(t)$ under periodic switching, i.e., $\sigma(t)=\sigma(t+\tau)\in\mathfrak{S}$. If there exists a convex combination of some matrix measure $\mu(\cdot):\mathbb{R}^{n\times n}\to\mathbb{R}$ that satisfies

$$\sum_{m=1}^{L}\nu_m\mu\left(\mathcal{A}_m\right)<0, \tag{37}$$

where $\nu_m=\frac{\tau_m}{\tau}$ with $\tau=\sum_{m=1}^{L}\tau_i$, then the switched system system is uniformly asymptotically stable for every positive τ.

The defense strategy of strategic topology switching is described as follows.

Strategic Topology Switching Against ZDA

Switching topologies: (21);(30);$\exists r\in\mathfrak{S}$,$\mathcal{L}_r$ has distinct eigenvalues.
Switching times: (24);(37); $\sigma(t_k)\leftarrow\mathfrak{S}(\mathrm{mod}(k,|\mathfrak{S}|)+1)$.

In the following theorem, whose proof appears in Appendix D, we present our main results.

Theorem 4: Consider the multi-agent system (8) and the observer (31), where the observer gain matrices Φ and Θ satisfy (36). Under the strategic topology switching,

1) without knowledge of the misbehaving agents and the attack-starting time, the observer (31) is able to detect ZDA in system (8), i.e., $r(t) \equiv \mathbf{0}_{|\mathbb{M}|}$ does not holds;
2) in the absence of ZDA, the agents in system (8) reach the second-order consensus asymptotically, and without imposing any constraints on the magnitudes of observer gains and the number of monitored agents, the observer (31) asymptotically tracks the actual system (8).

VI. Conclusion

This technical note studies strategic topology switching for a second-order multi-agent system under a ZDA. For the simplified control protocol that does need velocity measurements, we propose a strategy on switching times that addresses the problem: when the topology should switch such that the second-order consensus can be achieved in the absence of attack. We then propose a strategy on switching topologies that addresses the problem: what topology to switch to, such that the ZDA can be detected. Based on the two strategies, a Luenberger observer based attack detector is proposed.

Acknowledgment

The authors would like to thank Prof. Naira Hovakimyan, Prof. Lui Sha and Ayoosh Bansal for discussions on secure control and velocity measurement of automatous vehicles.

Appendix A: Proof of Theorem 1

We recall that the dynamics of fluctuations (7) is equivalent to (6). Hence, in the proof we consider only the system (7). We should note that the multi-agent system (7) can be described by (14), where $A_{\sigma(t)}$ is given by (11a), and $z(t) \triangleq [\widehat{x}_1(t), \ldots, \widehat{x}_n(t), \widehat{v}_1(t), \ldots, \widehat{v}_n(t)]^\top \in \mathbb{R}^{2n}$.

It follows from (24)–(26) that the minimum and maximum dwell times defined as

$$\tau_{\min} \triangleq \min_{k\in\mathbb{N}_0} \{t_{k+1} - t_k\}, \quad \tau_{\max} \triangleq \max_{k\in\mathbb{N}_0} \{t_{k+1} - t_k\}, \quad (38)$$

that respectively satisfy $\tau_{\max} < \frac{-\ln\beta}{\alpha} + \mathrm{m}\frac{T_{\sigma(t_k)}}{2}$ and

$$\tau_{\min} > \left(\beta^{-\frac{1}{\kappa}} - 1\right)\frac{\kappa}{\alpha - \xi}. \quad (39)$$

For each topology, we consider the positive definite matrix

$$P_{r,q} \triangleq \left[\begin{array}{c|c} \hat{P}_{r,q} & \mathbf{0}_{n\times n} \\ \hline \mathbf{0}_{n\times n} & \hat{P}_{r,q} \end{array}\right] > 0, \quad (40)$$

$$\hat{P}_{r,q} \triangleq \beta^{-\frac{q}{\kappa}} h\mathbf{I}_n, \quad q = 0, \ldots, \kappa, \forall r \in \mathfrak{S}. \quad (41)$$

with $h > 0$. It follows from (41) that

$$\hat{P}_{r,q} \triangleq \beta^{\frac{1}{\kappa}} \hat{P}_{r,q+1}, \quad q = 0, \ldots, \kappa - 1, \forall r \in \mathfrak{S}, \quad (42)$$

$$\hat{P}_{s,0} \triangleq \beta \hat{P}_{r,\kappa}, \qquad \forall r \neq s \in \mathfrak{S}. \quad (43)$$

Substituting the matrices $P_{r,q}$ (40) and $A_{\sigma(t)}$ (11a) into conditions (15)–(17) yields

$$R_{r,q} \triangleq \left[\begin{array}{c|c} Q_{r,q} & (\mathbf{I}_n - \mathcal{L}_r)\hat{P}_{r,q} \\ \hline (\mathbf{I}_n - \mathcal{L}_r)\hat{P}_{r,q} & Q_{r,q} \end{array}\right] < 0, \quad (44)$$

$$\breve{R}_{r,q} \triangleq \left[\begin{array}{c|c} \breve{Q}_{r,q} & (\mathbf{I}_n - \mathcal{L}_r)\hat{P}_{r,q+1} \\ \hline (\mathbf{I}_n - \mathcal{L}_r)\hat{P}_{r,q+1} & \breve{Q}_{r,q} \end{array}\right] < 0, \quad (45)$$

$$S_{r,\kappa} \triangleq \left[\begin{array}{c|c} -\alpha\hat{P}_{r,\kappa} & (\mathbf{I}_n - \mathcal{L}_r)\hat{P}_{r,\kappa} \\ \hline (\mathbf{I}_n - \mathcal{L}_r)\hat{P}_{r,\kappa} & -\alpha\hat{P}_{r,\kappa} \end{array}\right] < 0, \quad (46)$$

where $Q_{r,q} \triangleq \frac{\kappa}{\tau_{\min}}(\hat{P}_{r,q+1} - \hat{P}_{r,q}) - \alpha\hat{P}_{r,q}$ and $\breve{Q}_{r,q} \triangleq \frac{\kappa}{\tau_{\min}}(\hat{P}_{r,q+1} - \hat{P}_{r,q}) - \alpha\hat{P}_{r,q+1}$.

