Novel Stealthy Attack and Defense Strategies for Networked Control Systems

Yanbing Mao, Hamidreza Jafarnejadsani, Pan Zhao, Emrah Akyol, and Naira Hovakimyan

***Abstract*—This paper studies novel attack and defense strategies, based on a class of stealthy attacks, namely the zero-dynamics attack (ZDA), for multi-agent control systems. ZDA poses a formidable security challenge since its attack signal is hidden in the null-space of the state-space representation of the control system and hence it can evade conventional detection methods. An intuitive defense strategy builds on changing the aforementioned representation via switching through a set of carefully crafted topologies. In this paper, we propose realistic ZDA variations where the attacker is aware of this topology-switching strategy, and hence employs the following policies to avoid detection: (i) pause, update and resume ZDA according to the knowledge of switching topologies; (ii) cooperate with a concurrent stealthy topology attack that alters network topology at switching times, such that the original ZDA is feasible under the corrupted topology. We first systematically study the proposed ZDA variations, and then develop defense strategies against them under the realistic assumption that the defender has no knowledge of attack starting, pausing, and resuming times and the number of misbehaving agents. Particularly, we characterize conditions for detectability of the proposed ZDA variations, in terms of the network topologies to be maintained, the set of agents to be monitored, and the measurements of the monitored agents that should be extracted, while simultaneously preserving the privacy of the states of the non-monitored agents. We then propose an attack detection algorithm based on the Luenberger observer, using the characterized detectability conditions. We provide numerical simulation results to demonstrate our theoretical findings.**

***Index Terms*—Multi-agent systems, security, privacy, zero-dynamics attack, topology attack, attack detection.**

I. Introduction

COORDINATION and control of networked systems is a well-studied theoretical problem (see e.g., [2], [3]) with many practical applications including distributed optimization [4], power sharing for droop-controlled inverters in islanded microgrids [5], clock synchronization for sensor networks [6], as well as connected vehicles [7], spacecrafts [8], and electrical power networks [9].

Security concerns regarding the aforementioned networked systems pose a formidable threat to their wide deployment, as highlighted by the recent incidents including distributed denial-of-service (DDOS) attack on Estonian web sites [10] and Maroochy water breach [11]. The "networked" aspect exacerbates the difficulty of securing these systems, since centralized measurement (sensing) and control are not feasible for such large-scale systems [12], and hence require the development of decentralized approaches, which are inherently prone to attacks. Particularly, a special class of stealthy attacks, namely the "zero-dynamics attack" (ZDA), poses a significant security challenge [13]–[15]. The main idea behind ZDA is to hide the attack signal in the null-space of the state-space representation of the control system so that it cannot be detected by applying conventional detection methods on the observation signal. The objective of such an attack can vary from manipulating the controller to accept false data that would yield the system towards a desired (e.g., unstable) state to maliciously altering system dynamics (topology attack) to affect the system trajectory.

Recent research efforts have focused on variations of ZDA for systems with distinct properties. For stochastic cyber-physical systems, Park et al. [16] designed a robust ZDA, where the attack-detection signal is guaranteed to stay below a threshold over a finite horizon. In [17], Kim et al. proposed a discretized ZDA for the sampled-data control systems, where the attack-detection signal is constant zero at the sampling times. Another interesting line of research pertains to developing defense strategies [12], [18]–[21]. For example, Jafarnejadsani et al. [14] proposed a multi-rate $\mathcal{L}_1$ adaptive controller that can detect ZDA in sampled-data control systems by removing certain unstable zeros of discrete-time systems. Back et al. [22] used generalized hold strategy to mitigate the impact of ZDA.

Most of the prior work on defense strategies for the original ZDA in networked systems builds on rather restrictive assumptions regarding the connectivity of network topology and the number of the misbehaving agents (i.e., the agents under attack) [12], [18]–[20]. Teixeira et al. [23] showed that the strategic changes in system dynamics could be used by defender to detect ZDA. But the defense strategy requires the attack-starting times to be the initial time and known to defender, and the attacker has no capability of inferring the changed system dynamics. In other words, the defense strategy fails to work if the stealthy attack strategy is based on the newly inferred system dynamics. As a first step towards a practical ZDA defense strategy, in [24], strategic topology switching is proposed. This strategy is motivated by the feasibility of controlling communication topology driven due to recent developments in mobile computing, wireless com-

Y. Mao, P. Zhao and N. Hovakimyan are with the Department of Mechanical Science and Engineering, University of Illinois at Urbana–Champaign, Urbana, IL, 61801 USA (e-mail: {ybmao, panzhao2, nhovakim}@illinois.edu).

H. Jafarnejadsani is with the Department of Mechanical Engineering, Stevens Institute of Technology, Hoboken, NJ, 07310 USA (e-mail: hjafarne@stevens.edu).

E. Akyol is with the Department of Electrical and Computer Engineering, Binghamton University–SUNY, Binghamton, NY, 13902 USA (e-mail: eakyol@binghamton.edu).

Parts of this paper were presented at the 58th IEEE Conference on Decision and Control, 2019 [1]. This work is supported in part by NSF (award numbers CMMI-1663460 and ECCS-1739732), and Binghamton University–SUNY, Center for Collective Dynamics of Complex Systems ORC grant.

munication and sensing [25], [26]. We note, in passing, that the idea of using the changes in the state-space dynamics to detect ZDA first appeared in [23], albeit a realistic mechanism (e.g., switching the system topology) to achieve that objective was only very recently studied in [24]. However, the defense strategy in [24] still relies on a naive attacker that does not take the topology switching strategy of the defender into account.

In this paper, we systematically address this practically important problem: what kind of ZDA strategies can an informed attacker design against a topology-switching system and what are the optimal defense strategies, beyond switching the topology, against such intelligent attacks? We note that we study these questions under realistic assumptions on the capabilities of the defender, i.e., we assume that the defender does not know the start, pause and resume times of the attack or the number of misbehaving agents. We also assume that the attacker is aware of the strategic changes in system dynamics. Moreover, we assume that the defender has to preserve the privacy of the outputs of the non-monitored agents, since it is assumed that the attacker has access to the sensor outputs. The following example from coordination control illustrates our motivation to impose this privacy constraint.

For the coordination control of multi-agent systems, e.g., the connected autonomous vehicles, the data of initial positions and velocities can be used by the adversary to estimate target location [27], and the individual initial positions include individual home-base locations. Once the attacker has access to the outputs of monitored agents and the system is observable, the attacker can use current available data to infer the global initial condition and global real-time system state. From a perspective of stealthy topology attack design (e.g., topology attack in smart grids [28] and software-defined networks [29]), the attacker needs (estimated) real-time data of some agents' state to decide the target connection links to attack. Unfortunately, the inferred global real-time system state implies the largest scope of attackable connection links is exposed to the attacker. To reduce the feasible area of target links for ZDA in cooperation with a stealthy topology attack, monitored outputs have to be constrained to be unobservable to preserve the privacy of non-monitored agents' real-time states, consequently, the global system state and global initial condition.

Throughout this paper, we focus on the following policies which can be used by the attacker to evade detection:

1) intermittently pause attack if the incoming topology is unknown, and update (if necessary) and resume attack after the newly activated topology is inferred (intermittent ZDA).
2) cooperatively work with a stealthy topology attack, such that the original ZDA policy continues to be feasible under the corrupted topology (cooperative ZDA).

In this paper, we develop integrated defense strategies for both intermittent and cooperative ZDAs, in the presence of privacy considerations. More specifically, we develop defense strategies to address the following questions: what network topology should be maintained, which agents should be monitored and what measurements the monitored agents should output, such that the *intermittent* and *cooperative* ZDA variants are detectable, and at the same time, the privacy of non-monitored agents' real-time states are preserved? Based on the answers of the questions above, we next propose a strategic topology-switching algorithm to detect the ZDA.

The contributions are summarized as follows.

- To evade conventional detection methods that rely on naive attacker, we propose two ZDA variations: intermittent and cooperative ZDAs, where the attacker is aware of the defense strategy and has practical capability of inferring switching topologies.
- We systematically study the policies of ZDA variations that the attacker follows to devise stealthy attacks that lay the foundation for the novel defense strategies.
- We characterize conditions for detectability of the proposed ZDA variations, in terms of the network topologies to be maintained, the set of agents to be monitored, and the measurements of the monitored agents that should be extracted.
- Under the privacy-preserving constraint of non-monitored agents' states, we propose a strategic topology-switching algorithm for attack detection that is based on the detectability of ZDA variations using the Luenberger observer. The advantages of this approach include:
  - in achieving consensus and tracking real systems in the absence of attacks, it has no constraint on the magnitudes of coupling weights and observer gains;
  - in detecting ZDA variations, it allows the defender to be unaware of attack starting, pausing, and resuming times and the number of misbehaving agents;
  - in detecting ZDA variations, only one monitored agent is sufficient for intermittent ZDA and only two monitored agents are sufficient for cooperative ZDA.

This paper is organized as follows. We present the preliminaries and the problem formulation in Sections II and III, respectively. In Section IV, we analyze the proposed ZDA variations. In Section V, we characterize the conditions for detectability of these ZDA variations. Based on this characterization, we develop an attack detection algorithm in Section VI. Numerical simulation results are provided in Section VII, and the concluding remarks and the future research directions are discussed in Section VIII.

II. Preliminaries

A. Notation

We let $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ denote the set of n-dimensional real vectors and the set of $m \times n$-dimensional real matrices, respectively. Let $\mathbb{C}$ denote the set of complex numbers. $\mathbb{N}$ represents the set of the natural numbers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Let $\mathbf{1}_{n\times n}$ and $\mathbf{0}_{n\times n}$ be the $n \times n$-dimensional identity matrix and zero matrix, respectively. $\mathbf{1}_n \in \mathbb{R}^n$ and $\mathbf{0}_n \in \mathbb{R}^n$ denote the vector with all ones and the vector with 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., $\mu_P(A) = \frac{1}{2}\max_{i=1,\ldots,n}\left\{\lambda_i\left(P^{1/2}AP^{-1/2} + P^{-1/2}A^\top P^{1/2}\right)\right\}$. $\ker(Q) \triangleq \{y : Qy = \mathbf{0}_n, Q \in \mathbb{R}^{n\times n}\}$, $A^{-1}\mathbb{F} \triangleq \{y : Ay \in \mathbb{F}\}$. Also, $|\cdot|$

denotes the cardinality of a set, or the modulus of a number. $\mathbb{V}\backslash\mathbb{K}$ describes the complement set of $\mathbb{K}$ with respect to $\mathbb{V}$. $\lambda_i(M)$ is i^{th} eigenvalue of matrix M. $x^{(b)}(t)$ stands for the b^{th}-order time derivative of $x(t)$. For a matrix $W \in \mathbb{R}^{n\times n}$, W^k, $[W]_{i,j}$, $[W]_{i,:}$, and $[W]_{a:b,c:d}$ denote the k^{th} power of W, the element in row i and column j, the b^{th} row, and the sub-matrix formed by the entries in the a^{th} through b^{th} row and the c^{th} through d^{th} column of W, respectively.

The interaction among n agents is modeled by an undirected graph $\mathrm{G} \triangleq (\mathbb{V}, \mathbb{E})$, where $\mathbb{V} \triangleq \{1, 2, \ldots, n\}$ is the set of vertices that represents n agents and $\mathbb{E} \subseteq \mathbb{V} \times \mathbb{V}$ is the set of edges of the graph G. The weighted adjacency matrix $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{n\times n}$ of the graph G is defined as $a_{ij} = a_{ji} > 0$ if $(i, j) \in \mathbb{E}$, and $a_{ij} = a_{ji} = 0$ otherwise. Assume that there are no self-loops, i.e., for any $i \in \mathbb{V}$, $a_{ii} = 0$. The Laplacian matrix of graph G is defined as $\mathcal{L} \triangleq [l_{ij}] \in \mathbb{R}^{n\times n}$, where $l_{ii} \triangleq \sum_{j=1}^{n} a_{ij}$, and $l_{ij} \triangleq -a_{ij}$ for $i \neq j$. The diameter m of a graph is the longest shortest unweighted path between any two vertices in the graph.

B. Definitions

A second-order system consists of a population of n agents whose dynamics are governed by the following equations:

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

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 local control input. The broad applications of its coordination control is the main motivation of this paper considering the model (1), see e.g., [30]–[33]. For coordination control, we consider the more representative average consensus.

We recall the definitions of consensus and ZDA to review the control objective and the attack policy.

Definition 1: [34] The agents in the system (1) are said to achieve the asymptotic consensus with final zero common velocity if for any initial condition:

$$\lim_{t\to\infty} |x_i(t) - x_j(t)| = 0 \text{ and } \lim_{t\to\infty} |v_i(t)| = 0, \forall i, j \in \mathbb{V}. \tag{2}$$

Definition 2: [12], [35] Consider the system (with proper dimension) in the presence of attack signal $\breve{g}(t)$:

$$\dot{\breve{z}}(t) = A\breve{z}(t) + B\breve{g}(t), \tag{3a}$$
$$\breve{y}(t) = C\breve{z}(t) + D\breve{g}(t). \tag{3b}$$

The attack signal $\breve{g}(t) = ge^{\eta t}$ is a *zero-dynamics attack* if there exist a scalar $\eta \in \mathbb{C}$, and nonzero vectors $\mathbf{z}_0$ and g, that satisfy

$$\left[\begin{array}{c} \mathbf{z}_0 \\ \hdashline -g \end{array}\right] \in \ker\left(\left[\begin{array}{c:c} \eta\mathbf{1}_{n\times n} - A & B \\ \hdashline -C & D \end{array}\right]\right). \tag{4}$$

Moreover, the states and observed outputs of system (7) satisfy

$$\breve{y}(t) = y(t), t \geq 0 \tag{5}$$
$$\breve{z}(t) = z(t) + \mathbf{z}_0 e^{\eta t}, \tag{6}$$

where $y(t)$ and $z(t)$ are the output and state of the system (3) in the absence of attacks, i.e., the dynamics:

$$\dot{z}(t) = Az(t), \tag{7a}$$
$$y(t) = Cz(t). \tag{7b}$$

C. Control Protocol

We borrow a control protocol that involves topology switching from [34], [36] to achieve the consensus (2) for the agents in system (1):

$$u_i(t) = -v_i(t) + \sum_{j\in\mathbb{V}} a_{ij}^{\sigma(t)}(x_j(t) - x_i(t)), i \in \mathbb{V} \tag{8}$$

where $\sigma(t) : [t_0, \infty) \to \mathbb{S} \triangleq \{1, \ldots, \mathrm{s}\}$, is the switching signal of the interaction topology of the communication network; $a_{ij}^{\sigma(t)}$ is the entry of the weighted adjacency matrix that describes the activated topology of communication graph.

III. Problem Formulation

We let $\mathbb{K} \subseteq \mathbb{V}$ denote the set of misbehaving agents, i.e., the agents whose local control inputs are under attack. For simplicity, we let the increasingly ordered set $\mathbb{M} \triangleq \{1, 2, \ldots\} \subseteq \mathbb{V}$ denote the set of monitored agents for attack detection.

We make the following assumptions on the attacker and defender throughout this paper.

Assumption 1: The attacker

1) is aware that the changes in system dynamics are used by the defender (system operator);
2) knows the initial topology, output matrix and switching times;
3) needs a *non-negligible* time to exactly infer the newly activated topology, compute and update attack strategy;
4) records the newly inferred topology into memory;
5) knows the outputs of monitored agents in $\mathbb{M}$.

Assumption 2: The defender

1) designs the switching times and switching topologies;
2) chooses candidate agents to monitor, i.e., the monitored agent set $\mathbb{M}$, for attack detection;
3) has no knowledge of the attack starting, pausing and resuming times, and the misbehaving agents.

Remark 1: In Assumption 1, the assumed attacker's capability 2) is motivated by recent incidents, see e.g., the revenge sewage attack (cyber attack) that led to the Maroochy water breach, where the attacker had previously installed the industrial control systems for the water service network (consequently, he knew the control protocol and the locations of sensors) [37].

Remark 2: Strategically changing the system dynamics has been demonstrated to be an effective approach to detect system-based stealthy attacks, see e.g., ZDA [23], [24] and $\mathcal{C}_k/\mathcal{C}$ stealthy attacks [38]. The core idea behind this defense strategy is the intentional generation of mismatch between the models of the attacker and the defender. Specifically, the attacker uses the original system dynamics to make the stealthy attack decision (i.e., the computation (4)) before the system starts to operate, while the defender strategically changes the system dynamics at some operating point in time. However,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2997363, IEEE Transactions on Automatic Control

from the attacker's perspective, it is practical to become aware of this defense strategy, and hence try to infer the changed system dynamics to update the stealthy attack strategy and evade detection. This motivates the awareness capability 1) in Assumption 1.

Remark 3: Although the switching topologies are kept confidential from attackers, the developed topology inference algorithms [39], [40] enable the attacker to exactly infer the switching topologies from observation signals. Even with the global ability of observing all agents' states, the inference algorithms need to collect the state data over a time interval to obtain an exact topology solution, which explains the imposed *non-negligible* time in capability 3) in Assumption 1.

Remark 4: Since the sensor devices are embedded within an environment, they are frequently vulnerable to local eavesdropping, which is the motivation of capability 5) in Assumption 1. The ZDA policy (4) shows that the attacker does not need the capability 5) to obtain a feasible attack strategy consisting of the false data $\mathbf{z}_0$, and the parameters g and η of attack signal $\breve{g}(t)$. However, when ZDA seeks cooperation with a stealthy topology attack in response to strategic topology switching defense, then the attacker needs the real-time outputs indicated by the capability to identify the target links to attack.

Remark 5: As analyzed in [1], the defense strategy of strategically changing system dynamics [23] implicitly assumes that the attack-starting time must be the initial time and known to the defender. The capability 3) of defender in Assumption 2 removes this unrealistic assumption.

A. Topology Switching Strategy

The building block of our defense strategy is periodic topology switching, i.e., there exists a period τ such that

$$\sigma(t) = \sigma(t+\tau) \in \mathbb{S}. \tag{9}$$

- We note that (9) implies the building block belongs to the time-dependent topology switching. The critical reason that we do not consider state-dependent switching is the attack signals injected into control input may generate a Zeno behavior [41] that renders the control protocol (8) infeasible.
- If the topology switching is random, the defender needs to often send the generated "random" information of network topology to the detector/estimator/observer in the cyber layer as well, which will be subject to a cyber topology attack (incorrect information of network topology is transmitted) [28], [29], [42]. To avoid this type of cyber attack, the defender chooses here periodic topology switching, and preprogram the (repeated) periodic switching sequence into the controlled links, and hence avoids sending the topology information to the cyber layer during the system operation.

For our defense strategy based on the periodic topology switching (9), we define the following periodic sequence with length of l:

$$\mathbf{L} \triangleq \left\{ \underbrace{\sigma(t_0)}_{\tau_0}, \underbrace{\sigma(t_1)}_{\tau_1}, \ldots, \underbrace{\sigma(t_{l-2})}_{\tau_{l-2}}, \underbrace{\sigma(t_{l-1})}_{\tau_{l-1}} \right\}, \tag{10}$$

where τ_k denotes the dwell time of the activated topology indexed by $\sigma(t_k)$, i.e., $\tau_k = t_{k+1} - t_k$.

Next, we study whether the agents in the system (1) using control input (8) can reach consensus under periodic topology switching. We first recall the well-known property of Laplacian matrix $\mathcal{L}_r$ of a connected undirected graph from [43]:

$$Q_r^\top = Q_r^{-1}, \tag{11a}$$

$$[Q_r]_{1,1} = [Q_r]_{2,1} = \ldots = [Q_r]_{|\mathbb{V}|,1}, \tag{11b}$$

$$Q_r^\top \mathcal{L}_r Q_r = \text{diag}\{0, \lambda_2(\mathcal{L}_r), \ldots, \lambda_n(\mathcal{L}_r)\} \triangleq \Lambda_r, \tag{11c}$$

based on which, we define:

$$\Upsilon_{rs} \triangleq Q_r^\top \mathcal{L}_s Q_r, \tag{12a}$$

$$\mathcal{A}_s \triangleq \left[\begin{array}{c:c} \mathbf{0}_{(|\mathbb{V}|-1)\times(|\mathbb{V}|-1)} & \mathbf{1}_{(|\mathbb{V}|-1)\times(|\mathbb{V}|-1)} \\ \hdashline -[\Upsilon_{rs}]_{2:|\mathbb{V}|,2:|\mathbb{V}|} & -\mathbf{1}_{(|\mathbb{V}|-1)\times(|\mathbb{V}|-1)} \end{array}\right]. \tag{12b}$$

Proposition 1: Consider the second-order multi-agent system (1) with control input (8). If the sequence $\mathbf{L}$ in (10) includes one connected topology, there exists a periodic topology sequence that satisfies

$$\sum_{s=0}^{l-1} \nu_s \mu_P(\mathcal{A}_s) < 0, \tag{13}$$

where $\nu_s = \frac{\tau_s}{\tau}$ with $\tau = \sum_{i=0}^{l-1} \tau_i$. Moreover, under that periodic topology switching, the consensus (2) can be achieved.

Proof: See Appendix B. ∎

Remark 6: Proposition 1 implies that *periodic* topology switching has no constraint on the magnitudes of coupling weights in achieving consensus, i.e., for any coupling weights there exists a feasible periodic topology switching sequence for consensus. This is in sharp contrast with the *arbitrary* topology switching that imposes a strict condition on the magnitudes of coupling weights in achieving consensus [36].

B. System Description

Under periodic topology switching, the multi-agent system in (1), with the control input given by (8) and the outputs of monitored agents in $\mathbb{M}$ subject to the attack signal $g_i(t)$, can be written as

$$\dot{\breve{x}}_i(t) = \breve{v}_i(t) \tag{14a}$$

$$\dot{\breve{v}}_i(t) = -\breve{v}_i(t) + \sum_{j\in\mathbb{V}} a_{ij}^{\sigma(t)}(\breve{x}_j(t) - \breve{x}_i(t)) + \begin{cases} g_i(t), i \in \mathbb{K} \\ 0, \ i \in \mathbb{V}\backslash\mathbb{K} \end{cases} \tag{14b}$$

$$\breve{y}_i(t) = c_{i1}\breve{x}_i(t) + c_{i2}\breve{v}_i(t) + d_i g_i(t), i \in \mathbb{M} \tag{14c}$$

where c_{i1} and c_{i2} are constant coefficients designed by the defender (system operator), while constant coefficient d_i is designed by the attacker.

Remark 7: The model in (14b) with (1b) implies that there are two practical approaches to attack the local control inputs: (i) the attacker directly injects the attack signal to the control architectures of misbehaving agents (target agents) in $\mathbb{K}$; (ii) possibly through breaking the encryption algorithm that protects the communication channels with misbehaving agents, the attacker injects attack signals to the data sent to controller.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2997363, IEEE Transactions on Automatic Control

The system in (14) can be equivalently expressed as a switched system under attack:

$$\dot{\breve{z}}(t) = A_{\sigma(t)}\breve{z}(t) + \breve{y}(t) \tag{15a}$$
$$\breve{y}(t) = C\breve{z}(t) + D\breve{y}(t), \tag{15b}$$

where we define:

$$\breve{z}(t) \triangleq \left[\breve{x}_1(t) \,\vdots\, \dots \,\vdots\, \breve{x}_{|\mathbb{V}|}(t) \,\vdots\, \breve{v}_1(t) \,\vdots\, \dots \,\vdots\, \breve{v}_{|\mathbb{V}|}(t)\right]^\top, \tag{16a}$$
$$A_{\sigma(t)} \triangleq \begin{bmatrix} \mathbf{0}_{|\mathbb{V}|\times|\mathbb{V}|} & \mathbf{1}_{|\mathbb{V}|\times|\mathbb{V}|} \\ -\mathcal{L}_{\sigma(t)} & -\mathbf{1}_{|\mathbb{V}|\times|\mathbb{V}|} \end{bmatrix}, \tag{16b}$$
$$C \triangleq \begin{bmatrix} C_1 \,\vdots\, C_2 \end{bmatrix}, \tag{16c}$$
$$C_j \triangleq \left[\mathrm{diag}\{c_{1j}, \dots, c_{|\mathbb{M}|j}\} \,\vdots\, \mathbf{0}_{|\mathbb{M}|\times(|\mathbb{V}|-|\mathbb{M}|)}\right], j = 1, 2 \tag{16d}$$
$$D \triangleq \left[\mathbf{0}_{|\mathbb{M}|\times|\mathbb{V}|} \,\vdots\, \mathrm{diag}\{d_1, \dots, d_{|\mathbb{M}|}\} \,\vdots\, \mathbf{0}_{|\mathbb{M}|\times(|\mathbb{V}|-|\mathbb{M}|)}\right], \tag{16e}$$
$$\breve{y}(t) \triangleq \begin{bmatrix} \mathbf{0}_{|\mathbb{V}|}^\top \,\vdots\, \bar{g}^\top(t) \end{bmatrix}^\top, \tag{16f}$$
$$\bar{g}_i(t) \triangleq \begin{cases} g_i(t), i \in \mathbb{K} \\ 0, \quad i \in \mathbb{V}\backslash\mathbb{K}. \end{cases} \tag{16g}$$

In addition, we consider the system (15) in the absence of attacks, which is given by

$$\dot{z}(t) = A_{\sigma(t)} z(t), \tag{17a}$$
$$y(t) = Cz(t). \tag{17b}$$

C. Privacy of Initial Condition and Global System State

To fully secure multi-agent systems, e.g., connected autonomous vehicles, the initial conditions should be kept confidential from an adversary since the initial data could be utilized to estimate the target locations [27]. Moreover, individual initial positions contain the information of home-base locations. The following two examples illustrate that the global initial condition as well as the global system state play an important role in stealthy attacks.

Example 1 (Attack Objective): The state solution under attack (6) implies that if $\eta = 0$, attacker's objective is to modify the steady-state value. If the attack objective is to modify the target location to a new location that the attacker desires, the attacker must know the original target location in the absence of attacks. Under undirected communication, it is straightforward to verify from the system (1) with its control input (8) that the average position $\bar{x}(t) \triangleq \frac{1}{|\mathbb{V}|}\sum_{i\in\mathbb{V}} x_i(t)$ proceeds with the average velocity $\bar{v}(t) \triangleq \frac{1}{|\mathbb{V}|}\sum_{i\in\mathbb{V}} v_i(t) = e^{-t}\bar{v}(t_0)$, which indicates that when the consensus is achieved, all of the individual agents synchronize to the target location:

$$x^* = \lim_{t\to\infty}\left(\bar{x}(t_0) + \left(1 - e^{-t}\right)\bar{v}(t_0)\right) = \bar{x}(t_0) + \bar{v}(t_0). \tag{18}$$

Unfortunately, (18) shows that once the global initial condition is known (i.e., initial positions and velocities of all agents), the original target location can simply be computed through a simple mean computation.