We take W as the orthogonal matrix of $\mathcal{L}_r$, i.e.,

$$\Lambda_r \triangleq W^\top \mathcal{L}_r W = \mathrm{diag}\{0, \lambda_2(\mathcal{L}_r), \ldots, \lambda_n(\mathcal{L}_r)\}. \quad (47)$$

Then, considering the matrices $P_{r,q}$ and $\hat{P}_{r,q}$, in (40) and (41), the conditions (44)–(46) can be equivalently expressed as

$$\frac{\kappa}{\tau_{\min}}(\hat{P}_{r,q+1} - \hat{P}_{r,q}) - \alpha\hat{P}_{r,q} \pm (\mathbf{I}_n - \Lambda_r)\hat{P}_{r,q} < 0, \quad (48)$$

$$\frac{\kappa}{\tau_{\min}}(\hat{P}_{r,q+1} - \hat{P}_{r,q}) - \alpha\hat{P}_{r,q+1} \pm (\mathbf{I}_n - \Lambda_r)\hat{P}_{r,q+1} < 0, \quad (49)$$

$$-\alpha\hat{P}_{r,\kappa} \pm (\mathbf{I}_n - \Lambda_r)\hat{P}_{r,\kappa} < 0. \quad (50)$$

In the view of Lemma 2, to prove the second-order consensus, it suffices to verify that the conditions (15)–(19) are satisfied, as carried out in the following four steps.

Condition (18): It follows from the definitions in (40), (42), and (43) that $P_{s,0} = \beta P_{r,\kappa}, r \neq s \in \mathfrak{S}$.

Condition (19): Without loss of generality, we let $\sigma(t) = r \in \mathfrak{S}$ for $t \in [t_k, t_{k+1})$. To obtain Lemma 2, the considered discretized Lyapunov function for mode $r \in \mathfrak{S}$ in [28] is

$$V_r(t) \triangleq \begin{cases} z^\top(t) P_r^{(q)}(\zeta) z(t), t \in \mathfrak{N}_{k,q}, q = 0, 1, \ldots, \kappa - 1 \\ z^\top(t) P_{r,\kappa} z(t), \quad t \in [t_k + \tau_{\min}, t_{k+1}) \end{cases} \quad (51)$$

where $P_r^{(q)}(\zeta) \triangleq (1-\zeta)P_{r,q} + \zeta P_{r,q+1}$ with $\zeta = \frac{\kappa(t - t_k - \theta_q)}{\tau_{\min}}$, $\mathfrak{N}_{k,q} \triangleq [t_k + \theta_q, t_k + \theta_{q+1})$, $\theta_{q+1} \triangleq \frac{(q+1)\tau_{\min}}{\kappa}$, $P_{r,q} > 0$, $q = 0, 1, \ldots, \kappa - 1$. In [28], the purpose of the condition (19) is to guarantee that

$$V_{\sigma(t_k)}(t_{k+1}) \leq \beta^* V_{\sigma(t_k^-)}(t_k), \quad (52)$$

which is based on

$$V_{\sigma(t_k)}(t_k + \widehat{\tau}_{\max}) \leq \beta^* V_{\sigma(t_k^-)}(t_k), \quad (53)$$

with $1 > \beta^* > 0$. Noting that dwell time relation (24) is equivalent to $t_{k+1} = t_k + \widehat{\tau}_{\max} + \mathrm{m}\frac{T_r}{2}$, from (51) we have

$$V_r(t_{k+1}) = V_r(t_k + \widehat{\tau}_{\max} + \mathrm{m}\frac{T_r}{2}) \quad (54)$$
$$= z^\top(t_k + \widehat{\tau}_{\max} + \mathrm{m}\frac{T_r}{2}) P_{r,\kappa} z(t_k + \widehat{\tau}_{\max} + \mathrm{m}\frac{T_r}{2}), \forall r \in \mathfrak{S}.$$

Noting that $P_{r,\kappa} > 0$ and (23), we have

$$z^\top(t_k + \widehat{\tau}_{\max} + \mathrm{m}\frac{T_r}{2}) P_{r,\kappa} z(t_k + \widehat{\tau}_{\max} + \mathrm{m}\frac{T_r}{2}) \quad (55)$$
$$= z^\top(t_k + \widehat{\tau}_{\max}) P_{r,\kappa} z(t_k + \widehat{\tau}_{\max}) = V_r(t_k + \widehat{\tau}_{\max}), \forall r \in \mathfrak{S}.$$

Combining (54) with (55) yields

$$V_r(t_{k+1}) = V_r(t_k + \widehat{\tau}_{\max}), \forall r \in \mathfrak{S}. \quad (56)$$

We note that the condition (25) is equivalent to $\alpha\widehat{\tau}_{\max} + \ln\beta < 0$, which corresponds to the condition (19) in Lemma 2. Then, it follows from (53) and (56) that

$$V_{\sigma(t_k)}(t_{k+1}) = V_{\sigma(t_k)}(t_k + \widehat{\tau}_{\max}) \leq \beta^* V_{\sigma(t_k^-)}(t_k). \quad (57)$$

From (57) and (52), we conclude that the objective of $\mathrm{m}\frac{T_{\sigma(t_k)}}{2}, \mathrm{m} \in \mathbb{N}$, which is imposed on (24), is to maintain the original goal of the condition (19) through keeping (57) holding, while ensuring $\tau_{\max} \geq \tau_{\min}$, where $\tau_{\max}$ and $\tau_{\min}$

Limited circulation. For review only
IEEE-TAC Submission no.: 20-1698.1

are given in (38). This also means that it is the period $T_{\sigma(t_k)}$ that makes Lemma 2 applicable to (7).

Condition (17): Since $\hat{P}_{r,\kappa} > 0$, (27) implies $0 > -\alpha\hat{P}_{r,\kappa} + \xi\hat{P}_{r,\kappa}$, while (28) implies $\xi \geq \pm(\mathbf{I}_n - \Lambda_r)$. Thus,

$$0 > -\alpha\hat{P}_{r,\kappa} + \xi\hat{P}_{r,\kappa} > -\alpha\hat{P}_{r,\kappa} \pm (\mathbf{I}_n - \Lambda_r)\hat{P}_{r,\kappa}. \quad (58)$$

In light of (50), the condition (17) is satisfied.

Conditions (15) and (16): It follows from (27) and (39):

$$\frac{(\alpha-\xi)\tau_{\min}}{\kappa} + 1 > \beta^{-\frac{1}{\kappa}}. \quad (59)$$

Considering the fact of $h > 0$, from (42) and (59) we have $1 + \frac{(\alpha-\xi)\tau_{\min}}{\kappa} > \frac{\hat{P}_{r,q+1}}{\hat{P}_{r,q}} = \beta^{-\frac{1}{\kappa}}$, which is equivalent to

$$\frac{\kappa}{\tau_{\min}}(\hat{P}_{r,q+1} - \hat{P}_{r,q}) - (\alpha-\xi)\hat{P}_{r,q} < 0, q = 0, \ldots, \kappa-1. \quad (60)$$

Since $1 > \beta > 0$, (42) implies that $\hat{P}_{r,q} < \hat{P}_{r,q+1}$. Condition (27) implies that $\alpha - \xi > 0$. Therefore, (60) indicates

$$\frac{\kappa}{\tau_{\min}}(\hat{P}_{r,q+1} - \hat{P}_{r,q}) - (\alpha-\xi)\hat{P}_{r,q+1} < 0, \quad (61)$$

for $q = 0, \ldots, \kappa - 1$. By (28) and (58), we have:

$$0 > -\alpha\hat{P}_{r,q+1} + \xi\hat{P}_{r,q+1} > -\alpha\hat{P}_{r,q+1} \pm (\mathbf{I}_n - \Lambda_r)\hat{P}_{r,q+1}$$
$$0 > -\alpha\hat{P}_{r,q} + \xi\hat{P}_{r,q} > -\alpha\hat{P}_{r,q} \pm (\mathbf{I}_n - \Lambda_r)\hat{P}_{r,q},$$

which together with (60) and (61) imply (48) and (49), respectively. Thus, (15) and (16) in Lemma 2 hold.

Appendix B: Proof of Theorem 2

Let us first define $\breve{y}(t) \triangleq \tilde{y}(t) - y(t)$ and

$$\breve{z}(t) \triangleq \tilde{z}(t) - z(t) = \begin{bmatrix} \tilde{x}(t) \\ \tilde{v}(t) \end{bmatrix} - \begin{bmatrix} x(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} \breve{x}(t) \\ \breve{v}(t) \end{bmatrix}. \quad (62)$$

Under time-dependent topology switching, i.e., $\sigma(t) \equiv \tilde{\sigma}(t)$, from (10) and (12) we have:

$$\dot{\breve{z}}(t) = A_{\sigma(t)}\breve{z}(t) + g(t), \quad \breve{y}(t) = C\breve{z}(t),$$

based on which we define:

$$\mathcal{P}_r = \begin{bmatrix} \eta\mathbf{I}_{2n} - A_r & \mathbf{I}_{2n} \\ -C & \mathbf{0}_{|\mathbb{M}|\times 2n} \end{bmatrix}. \quad (63)$$

(Sufficient Condition) We now assume to the contrary that the system (10) is under ZDA. By Definition 1, we obtain $\left[\breve{z}^\top(\rho), -g^\top(\rho)\right]^\top \in \ker\left(\mathcal{P}_{\sigma(\rho)}\right)$. Meanwhile, we obtain from the condition (3) that $\tilde{z}(t) = \begin{cases} e^{At}\tilde{z}(0), & t \in [0,\rho] \\ z(t) + (\tilde{z}(\rho) - z(\rho))e^{\eta(t-\rho)}, & t \in (\rho,\infty) \end{cases}$, which together with (2) imply that $[\breve{z}(t), -g(t)] = e^{\eta(t-\rho)}[\breve{z}(\rho), -g(\rho)]$, $t \geq \rho$.