Example 2 (Stealthy Topology Attack Design): Stealthy topology attack design, as in smart grids [28] and power networks [42], requires (estimated) real-time data of system states to choose the target connection links to maliciously alter. Since attacker can record the newly obtained knowledge of the network topology, the attacker has the memory of the past topology sequence. Whenever the data on the global initial condition $z(t_0)$ (or real-time global state $z(t)$) is available, the attacker can infer the exact real-time global state $z(t)$ (or global initial condition $z(t_0)$) through

$$z(t) = e^{A_{\sigma(t_k)}(t-t_k)}\prod_{l=0}^{k-1} e^{A_{\sigma(t_l)}(t_{l+1}-t_l)} z(t_0), t \in [t_k, t_{k+1})$$

which indicates whenever ZDA seeks a cooperation with stealthy topology attack to evade detection, the attacker would have the largest scope of attackable links since the attacker knows all of agents' real-time state data. Therefore, the private global initial condition or system state can reduce the scope of target links for stealthy topology attack.

We next impose the following unobservability condition on the monitored outputs to preserve the privacy of non-monitored agents, such that the attacker cannot use the available (monitored) outputs to infer any non-monitored agent's full state (and consequently, the global system state and initial condition).

Lemma 1: For the system (17), $x_i(t)$ and $v_i(t)$, $\forall i \in \mathbb{V}\backslash\mathbb{M}$, are not simultaneously observable for any $t \in [t_0, t_m^+)$, if and only if

$$\exists p \in \mathbf{N}_0^m : |p_i| + \left|p_{i+|\mathbb{V}|}\right| \neq 0, \forall i \in \mathbb{V}\backslash\mathbb{M} \tag{19}$$

where

$$\mathbf{N}_m^m = \ker(\mathcal{O}_m), \tag{20}$$
$$\mathbf{N}_q^m = \ker(\mathcal{O}_q) \bigcap e^{-A_{\sigma(t_q)}\tau_q}\mathbf{N}_{q+1}^m, 0 \leq q \leq m-1 \tag{21}$$
$$\mathcal{O}_q = \left[C^\top \,\vdots\, (CA_{\sigma(t_q)})^\top \,\vdots\, \dots \,\vdots\, (CA_{\sigma(t_q)}^{2|\mathbb{V}|-1})^\top\right]^\top. \tag{22}$$

Proof: The condition in (19) implies that $\mathbf{N}_0^m \neq \{\mathbf{0}_{2|\mathbb{V}|}\}$. Using Theorem 1 in [44], it follows that the system in (17) is unobservable for any $t \in [t_0, t_m^+)$. Also, (19) implies that $p_i \neq 0$, and (or) $p_{i+|\mathbb{V}|} \neq 0$, and therefore the agent i's position and (or) velocity are (is) not partially observable. ■

Remark 8: Although the selection of the monitored output coefficients in (14c) subject to (19) renders the system (17) unobservable to preserve privacy, we will show that the proposed ZDA variations become detectable using the outputs $y_i(t)$'s by careful selection of switching topologies and the set of monitored agents.

IV. Stealthy Attack Model

In the scenario where the attacker is aware of the detection purpose of strategic changes in system dynamics induced by topology switching [24], the attacker can evolve the attack policies in response to the strategic changes at switching times to stay stealthy:

- "pause attack" before topology switching when the incoming topology is unknown or the attack policy (4) is infeasible under the known incoming topology, and "resume attack" after the feasibility of (updated if needed) attack policy under the inferred activated topology is verified;
- cooperate with a topology attack that maliciously alters network topology at switching times, such that the original attack policy (4) continues to be feasible under the corrupted topology.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2997363, IEEE Transactions on Automatic Control

In the following subsections, we present a systematic study on these ZDA variations.

A. Intermittent Zero-Dynamics Attack

For convenience, we refer to $\mathbb{T}$ as the set of topologies under which the attacker injects attack signals to control inputs, and we refer to ξ_k and ζ_k as the attack-resuming and attack-pausing times over the active topology intervals $[t_k, t_{k+1})$, $k \in \mathbb{N}_0$, respectively.

The ZDA signals injected into the control input and monitored output of system (14) with intermittent pausing and resuming behaviors are described as

$$g_i(t) = \begin{cases} g_i^{\sigma(t_k)} e^{\eta_{\sigma(t_k)}(t-\xi_k)}, & t \in [\xi_k, \zeta_k) \subseteq [t_k, t_{k+1}) \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

To analyze this ZDA, we review the monitored output (14c) at the first "pausing" time ζ_0:

$$\breve{y}_i\left(\zeta_0^-\right) = c_{i1}\breve{x}_i(\zeta_0^-) + c_{i2}\breve{v}_i(\zeta_0^-) + d_i g_i\left(\zeta_0^-\right), \forall i \in \mathbb{M}$$

which implies that $\breve{y}_i\left(\zeta_0^-\right) = \breve{y}_i\left(\zeta_0\right)$ if and only if $g_i\left(\zeta_0^-\right) = g_i\left(\zeta_0\right)$, since $\breve{v}_i\left(\zeta_0^-\right) = \breve{v}_i\left(\zeta_0\right)$ and $\breve{x}_i\left(\zeta_0^-\right) = \breve{x}_i\left(\zeta_0\right)$. Meanwhile, the velocity and position states are always continuous with respect to time, and hence the monitored outputs must be continuous as well. Therefore, to avoid the "jump" on monitored outputs to maintain the stealthy property (5), the attacker cannot completely pause the attack, i.e., whenever the attacker pauses injecting ZDA signals to control inputs at pausing time ζ_k, simultaneously continues to inject the same attack signals to monitored outputs (14c):

$$\breve{y}_i(t) = c_{i1}\breve{x}_i(t) + c_{i2}\breve{v}_i(t) + d_i \sum_{m=0}^{k} g_i\left(\zeta_m^-\right), t \in [\zeta_k, \xi_{k+1}) \tag{24}$$

or equivalently,

$$\breve{y}(t) = C\breve{z}(t) + D\sum_{m=0}^{k} \breve{g}\left(\zeta_m^-\right), t \in [\zeta_k, \xi_{k+1}). \tag{25}$$

Based on the above analysis, for ZDA policy consisting of "pause attack" and "resume attack" behaviors to remain stealthy, it should satisfy (25) and

$$\mathbf{z}(t_0) \in \widehat{\mathbf{N}}_0^k \bigcap \widetilde{\mathbf{N}}_0^k, \tag{26a}$$

$$\begin{bmatrix} \mathbf{z}(\xi_k) \\ \hdashline -\breve{g}(\xi_k) \end{bmatrix} \in \ker(\mathcal{P}_r), \forall \sigma(\xi_k) \in \mathbb{T} \tag{26b}$$

where

$$\widehat{\mathbf{N}}_k^k = \ker(\mathcal{O}_k), \tag{27}$$

$$\widehat{\mathbf{N}}_q^k = \ker(\mathcal{O}_q) \bigcap e^{-A_{\sigma(t_q)}(\tau_q - (\zeta_q - \xi_q))} \mathbf{N}_{q+1}^k, 0 \le q \le k-1 \tag{28}$$

$$\widetilde{\mathbf{N}}_k^k = \ker(\widetilde{\mathcal{O}}_k), \tag{29}$$

$$\widetilde{\mathbf{N}}_q^k = \ker(\widetilde{\mathcal{O}}_q) \bigcap e^{-A_{\sigma(t_q)}(\tau_q - (\zeta_q - \xi_q))} \mathbf{N}_{q+1}^k, 0 \le q \le k-1 \tag{30}$$

$$\widetilde{\mathcal{O}}_r \triangleq \left[(CA_r)^\top \;\vdots\; (CA_r^2)^\top \;\vdots\; \dots \;\vdots\; (CA_r^{2|\mathbb{V}|})^\top \right]^\top, \tag{31}$$

$$\mathcal{P}_r \triangleq \begin{bmatrix} \eta_r \mathbf{1}_{2|\mathbb{V}|\times 2|\mathbb{V}|} - A_r & \vdots & \mathbf{1}_{2|\mathbb{V}|\times 2|\mathbb{V}|} \\ \hdashline -C & \vdots & D \end{bmatrix}, \tag{32}$$

$$\mathbf{z} = \left[\mathbf{x}^\top \;\vdots\; \mathbf{v}^\top \right]^\top \triangleq \breve{z} - z = \left[\breve{x}^\top - x^\top \;\vdots\; \breve{v}^\top - v^\top \right]^\top, \tag{33}$$

and $\mathcal{O}_r$ is given by (22).

Proposition 2: Under the stealthy attack policy consisting of (25) and (26), the states and monitored outputs of the systems (17) and (15) in the presence of attack signal (23) satisfy

$$\breve{y}(t) = y(t), t \in [t_0, t_{k+1}), \tag{34}$$

$$\breve{z}(t) = z(t) + e^{\eta_{\sigma(t_k)}(t-\xi_k)} \mathbf{z}(\xi_k), t \in [\xi_k, \zeta_k). \tag{35}$$

Proof: See Appendix C. ∎

Remark 9: At first glance, it might seem that the intermittent ZDA is an asynchronous attack response to the strategic topology switching, which is due to the imposed *non-negligible* time on capability 3) in Assumption 1. We note however that the attacker can record the newly obtained topology knowledge into the memory. Since the defender switches topologies *periodically*, if the recorded length of topology sequence is sufficiently long, the attacker can learn from the recorded memory the (recurring) periodic sequence, i.e., the attacker knows all future switching topologies and times. The corresponding future synchronous attack policies can be obtained off-line. Therefore, a synchronous attack response is possible only after the attacker obtains the (recurring) periodic topology sequence from memory.

B. Cooperative Zero-Dynamics Attack

The objective of cooperation with stealthy topology attack is to make the ZDA policy (4) continue to hold under the corrupted topology. Stealthy topology attack can be of two types:

- Physical Topology Attack: the attacker maliciously alters the status of target connection links of physical systems, e.g., the bus interaction breaks in power networks [42] and link fabrication in software-defined networks [29].
- Cyber Topology Attack: the attacker maliciously alters the information of network topology sent to the estimator/observer/detector in cyber layer [28], [45].

As stated in Subsection III-A, the basis of our defense strategy is the periodic topology switching, and the defender (system operator) would preprogram the repeated switching times and topologies into the controlled links of the real system and observer/detector. In this case, the operator of the real system does not need to send the topology information to the observer/detector when the system operates. Therefore, the system under our defense strategy is not subject to a cyber topology attack, albeit it is subject to a physical topology attack.

We let t_{k+1} denote the switching time when ZDA cooperates with topology attack. The multi-agent system (15) in the presence of such cooperative attacks is described by

$$\dot{\breve{z}}(t) = \widehat{A}_{\sigma(t)}\breve{z}(t) + \breve{g}(t), t \in [t_{k+1}, t_{k+2}) \tag{36a}$$

$$\breve{y}(t) = C\breve{z}(t) + D\breve{g}(t), \tag{36b}$$

where $\widehat{A}_{\sigma(t)}$ is defined as

$$\widehat{A}_{\sigma(t)} \triangleq \begin{bmatrix} \mathbf{0}_{|\mathbb{V}|\times|\mathbb{V}|} & \vdots & \mathbf{1}_{|\mathbb{V}|\times|\mathbb{V}|} \\ \hdashline -\widehat{\mathcal{L}}_{\sigma(t)} & \vdots & -\mathbf{1}_{|\mathbb{V}|\times|\mathbb{V}|} \end{bmatrix}, \tag{37}$$

with $\widehat{\mathcal{L}}_{\sigma(t_{k+1})}$ denoting the Laplacian matrix of the corrupted topology. We describe its corresponding system in the absence

of ZDA, i.e., in the presence of the only physical topology attack, as

$$\dot{\widehat{z}}(t) = \widehat{A}_{\sigma(t)}\widehat{z}(t), t \in [t_{k+1}, t_{k+2}) \tag{38a}$$

$$\widehat{y}(t) = C\widehat{z}(t). \tag{38b}$$

If $\breve{y}(t)$ is a ZDA signal in systems (15) and (36) at times t_{k+1}^- and t_{k+1}, by (6) we have $\breve{z}(t_{k+1}) = \breve{z}(t_{k+1}^-) = z(t_{k+1}^-) + \mathbf{z}_0 e^{\eta t_{k+1}}$ and $\breve{z}(t_{k+1}^-) = \breve{z}(t_{k+1}) = \widehat{z}(t_{k+1}) + \mathbf{z}_0 e^{\eta t_{k+1}}$. Here, we conclude that

$$\widehat{z}(t_{k+1}) = z(t_{k+1}^-) = z(t_{k+1}), \tag{39}$$

otherwise, the system state $\breve{z}(t_{k+1})$ has "jump" behavior, which contradicts with the fact that $\breve{z}(\cdot)$ is continuous.

The equation (39) and the stealthy property (5) imply that $C\breve{z}(t_{k+1}) = Cz(t_{k+1}) = C\widehat{z}(t_{k+1})$, based on which, a necessary condition for the existence of ZDA under corrupted topology is stated formally in the following proposition.

Proposition 3: Consider the systems in (38) and (17). We have $y(t) = \widehat{y}(t)$ for any $t \in [t_{k+1}, t_{k+2})$, if and only if

$$\sum_{l=0}^{d} C\widehat{A}_{\sigma(t_{k+1})}^{l}(\widehat{A}_{\sigma(t_{k+1})} - A_{\sigma(t_{k+1})})z^{(d-l)}(t_{k+1}) = \mathbf{0}_{|\mathbb{M}|},\ \forall d \in \mathbb{N}_0. \tag{40}$$

Proof: See Appendix D. ∎

We set $d = 0, 1$ and expand (40) out to obtain:

$$C_2(\widehat{\mathcal{L}}_{\sigma(t_{k+1})} - \mathcal{L}_{\sigma(t_{k+1})})x(t_{k+1}) = \mathbf{0}_{|\mathbb{M}|}, \tag{41a}$$

$$C_2(\widehat{\mathcal{L}}_{\sigma(t_{k+1})} - \mathcal{L}_{\sigma(t_{k+1})})v(t_{k+1}) = \mathbf{0}_{|\mathbb{M}|}. \tag{41b}$$

The result (41) shows that like the stealthy topology attacks in smart grids [28], [45] and software-defined networks [29], the attacker needs some agents' real-time state data to decide the target links to attack, while according to Lemma 1, the attacker cannot simultaneously infer $x_i(t_{k+1})$ and $v_i(t_{k+1})$, $\forall i \in \mathbb{V}\backslash\mathbb{M}$. Therefore, there should be a scope of attackable connection links under the strategy (19).