Since $e^{\eta(t-\rho)} \neq 0$, we conclude that ZDA is equivalent to

$$[\breve{z}^\top(\rho), -g^\top(\rho)]^\top \in \bigcap_{r\in\mathfrak{S}} \ker(\mathcal{P}_r). \quad (64)$$

Substituting C in (11b), $\breve{z}(\rho)$ in (62), $g(\rho)$ in (11d) with (11e), A_r in (11a), and $\mathcal{P}_r$ in (63) into (64) and expanding it out yields

$$\begin{bmatrix} \eta\mathbf{I}_n & -\mathbf{I}_n & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \\ \mathcal{L}_r & \eta\mathbf{I}_n & \mathbf{0}_{n\times n} & \mathbf{I}_n \\ -\mathbf{e}_j^\top & \mathbf{0}_n^\top & \mathbf{0}_n^\top & \mathbf{0}_n^\top \end{bmatrix} \begin{bmatrix} \breve{x}(\rho) \\ \breve{v}(\rho) \\ \mathbf{0}_n \\ -a(\rho) \end{bmatrix} = \begin{bmatrix} \mathbf{0}_n \\ \mathbf{0}_n \\ 0 \end{bmatrix}, \forall j \in \mathbb{M}, \forall r \in \mathfrak{S}$$

which is equivalent to

$$\eta\breve{x}(\rho) - \breve{v}(\rho) = \mathbf{0}_n, \quad (65)$$
$$-a(\rho) + \mathcal{L}_r\breve{x}(\rho) + \eta\breve{v}(\rho) = \mathbf{0}_n, \forall r \in \mathfrak{S} \quad (66)$$
$$-a(\rho) + \mathcal{L}_s\breve{x}(\rho) + \eta\breve{v}(\rho) = \mathbf{0}_n, \forall s \in \mathfrak{S} \quad (67)$$
$$\breve{x}_j(\rho) = 0, \forall j \in \mathbb{M}. \quad (68)$$

Through elementary row transformation, the Laplacian matrix of union difference graph G_{diff} can be written as

$$\tilde{\mathcal{L}} \triangleq \mathrm{diag}\left\{\mathcal{L}(\mathbb{C}_1(\mathrm{G}_{\text{diff}})), \ldots, \mathcal{L}(\mathbb{C}_{\mathrm{d}}(\mathrm{G}_{\text{diff}}))\right\}, \quad (69)$$

where $\mathcal{L}(\mathbb{C}_{\mathrm{q}}(\mathrm{G}_{\text{diff}}))$ denotes the Laplacian matrix of the q^{th} component for $q \in \{1, 2, \ldots, \mathrm{d}\}$. Noting that equation (66) subtracting equation (67) results in $(\mathcal{L}_r - \mathcal{L}_s)\breve{x}(\rho) = \mathbf{0}_n, \forall r, s \in \mathfrak{S}$, which is equivalent to

$$\tilde{\mathcal{L}}\breve{x}(\rho) = \mathbf{0}_n, \quad (70)$$

where $\tilde{\mathcal{L}}$ is defined in (69). From [32], it is known that the Laplacian matrix of component $\mathcal{L}(\mathbb{C}_{\mathrm{q}}(\mathrm{G}_{\text{diff}}))$ has properties: i) zero is one of its eigenvalues with multiplicity one, ii) the eigenvector that corresponds to the eigenvalue zero is $\mathbf{1}_{|\mathbb{C}_q(\mathrm{G}_{\text{diff}})|}$, $\forall \mathrm{q} \in \{1, \ldots, \mathrm{d}\}$. It follows from (68) and (30) that *under the strategy* (30), the solution of (70) is obtained as $\breve{x}(\rho) = \mathbf{0}_n$, which together with (65) imply that $\breve{v}(\rho) = \mathbf{0}_n$, substituting which into (66) or (67) yields the same result as $a(\rho) = \mathbf{0}_n$, which, in conjunction with (11d), implies $g(\rho) = \mathbf{0}_{2n}$, indicating there is no ZDA by Definition 1. Thus, a contradiction occurs.

(Necessary Condition) Substituting (65) into (66) yields $\left(\mathcal{L}_r + \eta^2\mathbf{I}_n\right)\breve{x}(\rho) = a(\rho), \forall r \in \mathfrak{S}$, which is equivalent to

$$\left(\sum_{r=1}^{|\mathfrak{S}|} \alpha_r\mathcal{L}_r + \eta^2\mathbf{I}_n\right)\breve{x}(\rho) = a(\rho), \forall \alpha_r > 0, \sum_{r=1}^{|\mathfrak{S}|} \alpha_r = 1. \quad (71)$$

To prove necessary condition, we need to prove that if the condition (30) does not hold, there exist feasible attack strategies. To achieve this goal, we consider two simple cases.

Case A: $(\mathrm{Im}(\eta) = 0, \mathrm{Re}(\eta) \neq 0)$*:* We assume that (30) does not hold, and without of loss generality, we let $\mathbb{C}_1(\mathrm{G}_{\text{diff}}) \cap \mathbb{M} = \emptyset$ and define $\breve{x}_{\mathbb{C}_1} \triangleq \left[\breve{x}_{i\in\mathbb{C}_1(\mathrm{G}_{\text{diff}})}\right] \in \mathbb{R}^{|\mathbb{C}_1(\mathrm{G}_{\text{diff}})|}$. Following (69) and (70), we obtain $\mathcal{L}(\mathbb{C}_1(\mathrm{G}_{\text{diff}}))\breve{x}_{\mathbb{C}_1}(\rho) = \mathbf{0}_{|\mathbb{C}_1(\mathrm{G}_{\text{diff}})|}$, which, in conjunction with the fact that the eigenvector corresponding to the eigenvalue zero of $\mathcal{L}(\mathbb{C}_1(\mathrm{G}_{\text{diff}}))$ is $\mathbf{1}_{|\mathbb{C}_1(\mathrm{G}_{\text{diff}})|}$, leads to nonzero solution $\breve{x}_{\mathbb{C}_1}(\rho)$. Consequently, the solution of (70) satisfies