Without loss of generality, we express the difference of Laplacian matrices in the form:

$$\widehat{\mathcal{L}}_{\sigma(t_{k+1})} - \mathcal{L}_{\sigma(t_{k+1})} = \begin{bmatrix} \mathfrak{L}_{\sigma(t_{k+1})} & \mathbf{0}_{|\mathbb{D}|\times(|\mathbb{V}|-|\mathbb{D}|)} \\ \mathbf{0}_{(|\mathbb{V}|-|\mathbb{D}|)\times|\mathbb{D}|} & \mathbf{0}_{(|\mathbb{V}|-|\mathbb{D}|)\times(|\mathbb{V}|-|\mathbb{D}|)} \end{bmatrix}, \tag{42}$$

where $\mathbb{D}$ denotes the set of agents in the sub-graph formed by the target links to be possibly attacked, $\mathfrak{L}_{\sigma(t_{k+1})} \in \mathbb{R}^{|\mathbb{D}|\times|\mathbb{D}|}$ is the elementary row transformation of the Laplacian matrix of a subgraph $\mathcal{G}$ in the difference graph, which is generated by the corrupted graph $\widehat{\mathcal{G}}_{t_{k+1}}$ of the topology attacker and candidate graph $\mathcal{G}_{t_{k+1}}$ of the defender at time t_{k+1}.

Since $C_2 \in \mathbb{R}^{|\mathbb{M}|\times|\mathbb{V}|}$ and $\mathfrak{L}_{\sigma(t_{k+1})} \in \mathbb{R}^{|\mathbb{D}|\times|\mathbb{D}|}$, the relations in (19), (41), and (42) imply that the attacker can devise a stealthy topology attack (without knowing the measurements of the agents in $\mathbb{V}\backslash\mathbb{M}$ which are unavailable) only when the scope of target links satisfies:

$$\mathbb{D} \subseteq \mathbb{M}. \tag{43}$$

V. Detectability of Stealthy Attacks

Based on the systematic study of the attack behaviors and policies in Section IV, in this section, we investigate the detectability of the proposed ZDA variations.

A. Detectability of Intermittent Zero-Dynamics Attack

We first define

$$\mathcal{U}_{ri} \triangleq \text{diag}\left\{[Q_r]_{i,1}, \ldots, [Q_r]_{i,|\mathbb{V}|}\right\} Q_r^\top, \tag{44}$$

$$\mathbb{F} \triangleq \left\{ i | [Q_r]_{i,j} \neq 0, i \in \mathbb{M}, \forall j \in \mathbb{V}, \forall r \in \mathbf{L} \right\}, \tag{45}$$

where Q_r satisfies (11).

Defense Strategy Against Intermittent ZDA

Strategy on switching topologies: $\mathcal{L}_r$ has distinct eigenvalues for $\forall r \in \mathbf{L}$. (46)

Strategy on monitored-agent locations: $\mathbb{F} \neq \emptyset$. (47)

Theorem 1: Consider the system (14) in the presence of attack signals (23). Under the defense strategy against intermittent ZDA,

- if the monitored agents output the full observations of their velocities (i.e., $c_{i1} = 0$ and $c_{i2} \neq 0$ for $\forall i \in \mathbb{M}$), the intermittent ZDA is detectable and
$$\mathbf{N}_0^\infty = \left\{ \mathbf{0}_{2|\mathbb{V}|}, \left[\mathbf{1}_{|\mathbb{V}|}^\top \vdots \mathbf{0}_{|\mathbb{V}|}^\top \right]^\top \right\}; \tag{48}$$
- if the monitored agents output the full observations of their positions (i.e., $c_{i1} \neq 0$ and $c_{i2} = 0$ for $\forall i \in \mathbb{M}$), the intermittent ZDA is detectable but
$$\mathbf{N}_0^\infty = \left\{ \mathbf{0}_{2|\mathbb{V}|} \right\}; \tag{49}$$
- if the monitored agents output the partial observations (i.e., $c_{i1} \neq 0$ and $c_{i2} \neq 0$ for $\forall i \in \mathbb{M}$), and $c_{i1} = c_{i2}, \forall i \in \mathbb{M}$, the kernel of the observability matrix satisfies
$$\mathbf{N}_0^\infty = \left\{ \mathbf{0}_{2|\mathbb{V}|}, \left[\mathbf{1}_{|\mathbb{V}|}^\top \vdots -\mathbf{1}_{|\mathbb{V}|}^\top \right]^\top \right\}; \tag{50}$$
and the intermittent ZDA is detectable if
$$\xi_0 > t_0, \text{ or } D = \mathbf{0}_{|\mathbb{M}|\times 2|\mathbb{V}|}, \tag{51}$$
where $\mathbf{N}_0^\infty$ is computed recursively by (20) and (21).

Proof: See Appendix E. ∎

Under the defense strategy consisting of (46) and (47), the result (49) implies that if the monitored agents output full observations of position, the condition (19) is not satisfied. While the results (48) and (50) show that if the monitored agents output full observations of velocity or partial observations, the condition (19) is satisfied, and according to Lemma 1, the privacy of all states of non-monitored agents is preserved, which further implies that using the available data (5), the attacker cannot infer the global system state and the global initial condition. Therefore, for the purpose of privacy preserving of non-monitored agents' states, consequently, restricting the scope of attackable links to derive the defense strategies against the cooperative ZDA, we abandon full observation of position.

B. Detectability of Cooperative Zero-Dynamics Attack

Considering the matrix Q_r satisfying (11), we describe the defense strategy as follows:

Defense Strategy Against Cooperative ZDA

Strategy on switching topologies: (46).

Strategy on monitored-agent outputs: $c_{i2} > 0, \forall i \in \mathbb{M}$. (52)

Strategy on monitored-agent locations:

$[Q_r]_{i,m} - [Q_r]_{j,m} \neq 0, \forall m \in \mathbb{V}\backslash\{1\}, \forall r \in \mathbf{L}, \forall i \neq j \in \mathbb{M}$. (53)

Theorem 2: Consider the system (36) in the presence of zero-dynamics attack in cooperation with topology attack under (43). Under the defense strategy against cooperative ZDA, the attack is detectable.

Proof: See Appendix F. ∎

Remark 10: The common critical requirement of our defense strategies is that the communication network has distinct Laplacian eigenvalues. There indeed exist many topologies whose associated Laplacian matrices have distinct eigenvalues. The following lemma provides a guide to design such topologies:

Lemma 2 (Proposition 1.3.3 in [43]): Let G be a connected graph with diameter m. Then, G has at least $m+1$ distinct Laplace eigenvalues.

VI. Attack Detection Algorithm

Using the proposed defense strategies and the detectability conditions in Section V, this section focuses on the attack detection algorithm that is based on a Luenberger observer.

A. Luenberger Observer under Switching Topology

We now present a Luenberger observer [46]:

$$q_i(t) = w_i(t) \tag{54a}$$

$$\dot{w}_i(t) = -w_i(t) + \sum_{i\in\mathbb{V}} a_{ij}^{\sigma(t)}(q_j(t) - q_i(t)) - \begin{cases} r_i(t), & c_{i1}\neq 0, i\in\mathbb{M} \\ \int_{t_0}^{t} r_i(b)\mathrm{d}b, & c_{i1}=0, i\in\mathbb{M} \\ 0, & i\in\mathbb{V}\backslash\mathbb{M} \end{cases} \tag{54b}$$

$$r_i(t) = c_{i1}q_i(t) + c_{i2}w_i(t) - \breve{y}_i(t), i\in\mathbb{M} \tag{54c}$$

where $\breve{y}_i(t)$ is the monitored output of agent i in system (14), $r_i(t)$ is the attack-detection signal.

We next consider a system matrix related to the system (54) in the absence of attacks:

$$\widehat{\mathcal{A}}_r \triangleq \left[\begin{array}{c|c} \mathbf{0}_{|\mathbb{V}|\times|\mathbb{V}|} & \mathbf{1}_{|\mathbb{V}|\times|\mathbb{V}|} \\ \hline -\mathcal{L}_r - \widehat{C} & -\mathbf{1}_{|\mathbb{V}|\times|\mathbb{V}|} \end{array}\right], \tag{55}$$

where

$$\widehat{C} \triangleq \left[\begin{array}{c} C_1 \\ \hline \mathbf{0}_{(|\mathbb{V}|-|\mathbb{M}|)\times|\mathbb{V}|} \end{array}\right] \text{ or } \left[\begin{array}{c} C_2 \\ \hline \mathbf{0}_{(|\mathbb{V}|-|\mathbb{M}|)\times|\mathbb{V}|} \end{array}\right] \tag{56}$$

with C_1 and C_2 given by (16d). It is straightforward to obtain the following result regarding the matrix stability.

Lemma 3: The matrix $\widehat{\mathcal{A}}_r$ defined by (55) is Hurwitz, if $\mathcal{L}_r$ is the Laplacian matrix of a connected graph and

$$\mathbf{0}_{|\mathbb{V}|\times|\mathbb{V}|} \neq \widehat{C} \geq 0. \tag{57}$$

If the sequence (10) has one connected graph and gain matrix $\widehat{C}$ (56) satisfies (57), it follows from Lemma 3 that there exists a $P>0$, such that under convex linear combination, the matrix measure satisfies

$$\sum_{s=0}^{l-1} \nu_s \mu_P\left(\widehat{\mathcal{A}}_s\right) < 0. \tag{58}$$

Algorithm 1: Strategic Topology Switching

Input: Initial index k = 0, initial time $t_k = 0$, observer gains satisfying (57), periodic sequence $\mathbf{L}$ (10) with length of l satisfying (13) and (58).

1. Run the system (14) and the observer (54);
2. Update dwell time: $\tau_{\sigma(t_k)} \leftarrow \tau_{\sigma(t_{\bmod(k,L+1)})}$;
3. Switch topology of system (14) and observer (54) at time $t_k + \tau_{\sigma(t_k)}$: $\sigma(t_k+\tau_{\sigma(t_k)}) \leftarrow \mathbf{L}(\mathrm{mod}(k+1, L))$;
4. Update switching time: $t_k \leftarrow t_k + \tau_{\sigma(t_k)}$;
5. Update index: $k \leftarrow k+1$;
6. Go to Step 2.

B. Strategic Topology-Switching Algorithm

We next propose Algorithm 1 that describes when and which topology to switch to detect the ZDA variations.

Theorem 3: If the monitored agents satisfy (47), (52) and (53), and the switching topologies in $\mathbf{L}$ satisfy (46),

- without requiring the knowledge of the misbehaving agents and the start, pause, and resume times of the attack,
  1) with $c_{i1}=0$, $\forall i\in\mathbb{M}$, the observer (54) is able to detect the intermittent and cooperative ZDAs;
  2) with $c_{i1}=c_{i2}$, $\forall i\in\mathbb{M}$, the observer (54) is able to detect the cooperative ZDA and intermittent ZDA under (51);
- in the absence of attacks, the agents in system (14) achieve the asymptotic consensus, and the observer (54) asymptotically tracks the real system (15) if $c_{i1}=c_{i2}$, $\forall i\in\mathbb{M}$, or $c_{i1}=0$, $\forall i\in\mathbb{M}$.

Proof: See Appendix G. ∎

Remark 11: The modulo operations in steps 2 and 3 of Algorithm 1 describe the building block of our defense strategy, that is periodic topology switching. Given the length of topology switching sequence, i.e., l, and the length of the running time of the system (14) and the observer (54), denoted by $t_f - t_0$, the total number of topology switchings can roughly be computed as $\frac{t_f-t_0}{\tau}l$.

VII. Simulations

We consider a system with $n=16$ agents. The initial position and velocity conditions are chosen as $x(t_0) = \left[2\times\mathbf{1}_8^\top, 4\times\mathbf{1}_8^\top\right]^\top$ and $v(t_0) = \left[6\times\mathbf{1}_8^\top, 8\times\mathbf{1}_8^\top\right]^\top$. The coupling weights and observer gains are uniformly set to one. The considered network topologies are given in the following Figures 1 and 4 where the yellow nodes denote the monitored agents that output full observations of individual velocities.

A. Detection of Intermittent ZDA

We first consider the periodic topology switching scheme in Figure 1 (a). We denote the topologies with the controlled links $a_{17}^{\sigma(t)}$ in "On" and "Off" by 1 and 2, respectively. The considered corresponding periodic switching sequence is $\mathbf{L} = \left\{\underbrace{\sigma(t_0)=1}_{\tau_0=3}, \underbrace{\sigma(t_1)=2}_{\tau_1=6}\right\}$. It can be verified that with

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2997363, IEEE Transactions on Automatic Control

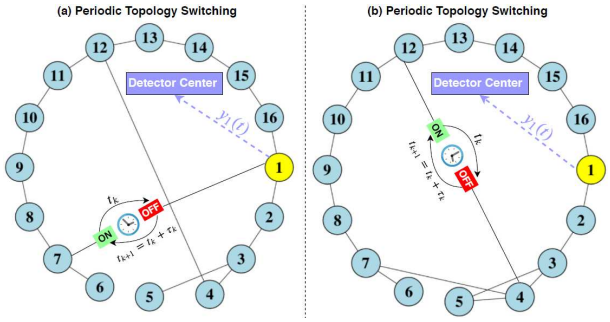

Figure 1. Two periodic topology switching schemes for intermittent ZDA.