$$\breve{x}(\rho) \neq \mathbf{0}_n. \quad (72)$$

The matrix $\sum_{r=1}^{|\mathfrak{S}|} \alpha_r\mathcal{L}_r + \eta^2\mathbf{I}_n$ is full-rank for nonzero $\eta \in \mathbb{R}$, and $\forall \alpha_r > 0$, $\sum_{r=1}^{|\mathfrak{S}|} \alpha_r = 1$. Then, considering (72), we obtain from (71) that $a(\rho) \neq \mathbf{0}_n$. Given nonzero $\eta \in \mathbb{R}$, we obtain from (72) and (65) that $\breve{v}(\rho) \neq \mathbf{0}_n$. We obtain a feasible ZDA consisting of nonzero η, $\breve{a}(\rho)$, $\breve{x}(\rho)$ and $\breve{v}(\rho)$.

Case B: $(\mathrm{Im}(\eta) = 0, \mathrm{Re}(\eta) = 0)$*:* We note that for the switching topologies in $\mathfrak{S}$, there exists one implied condition that is the union graph $\mathrm{G} \triangleq \left(\bigcup_{r\in\mathfrak{S}} \mathbb{V}_r, \bigcup_{r\in\mathfrak{S}} \mathbb{E}_r\right)$ is connected; otherwise, the asymptotic second-order consensus cannot be achieved, which is undesirable. Consequently, the eigenvector

associated with eigenvalue zero of $\sum_{r=1}^{|\mathfrak{S}|}\alpha_r\mathcal{L}_r$ is the only $\mathbf{1}_n$, for any $\alpha_r>0$, $\sum_{r=1}^{|\mathfrak{S}|}\alpha_r=1$. Thus, given $\breve{x}(\rho)$ that has non-identical entries, we obtain a non-zero vector $a(\rho)$ from (71) with $\eta=0$. Therefore, we obtain a feasible ZDA consisting of $\eta=0$, $\breve{v}(\rho)=\mathbf{0}_n$ (obtained from (65)), nonzero vector $\breve{a}(\rho)$, and vector $\breve{x}(\rho)$ that has non-identical entries.

Appendix C: Proof of Theorem 3

We let $\sigma(t)=s\in\mathfrak{S}$ for $t\in[t_k,t_{k+1}),k\in\mathbb{N}_0$. Since $\mathcal{L}_s$ is the Laplacian matrix of a connected graph and $\Phi\geq 0$, $\mathcal{L}_s+\Phi$ is positive definite. We define the following positive function for the system (32) with $a(t)\equiv\mathbf{0}_n$:

$$V_s(e(t))=\frac{1}{2}e_x^\top(t)(\mathcal{L}_s+\Phi)e_x(t)+e_v^\top(t)e_v(t).$$

Its time derivative is obtained as

$$\dot{V}_s(e(t))=-e_v^\top(t)\Theta e_v(t)\leq 0, \tag{73}$$

where the inequality is obtained by considering $\Theta\geq 0$. Since the dynamics (32) with $a(t)\equiv\mathbf{0}_n$ is equivalent to $\dot{e}(t)=\mathcal{A}_s e(t)$ with $e(t)\triangleq\left[\begin{array}{c|c}e_x^\top(t) & e_v^\top(t)\end{array}\right]^\top$, we conclude from (73) none of eigenvalues of $\mathcal{A}_r$ has positive real part.

We next prove $\mathcal{A}_r$ has neither zero nor purely imaginary eigenvalues. Using the formula $\det\left(\left[\begin{array}{c|c}A & B\\ \hline C & D\end{array}\right]\right)=\det(A)\det\left(D-CA^{-1}B\right)$, from (35) we have:

$$\begin{aligned}\det(\mathcal{A}_s-\lambda\mathbf{I}_{2n})&=\det\left(\left[\begin{array}{c|c}-\lambda\mathbf{I}_n & \mathbf{I}_n\\ \hline -\mathcal{L}_s-\Phi & -\Theta-\lambda\mathbf{1}_{n\times n}\end{array}\right]\right)\\&=\det(-\lambda\mathbf{I}_n)\det\left(-\Theta-\lambda\mathbf{I}_n-\frac{\mathcal{L}_s+\Phi}{\lambda}\right)\\&=\det\left(\lambda^2\mathbf{I}_n+\Theta\lambda+\mathcal{L}_s+\Phi\right).\end{aligned}\tag{74}$$

We define:

$$\phi_m\triangleq\sqrt{\psi_m+\lambda\theta_m}\mathbf{e}_m, \tag{75}$$

$$\mathcal{P}(m)\triangleq\lambda\widehat{\Theta}(m)+\widehat{\Phi}(m)=\sum_{p=m}^{|\mathbb{M}|}\phi_p\phi_p^\top, \tag{76}$$

where $\mathbf{e}_m\in\mathbb{R}^n$ is the m^{th} vector of the canonical basis, and

$$\begin{aligned}\widehat{\Theta}(m)&\triangleq\text{diag}\left\{0,\ldots,0,\theta_m,\ldots,\theta_{|\mathbb{M}|},0,\ldots 0\right\},\\\widehat{\Phi}(m)&\triangleq\text{diag}\left\{0,\ldots,0,\psi_m,\ldots,\psi_{|\mathbb{M}|},0,\ldots 0\right\}.\end{aligned}$$