$y_1(t) = v_1(t)$, neither of the switching topologies in Figure 1 (a) has distinct eigenvalues and $\mathbb{F} = \varnothing$, such that the defense strategy consisting of (46) and (47) does not hold. Therefore, the attacker can design an undetectable intermittent ZDA as follows:

- inject false data $\mathbf{z}(t_0) = \left[\mathbf{0}_3^\top, -1, 1, \mathbf{0}_{14}^\top, -0.08 - 2\mathrm{i}, 0.08 + 2\mathrm{i}, \mathbf{0}_{11}^\top\right]^\top$ to the data of initial condition sent to the observer (54);
- inject ZDA signals $\breve{g}_4(t) = (2.9136 + 2.32\mathrm{i})e^{(0.08-2\mathrm{i})(t-0.2)}$ and $\breve{g}_5(t) = (-2.9136 - 2.32\mathrm{i})e^{(0.08-2\mathrm{i})(t-0.2)}$ to the local control inputs of agents 4 and 5 for the initial Topology 1 at $\xi_0 = 0.2$;
- pause the ZDA if the incoming topology is unknown;
- update the attack strategy if necessary, and resume the feasible attack after newly switched topology is inferred;
- iterate the last two steps.

Some agents' velocities and the attack-detection signals in Figure 2 show that with $y_1(t) = v_1(t)$, when the defense strategy consisting of (46) and (47) does not hold, the attacker can design an intermittent ZDA that cannot be detected by the observer (54) under Algorithm 1 (constant zero detection signal), and the stealthy attack renders the system unstable (in the absence of attacks, $\lim_{t\to\infty} |v_i(t)| = 0, \forall i \in \mathbb{V}$).

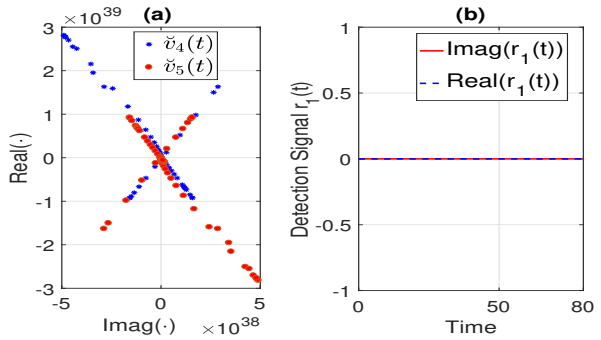

Figure 2. Individual velocities (a) and attack-detection signal (b).

The switching topologies in Figure 1 (b) satisfy the defense strategy consisting of (46) and (47). Hence, according to the first statement in Theorem 1, with $y_1(t) = v_1(t)$, i.e., $c_{11} = 0$, we can turn to the switching scheme at some time to detect the stealthy attack. Under the topology switching scheme in Figure 1 (b), the trajectory of attack-detection signal in Figure 3 (b) shows the observer (54) succeeds in detecting the intermittent ZDA (nonzero detection signal), which also demonstrates the first statement in Theorem 3.

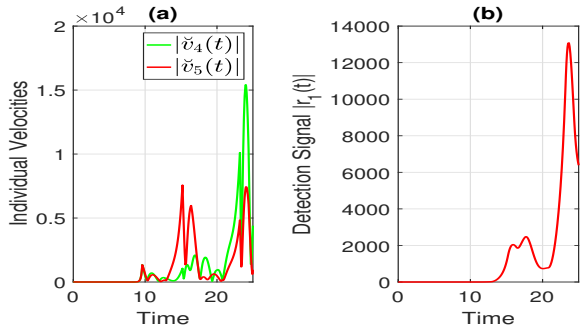

Figure 3. Trajectories of velocities (a) and attack-detection signal (b).

B. Detection of Cooperative ZDA

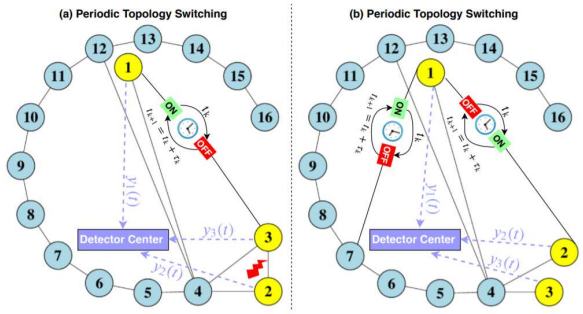

Figure 4. Network topologies for cooperative ZDA.

We denote the switching topologies in Figure 4 (a) by 3 and 4, in Figure 4 (b) by 5 and 6, respectively. It can be verified that with $y_i(t) = v_i(t)$, $i \in \mathbb{M} = \{1, 2, 3\}$, neither Topology 3 nor Topology 4 satisfies the defense strategy consisting of (46), (52) and (53). Therefore, under the periodic topology switching sequence $\mathbf{L} = \left\{\underbrace{\sigma(t_0) = 3}_{\tau_0=3}, \underbrace{\sigma(t_1) = 4}_{\tau_1=1}\right\}$, it is possible to design stealthy cooperative ZDA as follows:

- inject false data $\mathbf{z}(t_0) = [0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^\top$ to the data of initial condition sent to observer (54);
- inject ZDA signals $\breve{g}_1(t) = \breve{g}_3(t) = \breve{g}_7(t) = -e^t$, $\breve{g}_4(t) = 5e^t$, $\breve{g}_5(t) = 2e^t$ and $\breve{g}_6(t) = 3e^t$ to the local control inputs of agents 1, 3, 7, 4, 5 and 6, respectively, at initial time for Topology 3;
- inject false data $-e^t$ to the monitored outputs;
- maliciously control the connection between agents 2 and 3, such that the original ZDA policy maintains its feasibility under the corrupted topology at incoming switching times.

The trajectories of velocities and attack-detection signals in Figure 5 show that the designed attack makes system unstable without being detected (constant zero detection signals).

The switching topologies in Figure 4 (b) satisfy the defense strategy consisting of (46), (52) and (53). Therefore, with $y_i(t) = v_i(t)$, i.e., $c_{i1} = 0$, $i \in \mathbb{M} = \{1, 2, 3\}$, according to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2997363, IEEE Transactions on Automatic Control

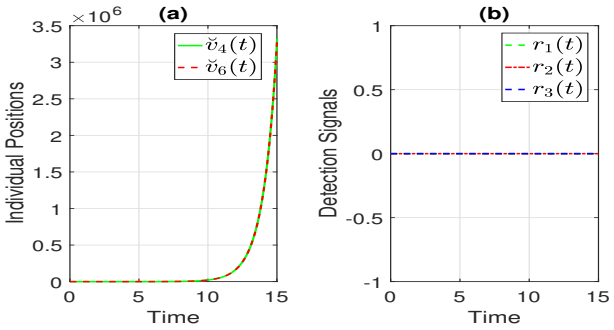

Figure 5. Trajectories of velocities (a) and attack-detection signals (b).

Theorem 2, to detect the cooperative ZDA we can consider the periodic topology switching sequence in Figure 4 (b): $\mathbf{L} = \left\{ \underbrace{\sigma(t_0) = 5}_{\tau_0=3}, \underbrace{\sigma(t_1) = 6}_{\tau_1=1} \right\}$. We assume that the attacker can modify any connection in the scope of attackable links. The trajectories of attack-detection signals in Figure 6 demonstrate that the observer (54) under Algorithm 1 succeeds in detecting the cooperative ZDA (nonzero detection signals).

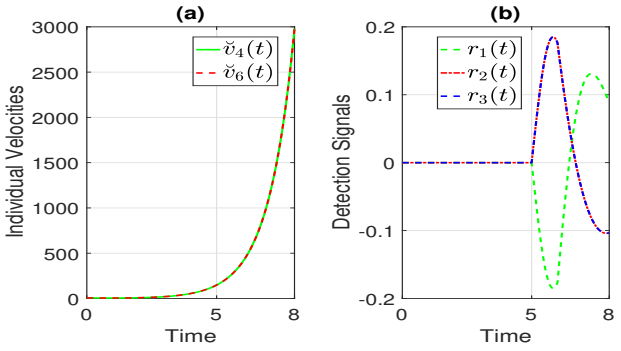

Figure 6. Trajectories of velocities (a) and attack-detection signals (b).

C. Comparison with Existing Works

The existing results on the detection of ZDA are summarized in Table I. Since $|\mathbb{M}| = 1$ in Figure 1 and $|\mathbb{K}| = 1$ for intermittent ZDA, $|\mathbb{M}| = 3$ in Figure 4 and $|\mathbb{K}| = 6$ for cooperative ZDA, and the connectivity of all network topologies are the same as 1, which violate the conditions in Table I. Defense strategies that rely on only strategically changing system dynamics [23], [24], while are effective against conventional ZDA and inspired us to analyze more sophisticated scenarios in this paper, implicitly assume that the attacker has no awareness of the aforementioned defense. Hence, the intermittent ZDA (when the system is unobservable) or cooperative ZDA (when the system is observable) cannot be detected by these methods. We also note that none of the prior work explicitly takes the issue of privacy/observability of initial/final states into account as we have pursued in this work.

VIII. Conclusion

In this paper, we have first introduced two ZDA variations for a scenario where the attacker is informed about the switching strategy of the defender: intermittent ZDA where the attacker pauses, updates and resumes ZDA in conjunction with the knowledge of switching topologies, and cooperative ZDA where the attacker employs a stealthy topology attack to render the switching topology defense ineffective. We have then studied conditions for a defender to detect these attacks, and subsequently based on these conditions, we have proposed an attack detection algorithm. The proposed defense strategy can detect both of the proposed ZDA variations, without requiring any knowledge of the set of misbehaving agents or the start, pause and resume times of the attack. Moreover, this strategy achieves asymptotic consensus and tracking in the absence of an attack without limiting the magnitudes of the coupling weights or the number of monitored agents.

Our analysis suggests an interesting trade-off among the switching cost, the duration of an undetected attack, the convergence speed to consensus and tracking. Analyzing this fundamental trade-off through the lens of game theory and multi-objective optimization constitutes a part of our future research.

Table I
Conditions for Detection of ZDA

Reference	Conditions	Dynamics
[12]	size of input-output linking is smaller than $\lvert\mathbb{K}\rvert$	Continuous Time
[18]	connectivity is not smaller than $2\lvert\mathbb{K}\rvert + 1$	Discrete Time
[19]	$\lvert\mathbb{K}\rvert$ is smaller than connectivity	Discrete Time
[20]	the minimum vertex separator is larger than $\lvert\mathbb{K}\rvert + 1$	Discrete Time
[21]	single attack, i.e., $\lvert\mathbb{K}\rvert = 1$	Continuous Time

Appendix A: Auxiliary Lemmas

In this section, we present auxiliary lemmas that are used in the proofs of the main results of this paper.

Lemma 4: [47] 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 that satisfies

$$\sum_{m=0}^{l-1} \nu_m \mu(\mathcal{A}_m) < 0, \tag{59}$$

where $\nu_m = \frac{\tau_m}{\sum_{i=0}^{l-1} \tau_i}$; then the switched system system is uniformly asymptotically stable for every positive $\tau = \sum_{i=0}^{l-1} \tau_i$.

Lemma 5: [48] Consider the Vandermonde matrix:

$$\mathcal{H} \triangleq \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & \cdots & \vdots \\ a_1^{n-1} & a_2^{n-1} & \cdots & a_n^{n-1} \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

Its determinant is $\det(\mathcal{H}) = (-1)^{\frac{n^2-n}{2}} \prod_{i<j} (a_i - a_j)$.

Lemma 6: Consider the matrix Q_r that satisfies (11). If $\lambda_2(\mathcal{L}_r) > 0$, then

$$\ker\left(\left[Q_r^\top \right]_{2:|\mathbb{V}|,:} \right) = \left\{ \mathbf{1}_{|\mathbb{V}|}, \mathbf{0}_{|\mathbb{V}|} \right\}. \tag{60}$$

Proof: The proof follows from a contradiction argument. We assume that (60) does not hold, i.e., there exists a vector $\psi = [\varphi_1, \ldots, \varphi_{|\mathbb{V}|}]^\top$ such that

$$\psi \notin \text{span}\left\{\mathbf{1}_{|\mathbb{V}|}, \mathbf{0}_{|\mathbb{V}|}\right\}, \tag{61}$$

and $\left[Q_r^\top\right]_{2:|\mathbb{V}|,:}\psi = \mathbf{0}_{|\mathbb{V}|-1}$. Then, it follows from (11) that

$$\mathcal{L}_r\psi = Q_r\Lambda_r Q_r^\top\psi = Q_r\mathbf{0}_{|\mathbb{V}|} = \mathbf{0}_{|\mathbb{V}|}. \tag{62}$$

From [43], we know that an undirected graph is connected if and only if $\lambda_2(\mathcal{L}_r) > 0$, and further the null space of the Laplacian matrix $\mathcal{L}_r$ of a connected graph is spanned by the vector $\mathbf{1}_{|\mathbb{V}|}$. We obtain from (62) that $\varphi_1 = \ldots = \varphi_{|\mathbb{V}|}$, which contradicts with (61). Thus, (60) holds. This concludes the proof. ∎

Appendix B: Proof of Proposition 1

Based on average variables $\bar{x}(t) \triangleq \frac{1}{|\mathbb{V}|}\sum_{i\in\mathbb{V}} x_i(t)$ and $\bar{v}(t) \triangleq \frac{1}{|\mathbb{V}|}\sum_{i\in\mathbb{V}} v_i(t)$, we define the following fluctuation terms:

$$\tilde{x}_i(t) \triangleq x_i(t) - \bar{x}(t), \tag{63a}$$

$$\tilde{v}_i(t) \triangleq v_i(t) - \bar{v}(t), \tag{63b}$$

which implies that

$$\mathbf{1}_{|\mathbb{V}|}^\top\tilde{x}(t) = 0, \text{ for } t \geq t_0 \tag{64a}$$

$$\mathbf{1}_{|\mathbb{V}|}^\top\tilde{v}(t) = 0, \text{ for } t \geq t_0. \tag{64b}$$