It follows from (75), (76), (33) and (34) that

$$\Theta\lambda+\Phi=\sum_{p=1}^{m}\phi_p\phi_p^\top+\mathcal{P}(m+1). \tag{77}$$

We recall the well-known matrix determinant lemma [33]:

$$\det\left(A+\chi uw^\top\right)=\det(A)\left(1+\chi w^\top A^{-1}u\right), \tag{78}$$

where A is an invertible matrix, χ is a scalar and w and u are vectors of appropriate dimensions.

With the consideration of (77), we have $\lambda^2\mathbf{I}_n+\Theta\lambda+\mathcal{L}_s+\Phi=\sum_{p=1}^{m}\phi_p\phi_p^\top+\mathcal{H}_s+\mathcal{P}(m+1)$, where $\mathcal{H}_s\triangleq\lambda^2\mathbf{I}_n+\mathcal{L}_s$. Then, by (78), we obtain from (74) that

$$\begin{aligned}&\det(\mathcal{A}_s-\lambda\mathbf{I}_{2n})\\&=\det(\mathcal{H}_s+\mathcal{P}(2))(1+\phi_1^\top(\mathcal{H}_s+\mathcal{P}(2))^{-1}\phi_1)\\&=\det(\mathcal{H}_s+\mathcal{P}(3))(1+\phi_1^\top(\mathcal{H}_s+\mathcal{P}(2))^{-1}\phi_1)(1+\phi_2^\top(\mathcal{H}_s\\&\qquad+\mathcal{P}(3))^{-1}\phi_2)\\&=\ldots\\&=\det(\mathcal{H}_s)\prod_{m=1}^{|\mathbb{M}|}(1+\phi_m^\top(\mathcal{H}_s+\mathcal{P}(m+1))^{-1}\phi_m),\end{aligned}\tag{79}$$

with $\mathcal{P}(|\mathbb{M}|+1)=\mathbf{0}_{n\times n}$. Since $\mathcal{L}_r$ is symmetric, there exists an orthogonal matrix $Q\triangleq[q_1;\ldots;q_n]\in\mathbb{R}^{n\times n}$ with $q_i\triangleq\left[\begin{array}{c|c|c|c}q_{i1} & q_{i2} & \ldots & q_{in}\end{array}\right]^\top\in\mathbb{R}^n$, $i\in\mathbb{V}$, such that

$$Q^\top=Q^{-1}, \tag{80a}$$

$$Q^\top\mathcal{H}_sQ=\text{diag}\left\{\lambda^2+\lambda_1(\mathcal{L}_s),\ldots,\lambda^2+\lambda_n(\mathcal{L}_s)\right\}. \tag{80b}$$

Considering (36) and (34), without loss of generality, we let

$$\theta_{|\mathbb{M}|}\neq 0. \tag{81}$$

It follows from (80) and (75) that $\phi_{|\mathbb{M}|}^\top(\mathcal{H}_s)^{-1}\phi_{|\mathbb{M}|}=\sum_{i=1}^{n}\frac{(\psi_{|\mathbb{M}|}+\lambda\theta_{|\mathbb{M}|})q_{i|\mathbb{M}|}^2}{\lambda_i(\mathcal{L}_s)+\lambda^2}$ and $\det(\mathcal{H}_s)=\prod_{i=1}^{n}\left(\lambda_i(\mathcal{L}_s)+\lambda^2\right)$, from which we arrive at

$$\begin{aligned}\left(1+\phi_{|\mathbb{M}|}^\top(\mathcal{H}_s)^{-1}\phi_{|\mathbb{M}|}\right)\det(\mathcal{H}_s)&=\prod_{i=1}^{n}\left(\lambda_i(\mathcal{L}_s)+\lambda^2\right)\\&+\sum_{i=1}^{n}\prod_{j\neq i}^{n}\left(\lambda_j(\mathcal{L}_s)+\lambda^2\right)\left(\psi_{|\mathbb{M}|}+\lambda\theta_{|\mathbb{M}|}\right)q_{i|\mathbb{M}|}^2.\end{aligned}\tag{82}$$

We define:

$$\mathcal{Q}(\lambda)\triangleq\prod_{m=1}^{|\mathbb{M}|-1}\left(1+\phi_m^\top(\mathcal{H}_s+\mathcal{P}(m+1))^{-1}\phi_m\right). \tag{83}$$

Substituting (82) and (83) into (79) yields

$$\begin{aligned}\det(\mathcal{A}_s-\lambda\mathbf{I}_{2n})&=\mathcal{Q}(\lambda)\Bigg(\prod_{i=1}^{n}\left(\lambda_i(\mathcal{L}_s)+\lambda^2\right)\\&+\sum_{i=1}^{n}\prod_{j\neq i}^{n}\left(\lambda_j(\mathcal{L}_s)+\lambda^2\right)\left(\psi_{|\mathbb{M}|}+\lambda\theta_{|\mathbb{M}|}\right)q_{i|\mathbb{M}|}^2\Bigg),\end{aligned}\tag{84}$$

based on which, we consider two different cases.