Considering (1b), (8) and $a_{ij}^{\sigma(t)} = a_{ji}^{\sigma(t)}$, we have

$$\begin{aligned}\dot{\bar{v}}(t) &= \frac{1}{|\mathbb{V}|}\sum_{i\in\mathbb{V}}\dot{v}_i(t) = \frac{1}{|\mathbb{V}|}\sum_{i\in\mathbb{V}}u_i(t)\\ &= \frac{1}{|\mathbb{V}|}\sum_{i\in\mathbb{V}}(-v_i(t) + \sum_{j\in\mathbb{V}}a_{ij}^{\sigma(t)}(x_j(t) - x_i(t)))\\ &= -\frac{1}{|\mathbb{V}|}\sum_{i\in\mathbb{V}}v_i(t) = -\bar{v}(t),\end{aligned}$$

which, in conjunction with (63b), leads to

$$\begin{aligned}&\dot{\tilde{v}}_i(t)\\ &= \dot{v}_i(t) - \dot{\bar{v}}(t) = u_i(t) + \bar{v}(t)\\ &= -v_i(t) + \sum_{j\in\mathbb{V}}a_{ij}^{\sigma(t)}(x_j(t) - x_i(t)) + \bar{v}(t)\\ &= -(v_i(t) - \bar{v}(t)) + \sum_{j\in\mathbb{V}}a_{ij}^{\sigma(t)}((x_j(t) - \bar{x}(t)) - (x_i(t) - \bar{x}(t)))\\ &= -\tilde{v}_i(t) + \sum_{j\in\mathbb{V}}a_{ij}^{\sigma(t)}(\tilde{x}_j(t) - \tilde{x}_i(t)), i \in \mathbb{V}.\end{aligned} \tag{65}$$

The dynamics of the second-order multi-agent system (1) with control input (8) can now be expressed equivalently as

$$\dot{\tilde{x}}(t) = \tilde{v}(t) \tag{66a}$$

$$\dot{\tilde{v}}(t) = -\tilde{v}(t) - \mathcal{L}_{\sigma(t)}\tilde{x}(t), \tag{66b}$$

where (66b) considers its equivalent form (65).

Let us define $\hat{x} \triangleq Q_r^\top\tilde{x}$ and $\hat{v} \triangleq Q_r^\top\tilde{v}$. Noting (11), the dynamics (66) can equivalently transform to

$$\dot{\hat{x}}(t) = \hat{v}(t) \tag{67a}$$

$$\dot{\hat{v}}(t) = -\hat{v}(t) - \Upsilon_{rs}\hat{x}(t), r, s \in \mathbb{S} \tag{67b}$$

where Υ_{rs} is defined in (12a). We note that it follows from (64) and (11b) that $\hat{x}_1(t) = \hat{v}_1(t) = 0$, $[\Upsilon_{rs}]_{1,:} = \mathbf{0}_{|\mathbb{V}|}^\top$ and $[\Upsilon_{rs}]_{:,1} = \mathbf{0}_{|\mathbb{V}|}$. Let us define $\theta \triangleq [\hat{x}_2 \ \vdots \ \ldots \ \vdots \ \hat{x}_{|\mathbb{V}|} \ \vdots \ \hat{v}_2 \ \vdots \ \ldots \ \vdots \ \hat{v}_{|\mathbb{V}|}]^\top$. Thus, the system (67) equivalently reduces to

$$\dot{\theta}(t) = \mathcal{A}_s\theta(t), s \in \mathbb{S} \tag{68}$$

with $\mathcal{A}_s$ given in (12b). Meanwhile, it is straightforward to verify that when $r = s$, $\mathcal{A}_s$ is Hurwitz. Therefore, the exists a $P > 0$ such that $\mu_P(\mathcal{A}_r) < 0$. Through setting on the dwell time of the topology indexed by r, (59) can be satisfied. By Lemma 4, the system (68) is uniformly asymptotically stable, i.e., for any initial condition, $\lim_{t\to\infty}\theta(t) = \mathbf{0}_{2|\mathbb{V}|-2}$, which implies that $\lim_{t\to\infty}Q^\top\tilde{x}(t) = \lim_{t\to\infty}Q^\top\tilde{v}(t) = \mathbf{0}_{|\mathbb{V}|}$. Since Q is full-rank, we have $\lim_{t\to\infty}\tilde{x}(t) = \lim_{t\to\infty}\tilde{v}(t) = \mathbf{0}_{|\mathbb{V}|}$. Then, (63) implies that $\lim_{t\to\infty}\tilde{x}_i(t) = \lim_{t\to\infty}\tilde{x}_j(t)$ and $\lim_{t\to\infty}\tilde{v}_i(t) = \lim_{t\to\infty}\tilde{v}_j(t), \forall i \neq j \in \mathbb{V}$. Here, we conclude that the second-order consensus is achieved, and we define $v^* = \lim_{t\to\infty}\tilde{v}_i(t), \forall i \in \mathbb{V}$. Then, substituting the second-order consensus into the system (1) with control input (8) yields the dynamics $\dot{v}^* = -v^*$, which implies a common zero velocity at steady state.

Appendix C: Proof of Proposition 2

Let us first define:

$$\mathbf{y} \triangleq \breve{y} - y. \tag{69}$$

It is straightforward to obtain dynamics from (3) and (7) as

$$\dot{\mathbf{z}}(t) = A_{\sigma(t)}\mathbf{z}(t) + \breve{g}(t) \tag{70a}$$

$$\mathbf{y}(t) = C\mathbf{z}(t) + D\breve{g}(t), \tag{70b}$$

where $\mathbf{z}(t)$ is defined in (33).

*1) **Proof of** (35):* Since $[\xi_k, \zeta_k) \subseteq [t_k, t_{k+1})$, $\sigma(t) = r$ for $t \in [\xi_k, \zeta_k)$. We denote $\Xi(s) \triangleq \mathfrak{L}\{\mathbf{z}(t)\}$, where $\mathfrak{L}(\cdot)$ stands for the Laplace transform operator. It follows from the attack signal (23) that $\mathfrak{L}\{\breve{g}(t)\} = (e^{-\xi_k s} - e^{-\zeta_k s})\frac{\breve{g}(\xi_k)}{s-\eta_r}$, $t \in [\xi_k, \zeta_k)$. Without loss of generality, we let $\sigma(t) = r$ for $t \in [t_k, t_{k+1})$. Then, the Laplace transform of the dynamics in (70) is obtained as

$$\begin{aligned}&(e^{-\xi_k s} - e^{-\zeta_k s})(s\Xi(s) - \mathbf{z}(\xi_k))\\ &= (e^{-\xi_k s} - e^{-\zeta_k s})A_r\Xi(s) + (e^{-\xi_k s} - e^{-\zeta_k s})\frac{\breve{g}(\xi_k)}{s-\eta_r},\end{aligned}$$

which is equivalent to

$$(e^{-\xi_k s} - e^{-\zeta_k s})\Xi(s) = \frac{(e^{-\xi_k s} - e^{-\zeta_k s})}{s\mathbf{1}_{2|\mathbb{V}|\times 2|\mathbb{V}|} - A_r}\left(\mathbf{z}(\xi_k) + \frac{\breve{g}(\xi_k)}{s-\eta_r}\right). \tag{71}$$

Expanding (26b) out yields

$$C\mathbf{z}(\xi_k) + D\breve{g}(\xi_k) = \mathbf{0}_{|\mathbb{M}|}, \tag{72}$$

$$\eta_r\mathbf{z}(\xi_k) - A_r\mathbf{z}(\xi_k) = \breve{g}(\xi_k), r \in \mathbb{T}. \tag{73}$$

Substituting (73) into (71) yields $(e^{-\xi_k s} - e^{-\zeta_k s})\Xi(s) = \frac{(e^{-\xi_k s} - e^{-\zeta_k s})}{s-\eta_r}\mathbf{z}(\xi_k)$, and the inverse Laplace transform of it gives (35).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2997363, IEEE Transactions on Automatic Control

*2) **Proof of** (34):* It follows from (35) and (70) that

$$\mathbf{y}(t)=e^{\eta_r(t-\xi_k)}\left(C\mathbf{z}(\xi_k)+D\breve{g}(\xi_k)\right),t\in[\xi_k,\zeta_k),k\in\mathbb{N}_0 \quad (74)$$

which combined with (72) results in $\mathrm{y}(t)=\mathbf{0}_{|\mathbb{M}|}$, or equivalently, $\breve{y}(t)=y(t)$, for any $t\in[\xi_k,\zeta_k)$.

We next prove (5) over non-attack interval of ZDA $[\zeta_k,\xi_{k+1})$. From (23) and (25), the dynamics (70) over such non-attack intervals of ZDA (subject to the monitored output attack as (25)) is described by

$$\dot{\mathbf{z}}(t)=A_{\sigma(t)}\mathbf{z}(t) \quad (75a)$$

$$\mathbf{y}(t)=C\mathbf{z}(t)+D\sum_{m=0}^{k}\breve{g}\left(\zeta_m^-\right),t\in[\zeta_k,\xi_{k+1}). \quad (75b)$$

It follows from (35) and (75a) that

$$\mathbf{z}(t) \quad (76)$$
$$=\begin{cases}e^{A_{\sigma(t_k)}(t-t_k)}\mathbf{z}(t_k), & t\in[t_k,\xi_k)\\ e^{\mathbf{1}_{2|\mathbb{V}|\times 2|\mathbb{V}|}\eta_r(t-\xi_k)+A_{\sigma(t_k)}(\xi_k-t_k)}\mathbf{z}(t_k), & t\in[\xi_k,\zeta_k)\\ e^{A_{\sigma(t_k)}(t-t_k-(\zeta_k-\xi_k))+\mathbf{1}_{2|\mathbb{V}|\times 2|\mathbb{V}|}\eta_r(\zeta_k-\xi_k)}\mathbf{z}(t_k), & t\in[\zeta_k,t_{k+1}).\end{cases}$$

We conclude from (69) that (34) is equivalent to

$$\mathbf{y}(t)\equiv\mathbf{0}_{|\mathbb{M}|}\ \text{on}\ [t_0,t_{k+1}). \quad (77)$$

For $D=\mathbf{0}_{|\mathbb{M}|\times 2|\mathbb{V}|}$, we note that (77) implies that the system (75) is unobservable for any $t\in[t_0,t_{k+1})$, $k\in\mathbb{N}_0$. It is immediate that

$$\mathbf{z}(t_k)\in\ker(\mathcal{O}_k)=\widehat{\mathbf{N}}_k^k,\ k\in\mathbb{N}_0. \quad (78)$$

We next show that $\mathbf{z}(t_{q-1})\in\widehat{\mathbf{N}}_{q-1}^k$ for $0\le q-1\le k$, through inductive argument. Let us suppose $\mathbf{z}(t_q)\in\widehat{\mathbf{N}}_q^k$. We obtain from (76) that $\mathbf{z}(t_q)=\mathbf{z}(t_q^-)=e^{\eta_{\sigma(t_{q-1})}(\zeta_{q-1}-\xi_{q-1})}e^{A_{\sigma(t_{q-1})}(\tau_{q-1}-(\zeta_{q-1}-\xi_{q-1}))}\mathbf{z}(t_{q-1})$, which, in conjunction with the fact of $e^{\eta_{\sigma(t_{q-1})}(\zeta_{q-1}-\xi_{q-1})}\neq 0$, leads to $\mathbf{z}(t_{q-1})\in e^{-A_{\sigma(t_{q-1})}(\tau_{q-1}-(\zeta_{q-1}-\xi_{q-1}))}\widehat{\mathbf{N}}_q^k$. Moreover, we note that (78) implies that $\mathbf{z}(t_{q-1})\in\ker(\mathcal{O}_{q-1})$. Therefore,

$$\mathbf{z}(t_{q-1})\in e^{-A_{\sigma(t_{q-1})}(\tau_{q-1}-(\zeta_{q-1}-\xi_{q-1}))}\widehat{\mathbf{N}}_q^k\cap\ker(\mathcal{O}_{q-1}), \quad (79)$$

where the right-hand expression is, in fact, the computation of $\widehat{\mathbf{N}}_{q-1}^k$, i.e., the unobservable space given by (28). Let $q=1$, we have $\mathbf{z}(t_0)\in\widehat{\mathbf{N}}_0^k$. Then, following the same steps in the proof of necessary condition in Theorem 1 of [44], we conclude that (34) holds if and only if there exists a non-zero vector $\mathbf{z}(t_0)$ such that

$$\mathbf{z}(t_0)\in\widehat{\mathbf{N}}_0^k. \quad (80)$$

For $D\neq\mathbf{0}_{|\mathbb{M}|\times 2|\mathbb{V}|}$, it follows from (72) and (75b) that $\mathbf{y}(\zeta_k)=\mathbf{y}\left(\zeta_k^-\right)=\mathbf{0}_{|\mathbb{M}|}$. Therefore, in this scenario, (77) holds only when $\dot{\mathbf{y}}(t)\equiv\mathbf{0}_{|\mathbb{M}|}$ on $[t_0,t_{k+1})$. Updating the observability matrix $\mathcal{O}_q$ in (22) by $\widetilde{\mathcal{O}}_q$ in (31) and following the same steps to derive (80), we conclude that (34) holds if and only if

$$\mathbf{z}(t_0)\in\widetilde{\mathbf{N}}_0^k, \quad (81)$$

where $\widetilde{\mathbf{N}}_0^k$ is recursively computed by (29) and (30).

In addition to (80) and (81), we conclude that if (26a) and (72) hold, regardless of $D\sum_{m=0}^{k}\breve{g}(\zeta_m^-)\neq\mathbf{0}_{|\mathbb{M}|}$ or $=\mathbf{0}_{|\mathbb{M}|}$, (34) always holds.

Appendix D: Proof of Proposition 3

Let us define $\tilde{e}\triangleq\left[\tilde{e}_x^\top\ \vdots\ \tilde{e}_v^\top\right]^\top\triangleq\widehat{z}-z$. Without loss of generality, we let $\sigma(t_{k+1})=s$. Noticing (39), we obtain from the dynamics (38) and (17) that

$$\dot{\tilde{e}}(t)=\widehat{A}_s\tilde{e}(t)+(\widehat{A}_s-A_s)z(t),t\in[t_{k+1},t_{k+2}) \quad (82a)$$
$$\widehat{y}(t)-y(t)=C\tilde{e}(t), \quad (82b)$$
$$\tilde{e}(t_{k+1})=\mathbf{0}_{|\mathbb{M}|}, \quad (82c)$$

from which we have

$$\widehat{y}(t)-y(t)=Ce^{\widehat{A}_s(t-t_{k+1})}\int_{t_{k+1}}^{t}e^{-\widehat{A}_s(\tau-t_{k+1})}((\widehat{A}_s-A_s)z(\tau))\mathrm{d}\tau,$$

and the corresponding derivatives

$$\begin{aligned}&\widehat{y}^{(d)}(t)-y^{(d)}(t)\\&=C\widehat{A}_s^de^{\widehat{A}_s(t-t_{k+1})}\int_{t_{k+1}}^{t}e^{-\widehat{A}_s(\tau-t_{k+1})}(\widehat{A}_s-A_s)z(\tau)\mathrm{d}\tau\\&\quad+\sum_{l=0}^{d-1}C\widehat{A}_s^l((\widehat{A}_s-A_s)z^{(d-1-l)}(t)).\end{aligned} \quad (83)$$

We note that under corrupted topology, the stealthy property $\widehat{y}(t)-y(t)=\mathbf{0}_{|\mathbb{M}|}$ for any $t\in[t_{k+1},t_{k+2})$ is equivalent to $\widehat{y}^{(d)}(t_{k+1})-y^{(d)}(t_{k+1})=\mathbf{0}_{|\mathbb{M}|}$ for $\forall d\in\mathbb{N}_0$, which is further equivalent to (40) by considering the solution (83).

Appendix E: Proof of Theorem 1

Without loss of generality, we let $\sigma(\zeta_k)=r\in\mathbb{T}$, and $\zeta_k<t_{k+1}$, $k\in\mathbb{N}$, i.e., attacker "pauses" ZDA at ζ_k. We now prove this theorem via a contradiction. We assume that the attack is not detectable in $[\zeta_k^-,\xi_{k+1})$, which is equivalent to

$$\mathbf{y}(t)=\mathbf{0}_{|\mathbb{M}|}\ \text{for any}\ t\in[\zeta_k^-,\xi_{k+1}), \quad (84)$$

where $\mathbf{y}(t)$ is defined in (69).

Considering the fact that given a differentiable function $f(t)$, $f(t)=0$ for any $t\in[a,b]$, if and only if $f(a)=0$ and $f^{(d)}(a)=0$, $\forall d\in\mathbb{N}$. We conclude from (75) that (84) at time ζ_k is equivalent to

$$\mathbf{y}^{(d)}(\zeta_k)=\begin{cases}C\mathbf{z}(\zeta_k)+D\sum_{m=0}^{k}\breve{g}(\zeta_m^-)=\mathbf{0}_{|\mathbb{M}|}, & d=0\\ CA_r^d\mathbf{z}(\zeta_k)=\mathbf{0}_{|\mathbb{M}|}, & \forall d\in\mathbb{N}.\end{cases} \quad (85)$$

With the definitions of A_r, C, D and $\mathbf{z}(\cdot)$ in (16b), (16c), (16e) and (33), the relation (85) can be further rewritten under different forms of observation as follows:

- Full Observation of Velocity, i.e., $c_{i1}=0$, $\forall i\in\mathbb{M}$,

$$C_2\mathbf{v}(\zeta_k)+D\sum_{m=0}^{k}\breve{g}\left(\zeta_m^-\right)=\mathbf{0}_{|\mathbb{M}|} \quad (86a)$$
$$C_2\mathbf{v}(\zeta_k)+C_2\mathcal{L}_r\mathbf{x}(\zeta_k)=\mathbf{0}_{|\mathbb{M}|} \quad (86b)$$
$$C_2\mathcal{L}_r^e\mathbf{v}(\zeta_k)=\mathbf{0}_{|\mathbb{M}|},\forall e\in\mathbb{N} \quad (86c)$$
$$C_2\mathcal{L}_r^d\mathbf{x}(\zeta_k)=\mathbf{0}_{|\mathbb{M}|},\forall d\in\mathbb{N}_{\ge 2} \quad (86d)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2997363, IEEE Transactions on Automatic Control

- Full Observation of Position, i.e., $c_{i2}=0$, $\forall i\in\mathbb{M}$,

$$C_1\mathbf{x}\left(\zeta_k\right)+D\sum_{m=0}^{k}\breve{g}\left(\zeta_m^-\right)=\mathbf{0}_{|\mathbb{M}|} \tag{87a}$$

$$C_1\mathcal{L}_r^e\mathbf{x}\left(\zeta_k\right)=\mathbf{0}_{|\mathbb{M}|},\forall e\in\mathbb{N} \tag{87b}$$

$$C_1\mathcal{L}_r^d\mathbf{v}\left(\zeta_k\right)=\mathbf{0}_{|\mathbb{M}|},\forall d\in\mathbb{N}_0 \tag{87c}$$

- Partial Observation, i.e., $c_{i1}\neq 0$ and $c_{i2}\neq 0$, $\forall i\in\mathbb{M}$,

$$C_1\mathbf{x}\left(\zeta_k\right)+C_2\mathbf{v}\left(\zeta_k\right)+D\sum_{m=0}^{k}\breve{g}\left(\zeta_m^-\right)=\mathbf{0}_{|\mathbb{M}|}, \tag{88a}$$

$$C_1\mathcal{L}_r^e\mathbf{x}\left(\zeta_k\right)+C_2\mathcal{L}_r^e\mathbf{v}\left(\zeta_k\right)=\mathbf{0}_{|\mathbb{M}|},\forall e\in\mathbb{N} \tag{88b}$$

$$(C_1-C_2)\mathcal{L}_r^d\mathbf{v}(\zeta_k)-C_2\mathcal{L}_r^{d+1}\mathbf{x}(\zeta_k)=\mathbf{0}_{|\mathbb{M}|},\forall d\in\mathbb{N}_0. \tag{88c}$$

Considering the definition of the vector $\mathbf{z}(t)$ in (33), and its continuity with respect to time, i.e., $\mathbf{z}\left(\zeta_k^-\right)=\mathbf{z}\left(\zeta_k\right)$, it follows from (35) and (23) that at time ζ_k^-,

$$\begin{bmatrix}\mathbf{z}\left(\zeta_k\right)\\ \hdashline -\breve{g}\left(\zeta_k^-\right)\end{bmatrix}=e^{\eta_r(\zeta_k^--\xi_k)}\begin{bmatrix}\mathbf{z}\left(\xi_k\right)\\ \hdashline -\breve{g}\left(\xi_k\right)\end{bmatrix}, \tag{89}$$

which, in conjunction with the fact of $e^{\eta_r(\zeta_k^--\xi_k)}\neq 0$ and the condition (26b), results in

$$\begin{bmatrix}\mathbf{z}\left(\zeta_k\right)\\ \hdashline -\breve{g}\left(\zeta_k^-\right)\end{bmatrix}\in\ker\left(\mathcal{P}_k\right). \tag{90}$$

With variables $\breve{g}(\zeta_k^-)$, $\bar{g}(\zeta_k^-)$, $\mathbf{z}\left(\zeta_k\right)$, A_r and $\mathcal{P}_k$ defined in (16f), (16g), (33), (16b) and (32), respectively, expanding (90) yields

$$\eta_r\mathbf{x}\left(\zeta_k\right)-\mathbf{v}\left(\zeta_k\right)=\mathbf{0}_{|\mathbb{V}|}, \tag{91}$$

$$-\bar{g}\left(\zeta_k^-\right)+\mathbf{v}(\zeta_k)+\mathcal{L}_r\mathbf{x}\left(\zeta_k\right)+\eta_r\mathbf{v}(\zeta_k)=\mathbf{0}_{|\mathbb{V}|}. \tag{92}$$

Before proceeding the rest of proof, we define the variables:

$$H_i\triangleq\left[\mathcal{U}_{ri}\mathbf{x}\left(\zeta_k\right)\right]_{2:|\mathbb{V}|}, \tag{93a}$$

$$\mathcal{D}_r\triangleq\text{diag}\left\{\lambda_2^2\left(\mathcal{L}_r\right),\ldots,\lambda_{|\mathbb{V}|}^2\left(\mathcal{L}_r\right)\right\}, \tag{93b}$$

$$\widetilde{\mathcal{H}}_r\triangleq\begin{bmatrix}\lambda_2^2(\mathcal{L}_r) & \cdots & \lambda_{|\mathbb{V}|}^2(\mathcal{L}_r)\\ \lambda_2^3(\mathcal{L}_r) & \cdots & \lambda_{|\mathbb{V}|}^3(\mathcal{L}_r)\\ \vdots & \cdots & \vdots\\ \lambda_2^{|\mathbb{V}|}(\mathcal{L}_r) & \cdots & \lambda_{|\mathbb{V}|}^{|\mathbb{V}|}(\mathcal{L}_r)\end{bmatrix}, \tag{93c}$$

$$\mathcal{H}_r\triangleq\begin{bmatrix}1 & \cdots & 1\\ \lambda_2(\mathcal{L}_r) & \cdots & \lambda_{|\mathbb{V}|}(\mathcal{L}_r)\\ \vdots & \cdots & \vdots\\ \lambda_2^{|\mathbb{V}|-2}(\mathcal{L}_r) & \cdots & \lambda_{|\mathbb{V}|}^{|\mathbb{V}|-2}(\mathcal{L}_r)\end{bmatrix}, \tag{93d}$$

where $\mathcal{U}_{ri}$ is given in (44).

A. Under Full Observation of Position or Velocity

Let us start with full observation of velocity. It follows from (11) that $\mathcal{L}_r^d=Q_r\Lambda_r^dQ_r^\top$ with Λ_r given in (11). Thus, (86d) is equivalent to $C_2Q_r\Lambda_r^dQ_r^\top\mathbf{x}\left(\zeta_k\right)=\mathbf{0}_{|\mathbb{M}|},\forall d\in\mathbb{N}_{\geq 2}$, which is further equivalent to

$$\sum_{l=1}^{|\mathbb{V}|}\lambda_l^d(\mathcal{L}_r)[Q_r]_{i,l}\left[Q_r^\top\right]_{l,:}\mathbf{x}\left(\zeta_k\right)=0,\forall d\in\mathbb{N},\forall i\in\mathbb{M} \tag{94}$$

with the consideration of the matrix C_2 defined in (16d) with $c_{i2}\neq 0,\forall i\in\mathbb{M}$. Further, recalling $\widetilde{\mathcal{H}}_r$, H_i and $\mathcal{U}_{ri}$ from (93c), (93a) and (44), from (94) we have

$$\widetilde{\mathcal{H}}_rH_i=\mathbf{0}_{|\mathbb{V}|-1},\forall i\in\mathbb{M}. \tag{95}$$

It can be verified from (93b)–(93d) that $\widetilde{\mathcal{H}}_r=\mathcal{H}_r\mathcal{D}_r$, from which we have $\det(\widetilde{\mathcal{H}}_r)=\det(\mathcal{H}_r)\det(\mathcal{D}_r)$. The matrix defined in (93b) shows if $\mathcal{L}_r$ has distinct eigenvalues, $\mathcal{D}_r$ is full-rank. In addition, by Lemma 5, the Vandermonde matrix $\mathcal{H}_\text{r}$ is full-rank; thus, $\widetilde{\mathcal{H}}_r$ is full-rank. Therefore, the solution of (95) is

$$H_i=\mathbf{0}_{|\mathbb{V}|-1},\forall i\in\mathbb{M}. \tag{96}$$

With the definitions in (44) and (93a), the equation (96) indicates that for $\forall i\in\mathbb{M}$,

$$\text{diag}\Big\{[Q_r]_{i,2},\ldots,[Q_r]_{i,|\mathbb{V}|}\Big\}\left[Q_r^\top\right]_{2:|\mathbb{V}|,:}\mathbf{x}\left(\zeta_k\right)=\mathbf{0}_{|\mathbb{V}|-1}. \tag{97}$$

We note that (44), (45) and (47) imply that $\exists i\in\mathbb{M}$: diag $\left\{[Q_r]_{i,2},\ldots,[Q_r]_{i,|\mathbb{V}|}\right\}$ is full-rank. Thus, from (97) we have $\left[Q_r^\top\right]_{2:|\mathbb{V}|,:}\mathbf{x}\left(\zeta_k\right)=\mathbf{0}_{|\mathbb{V}|-1}$. By Lemma 6, the solution of (97) is

$$\mathbf{x}_1\left(\zeta_k\right)=\ldots=\mathbf{x}_{|\mathbb{V}|}\left(\zeta_k\right). \tag{98}$$

Considering (86c), using the same method to derive (98), we obtain

$$\mathbf{v}_1\left(\zeta_k\right)=\ldots=\mathbf{v}_{|\mathbb{V}|}\left(\zeta_k\right). \tag{99}$$

Substituting (98) into (86b) yields $C_2\mathbf{v}\left(\zeta_k\right)=\mathbf{0}_{|\mathbb{M}|}$, which together with (99) results in

$$\mathbf{v}_1\left(\zeta_k\right)=\ldots=\mathbf{v}_{|\mathbb{V}|}\left(\zeta_k\right)=0. \tag{100}$$

For the full observation of position, using nearly the same analysis method employed above, we obtain the same results as (98) and (100).