Case One: assume $\mathcal{A}_s$ has zero eigenvalue: In this case, it follows from (75), (76), (83) and the conditions $\theta_m\geq 0$ and $\psi_m\geq 0$, $\forall m\in\mathbb{M}$, that $\mathcal{Q}(\lambda)>0$. We conclude from (84) that $\det(\mathcal{A}_s-\lambda\mathbf{I}_{2n})|_{\lambda=0}>0$, i.e., $\mathcal{A}_r$ does not have a zero eigenvalue.

Case Two: assume $\mathcal{A}_s$ has a purely imaginary eigenvalue: This case implies that $\lambda=\varpi\mathrm{i}$ with $0\neq\varpi\in\mathbb{R}$. It follows from (75) and (76) that

$$1+\phi_m^\top(\mathcal{H}_s+P(m+1))^{-1}\phi_m\neq 0,\forall m\in\mathbb{M}$$

thus, $\mathcal{Q}(\lambda)|_{\lambda=\varpi\mathrm{i}}\neq 0$. Then, we conclude from (84) that $\det(\mathcal{A}_s-\mathrm{i}\varpi\mathbf{I}_{2n})=0$ is equivalent to

$$\begin{aligned}&\prod_{i=1}^{n}\left(\lambda_i(\mathcal{L}_s)-\varpi^2\right)+\sum_{i=1}^{n}\prod_{j\neq i}^{n}\left(\lambda_j(\mathcal{L}_s)-\varpi^2\right)\psi_{|\mathbb{M}|}q_{i|\mathbb{M}|}^2\\&+\mathrm{i}\sum_{i=1}^{n}\prod_{j\neq i}^{n}\left(\lambda_j(\mathcal{L}_s)-\varpi^2\right)\varpi\theta_{|\mathbb{M}|}q_{i|\mathbb{M}|}^2=0.\end{aligned}\tag{85}$$

We note that (85) implies

$$\sum_{i=1}^{n}\prod_{j\neq i}^{n}\left(\lambda_j\left(\mathcal{L}_s\right)-\varpi^2\right)\varpi\theta_{|\mathbb{M}|}q_{i|\mathbb{M}|}^2=0, \tag{86}$$

which with (81) result in $\sum_{i=1}^{n}\prod_{j\neq i}^{n}\left(\lambda_j\left(\mathcal{L}_s\right)-\varpi^2\right)\varpi q_{i|\mathbb{M}|}^2=0$ which further implies that $\sum_{i=1}^{n}\prod_{j\neq i}^{n}\left(\lambda_j(\mathcal{L}_s)-\varpi^2\right)\psi_{|\mathbb{M}|}q_{i|\mathbb{M}|}^2=0$. Thus, from (85) we have $\prod_{i=1}^{n}\left(\lambda_i(\mathcal{L}_s)-\varpi^2\right)=0$, which means that $\exists i\in\{1,\ldots,n\}:\ \varpi^2=\lambda_i\left(\mathcal{L}_s\right)$, which indicates that if and only if $\mathcal{L}_s$ has distinct eigenvalues, $\sum_{i=1}^{n}\prod_{j\neq i}^{n}\left(\lambda_j\left(\mathcal{L}_s\right)-\varpi^2\right)\varpi\theta_{|\mathbb{M}|}q_{i|\mathbb{M}|}^2\neq 0$, substituting which into the left-hand side of (85) leads to $\det\left(\mathcal{A}_s-\mathrm{i}\varpi\mathbf{I}_{2n}\right)\neq 0$, which contradicts with assumption that $\mathcal{A}_s$ has a purely imaginary eigenvalue (that is equivalent to $\det\left(\mathcal{A}_s-\mathrm{i}\varpi\mathbf{I}_{2n}\right)=0$).

Appendix D: Proof of Theorem 4

Replacing $A_{\sigma(t)}$ by $\mathcal{A}_{\sigma(t)}$ (defined in (35)) in the steps to derive (65)–(68), we have

$$\eta e_x(\rho)-e_v(\rho)=\mathbf{0}_n, \tag{87}$$

$$a(\rho)+\mathcal{L}_r e_x(\rho)+\eta e_x(\rho)=\mathbf{0}_n,\forall r\in\mathfrak{S} \tag{88}$$

$$a(\rho)+\mathcal{L}_s e_x(\rho)+\eta e_x(\rho)=\mathbf{0}_n,\forall s\in\mathfrak{S} \tag{89}$$

$$e_{x_j}(\rho)=0,\forall j\in\mathbb{M}. \tag{90}$$

Therefore, the proof of 1) follows from Theorem 2.

In the absence of ZDA, the system matrix of system (32) is $\mathcal{A}_{\sigma(t)}$ defined in (35). In light of Theorem 3 and Lemma 3 and the conditions in (36), $\mathcal{A}_s$ is Hurwitz. Thus, there exists $P>0$ such that $\mu_P\left(\mathcal{A}_s\right)<0$. Through setting on the switching times (dwell times) by (24), (37) can be satisfied. By Lemma 3, the switched linear system (32) is uniformly asymptotically stable. The claim of consensus directly follows from Theorem 1.

References

[1] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: Algorithms and theory," *IEEE Transactions on automatic control*, vol. 51, no. 3, pp. 401–420, 2006.

[2] V. Gazi and K. M. Passino, "Stability analysis of social foraging swarms," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 539–557, 2004.

[3] H. G. Tanner, A. Jadbabaie, and G. J. Pappas, "Flocking in fixed and switching networks," *IEEE Transactions on Automatic control*, vol. 52, no. 5, pp. 863–868, 2007.

[4] K. You and L. Xie, "Network topology and communication data rate for consensusability of discrete-time multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2262–2275, 2011.

[5] J. Nazario, "Politically motivated denial of service attacks," *The Virtual Battlefield: Perspectives on Cyber Warfare*, pp. 163–181, 2009.

[6] J. Meserve, "Sources: Staged cyber attack reveals vulnerability in power grid," http://www.cnn.com/2007/US/09/26/power.at.risk/, accessed 2007-09-26.

[7] M. Naghnaeian, N. Hirzallah, and P. G. Voulgaris, "Dural rate control for security in cyber-physical systems," in *Proceedings of the 54th IEEE Conference on Decision and Control*, pp. 1415–1420, 2015.

[8] H. Jafarnejadsani, H. Lee, N. Hovakimyan, and P. Voulgaris, "A multirate adaptive control for mimo systems with application to cyber-physical security," in *Proceedings of the 57th IEEE Conference on Decision and Control*, pp. 6620–6625, 2018.

[9] N. H. Hirzallah and P. G. Voulgaris, "On the computation of worst attacks: a lp framework," in *2018 Annual American Control Conference*, pp. 4527–4532, 2018.

[10] J. Kim, G. Park, H. Shim, and Y. Eun, "Zero-stealthy attack for sampled-data control systems: The case of faster actuation than sensing," in *Proceedings of the 55th IEEE Conference on Decision and Control*, pp. 5956–5961, 2016.

[11] G. Park, H. Shim, C. Lee, Y. Eun, and K. H. Johansson, "When adversary encounters uncertain cyber-physical systems: Robust zero-dynamics attack with disclosure resources," in *Proceedings of the 55th IEEE Conference on Decision and Control*, pp. 5085–5090, 2016.

[12] M. Naghnaeian, N. H. Hirzallah, and P. G. Voulgaris, "Security via multirate control in cyber–physical systems," *Systems & Control Letters*, vol. 124, pp. 12–18, 2019.

[13] J. Back, J. Kim, C. Lee, G. Park, and H. Shim, "Enhancement of security against zero dynamics attack via generalized hold," in *Proceedings of the 56th IEEE Conference on Decision and Control*, pp. 1350–1355, 2017.

[14] S. Sundaram and C. N. Hadjicostis, "Distributed function calculation via linear iterative strategies in the presence of malicious agents," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1495–1508, 2011.

[15] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 90–104, 2012.

[16] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[17] S. Weerakkody, X. Liu, and B. Sinopoli, "Robust structural analysis and design of distributed control systems to prevent zero dynamics attacks," in *Proceedings of the 56th IEEE Conference on Decision and Control*, pp. 1356–1361, 2017.

[18] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in *Proceedings of the 50th Allerton Conference on Communication, Control, and Computing*, pp. 1806–1813, 2012.

[19] J. Mei, W. Ren, and J. Chen, "Distributed consensus of second-order multi-agent systems with heterogeneous unknown inertias and control gains under a directed graph," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2019–2034, 2016.

[20] R. G. Lins, S. N. Givigi, and P. R. G. Kurka, "Velocity estimation for autonomous vehicles based on image analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 1, pp. 96–103, 2015.

[21] W. Yu, W. X. Zheng, G. Chen, W. Ren, and J. Cao, "Second-order consensus in multi-agent dynamical systems with sampled position data," *Automatica*, vol. 47, no. 7, pp. 1496–1503, 2011.

[22] N. Huang, Z. Duan, and G. R. Chen, "Some necessary and sufficient conditions for consensus of second-order multi-agent systems with sampled position data," *Automatica*, vol. 63, pp. 148–155, 2016.

[23] A. Abdessameud and A. Tayebi, "On consensus algorithms for double-integrator dynamics without velocity measurements and with input constraints," *Systems & Control Letters*, vol. 59, no. 12, pp. 812–821, 2010.

[24] Y. Mao, E. Akyol, and Z. Zhang, "Strategic topology switching for security-Part II: Detection & switching topologies," *arXiv:1711.11181*, 2017.

[25] ——, "Novel defense strategy against zero-dynamics attack in multi-agent systems," in *Proceedings of the 58th IEEE Conference on Decision and Control*, pp. 3563–3568, 2019.

[26] ——, "Second-order consensus for multi-agent systems by time-dependent topology switching," in *Proceedings of the 57th IEEE Conference on Decision and Control*, pp. 6151–6156, 2018.

[27] ——, "Strategic topology switching for security-Part I: Consensus & switching times," *arXiv:1711.11183*.

[28] W. Xiang and J. Xiao, "Stabilization of switched continuous-time systems with all modes unstable via dwell time switching," *Automatica*, vol. 50, no. 3, pp. 940–945, 2014.

[29] M. Newman, *Networks: an introduction*. Oxford university press, 2010.

[30] D. G. Luenberger, "Observing the state of a linear system," *IEEE Transactions on Military Electronics*, vol. 8, no. 2, pp. 74–80, 1964.

[31] M. Porfiri, D. G. Roberson, and D. J. Stilwell, "Fast switching analysis of linear switched systems using exponential splitting," *SIAM Journal on Control and Optimization*, vol. 47, no. 5, pp. 2582–2597, 2008.

[32] A. E. Brouwer and W. H. Haemers, *Spectra of graphs*. Springer Science & Business Media, 2011.

[33] D. A. Harville, *Matrix algebra from a statistician's perspective*. Taylor & Francis Group, 1998.