Substituting (98) and (100) into (92) yields $\bar{g}\left(\zeta_k^-\right)=\mathbf{0}_{|\mathbb{V}|}$, and consequently, $\breve{g}\left(\zeta_k^-\right)=\mathbf{0}_{2|\mathbb{V}|}$. This means that there is no ZDA on the system at ζ_k^-, which contradicts the assumption that the attack is applied until ζ_k. Therefore, we conclude that under the full observation of position or velocity, the intermittent ZDA is detectable.

*1) **Full Observation of Velocity**:* To proceed with the proof of (48), we first need to obtain $\ker(\mathcal{O}_k)$ of the system (17) given in (22). The analysis of the kernel of the observability matrix $\mathcal{O}_k$ can follow the relation (85) with the setting of $D=\mathbf{0}_{|\mathbb{M}|\times 2|\mathbb{V}|}$. We note that (85) is equivalently represented by (86), (87) and (88). The results (98) and (99) are obtained without considering (86a), (87a) and (88a) which are the only terms involving D. Then, results similar to (98) and (99) can be obtained for the system in (17) as

$$x_1\left(\zeta_k\right)=\ldots=x_{|\mathbb{V}|}\left(\zeta_k\right)\ \text{ and }\ v_1\left(\zeta_k\right)=\ldots=v_{|\mathbb{V}|}\left(\zeta_k\right). \tag{101}$$

Further, with $D=\mathbf{0}_{|\mathbb{M}|\times 2|\mathbb{V}|}$, from (86a) with $\mathbf{v}\left(\zeta_k\right)$ replaced by $v\left(\zeta_k\right)$, we have $C_2v\left(\zeta_k\right)=\mathbf{0}_{|\mathbb{M}|}$, which combined with (101) yields $x_1\left(\zeta_k\right)=\ldots=x_{|\mathbb{V}|}\left(\zeta_k\right)$ and $v_1\left(\zeta_k\right)=\ldots=v_{|\mathbb{V}|}\left(\zeta_k\right)=0$. Thus, $\ker(\mathcal{O}_k)=\left\{\mathbf{0}_{2|\mathbb{V}|},\left[\mathbf{1}_{|\mathbb{V}|}^\top\ \vdots\ \mathbf{0}_{|\mathbb{V}|}^\top\right]^\top\right\}$. Since all of the elements in $\ker(\mathcal{O}_k)$ are the equilibrium points of the system (17), through the recursive computation of (20) and (21), we arrive at (48).

*2) **Full Observation of Position**:* To obtain $\ker(\mathcal{O}_k)$ under full observation of position, we can consider (87) with $D = \mathbf{0}_{|\mathbb{M}|\times 2|\mathbb{V}|}$. From (87a) and (98) we have $x_1(\zeta_k) = \ldots = x_{|\mathbb{V}|}(\zeta_k) = 0$. Then, we obtain from (100) (replace $\mathbf{v}_i(\zeta_k)$ by $v_i(\zeta_k)$) that $\ker(\mathcal{O}_k) = \{\mathbf{0}_{2|\mathbb{V}|}\}$, which means that if the monitored agents output full observation of positions, the system (17) is observable at t_k; thus (49) is obtained by the recursive computation of (20) and (21).

B. Under Partial Observation

The analysis of observability follows the same steps of that under full observation. With $C_1 = C_2$, from (88c) we have $C_2\mathcal{L}_r^{d+1}\mathbf{x}(\zeta_k) = 0, \forall d \in \mathbb{N}_0$. Employing the same steps to derive (98) under full observation of velocity, we obtain (98) as well under partial observation. Moreover, substituting (98) into (88b) and repeating the same steps, we arrive at (99). It is straightforward to verify from the dynamics (17) that $\mathbf{x}_1(t) = \ldots = \mathbf{x}_{|\mathbb{V}|}(t)$ and $\mathbf{v}_1(t) = \ldots = \mathbf{v}_{|\mathbb{V}|}(t)$ for any $t \geq t_0$, if and only if (99) and (98) hold. Finally, considering (88a) with the setting of $D = \mathbf{0}_{|\mathbb{M}|\times 2|\mathbb{V}|}$, we have $C_1\mathbf{x}(\zeta_k) + C_2\mathbf{v}(\zeta_k) = \mathbf{0}_{|\mathbb{M}|}$, from which we have $\ker(\mathcal{O}_k) = \left\{\mathbf{0}_{2|\mathbb{V}|}, \left[\mathbf{1}_{|\mathbb{V}|}^\top \vdots -\mathbf{1}_{|\mathbb{V}|}^\top\right]^\top\right\}, \forall k \in \mathbb{N}_0$, and then (50) is obtained by computation of (20) and (21).

Under the condition (51), $\mathbf{z}(\zeta_k) \in \mathbf{N}_0^k$, which in conjunction with (91) implies $\eta_r = -1$. Substituting (98), (99) and $\eta_r = -1$ into (92) yields $\bar{g}(\zeta_k^-) = \mathbf{0}_{|\mathbb{V}|}$, and consequently, $\breve{g}(\zeta_k^-) = \mathbf{0}_{2|\mathbb{V}|}$. This means that there is no ZDA on the system at ζ_k^-, which contradicts the assumption that the attack is applied until ζ_k.

APPENDIX F: PROOF OF THEOREM 2

With the definition of $C_j, j = 1, 2$, in (16d), we can rewrite (82) as

$$\dot{\tilde{e}}_x(t) = \tilde{e}_v(t), \tag{102a}$$

$$\dot{\tilde{e}}_v(t) = -\tilde{e}_v(t) - \widehat{\mathcal{L}}_s\tilde{e}_x(t) - \left(\widehat{\mathcal{L}}_s - \mathcal{L}_s\right)x(t), \tag{102b}$$

$$\widehat{y}(t) - y(t) = C_1\tilde{e}_x(t) + C_2\tilde{e}_v(t), t \in [t_{k+1}, t_{k+2}) \tag{102c}$$

$$\tilde{e}_x(t_{k+1}) = \mathbf{0}_{|\mathbb{V}|}, \tilde{e}_v(t_{k+1}) = \mathbf{0}_{|\mathbb{V}|}. \tag{102d}$$

We define $\mathcal{C} \triangleq \text{diag}\{c_{12}, \ldots, c_{|\mathbb{D}|2}\}$, where the diagonal entries are from C_2 defined in (16d). According to (52) and $|\mathbb{D}| \leq |\mathbb{M}|$ (implied by (43)), the matrix $\mathcal{C}$ is invertible. Now, considering (42), we have

$$C_2\left(\widehat{\mathcal{L}}_s - \mathcal{L}_s\right) = \left[\begin{array}{c:c} \mathcal{C}\mathfrak{L}_s & \mathbf{0}_{|\mathbb{D}|\times(|\mathbb{M}|-|\mathbb{D}|)} \\ \hdashline \mathbf{0}_{(|\mathbb{M}|-|\mathbb{D}|)\times|\mathbb{D}|} & \mathbf{0}_{(|\mathbb{M}|-|\mathbb{D}|)\times(|\mathbb{M}|-|\mathbb{D}|)} \end{array}\right], \tag{103}$$

which, in conjunction with invertible matrix $\mathcal{C}$ and the definitions of A_s in (16b) and $\widehat{A}_s$ in (37), implies that if $C_2\left(\widehat{\mathcal{L}}_s - \mathcal{L}_s\right)x^{(d)}(t_{k+1}) = \mathbf{0}_{|\mathbb{M}|}, \forall d \in \mathbb{N}_0$, then

$$\left(\widehat{A}_s - A_s\right)z^{(d)}(t_{k+1}) = \mathbf{0}_{2|\mathbb{V}|}, \forall d \in \mathbb{N}_0. \tag{104}$$

Under the dynamics (102) and the relation (104), the necessary condition (40) of guaranteeing stealthy property of cooperative ZDA is equivalently written as

$$C_2\left(\widehat{\mathcal{L}}_s - \mathcal{L}_s\right)\mathcal{L}_s^d x(t_{k+1}) = \mathbf{0}_{|\mathbb{M}|}, \forall d \in \mathbb{N}_0 \tag{105a}$$

$$C_2\left(\widehat{\mathcal{L}}_s - \mathcal{L}_s\right)\mathcal{L}_s^d v(t_{k+1}) = \mathbf{0}_{|\mathbb{M}|}, \forall d \in \mathbb{N}_0. \tag{105b}$$

We assume that the topology attack in system (36) can ensure that the stealthy property (5) of ZDA holds. Noticing (103) and the dynamics (17), the equation (105) is equivalent to $\mathcal{C}\mathfrak{L}_{\sigma(t_{k+1})}\chi^{(m)}(t_{k+1}) = \mathbf{0}_{|\mathbb{D}|}, \forall m \in \mathbb{N}_0$, where $\chi(t_{k+1}) \triangleq \left[x_1(t_{k+1}) \vdots \ldots \vdots x_{|\mathbb{D}|}(t_{k+1})\right]^\top$. Since $\mathcal{C}$ is invertible, we have

$$\mathfrak{L}_{\sigma(t_{k+1})}\chi^{(m)}(t_{k+1}) = \mathbf{0}_{|\mathbb{D}|}, \forall m \in \mathbb{N}_0. \tag{106}$$

As $\mathfrak{L}_{\sigma(t_{k+1})}$ is the elementary row transformation of a Laplacian matrix, there exists an elementary row operator $E \in \mathbb{R}^{|\mathbb{D}|\times|\mathbb{D}|}$ such that $\widehat{\mathfrak{L}}_{\sigma(t_{k+1})} \triangleq E\mathfrak{L}_{\sigma(t_{k+1})}$ is a Laplacian matrix. Pre-multiplying both sides of (106) by E yields

$$\widehat{\mathfrak{L}}_{\sigma(t_{k+1})}\chi^{(m)}(t_{k+1}) = \mathbf{0}_{|\mathbb{D}|}, \forall m \in \mathbb{N}_0. \tag{107}$$

It is well-known that the null space of the Laplacian matrix of a connected graph is spanned by the vector with all ones. From (107) we conclude that $\exists i, j \in \mathbb{D} : x_i^{(m)}(t_{k+1}) = x_j^{(m)}(t_{k+1}), t_{k+1} \geq t_0, \forall m \in \mathbb{N}_0$, which can be rewritten as

$$\left(\mathrm{e}_i^\top - \mathrm{e}_j^\top\right)x^{(m)}(t_{k+1}) = 0, \forall m \in \mathbb{N}_0 \tag{108}$$

where e_i denotes a vector of length $|\mathbb{D}|$ with a single nonzero entry with value 1 in its ith position.

Due to the dynamics (17), the equation (108) leads to

$$\left(\mathrm{e}_i^\top - \mathrm{e}_j^\top\right)\mathcal{L}_r^m x(t_{k+1}) = 0, \forall m \in \mathbb{N}_0 \tag{109a}$$

$$\left(\mathrm{e}_i^\top - \mathrm{e}_j^\top\right)\mathcal{L}_r^m v(t_{k+1}) = 0, \forall m \in \mathbb{N}_0. \tag{109b}$$

It follows from (11) that $\mathcal{L}_r^d = Q_r\Lambda_r^d Q_r^\top$ with Λ_r given in (11c), substituting which into (109) yields that for $\forall m \in \mathbb{N}$,

$$\sum_{l=2}^{|\mathbb{V}|}\lambda_l^m(\mathcal{L}_r)\Big([Q_r]_{i,l} - [Q_r]_{j,l}\Big)\big[Q_r^\top\big]_{l,:}x(t_{k+1}) = 0, \tag{110a}$$

$$\sum_{l=2}^{|\mathbb{V}|}\lambda_l^m(\mathcal{L}_r)\Big([Q_r]_{i,l} - [Q_r]_{j,l}\Big)\big[Q_r^\top\big]_{l,:}v(t_{k+1}) = 0. \tag{110b}$$

Then, with the definitions

$$\mathcal{D}_{ij} \triangleq \text{diag}\Big\{[Q_r]_{i,2} - [Q_r]_{j,2}, \ldots, [Q_r]_{i,|\mathbb{V}|} - [Q_r]_{j,|\mathbb{V}|}\Big\}, \tag{111}$$

$$f \triangleq [Q_r]_{2:|\mathbb{V}|,:}^\top x(t_{k+1}), \tag{112}$$

following the same derivations from (94) to (95), we arrive at

$$\widetilde{\mathcal{H}}_r\mathcal{D}_{ij}f = \mathbf{0}_{|\mathbb{V}|-1}, \forall i \in \mathbb{M}, \tag{113}$$

where $\widetilde{\mathcal{H}}_r$ is given in (93c). Using the same analysis to derive (96), we conclude that under the condition (46), the solution of (113) is $\mathcal{D}_{ij}f = \mathbf{0}_{|\mathbb{V}|-1}$. Since $\mathcal{D}_{ij}$ given by (111) is full-rank under the condition (53), we have $f = \mathbf{0}_{|\mathbb{V}|-1}$. Then, noticing (112), by Lemma 6 we arrive at

$$x_1(t_{k+1}) = \ldots = x_{|\mathbb{V}|}(t_{k+1}). \tag{114}$$

Repeating the same procedure of deriving (114) from (110a), we conclude $v_1(t_{k+1}) = \ldots = v_{|\mathbb{V}|}(t_{k+1})$ from (110b), which means that the second-order consensus is achieved at t_{k+1}, i.e., $x_i(t_{k+1}) = x_j(t_{k+1})$ and $v_i(t_{k+1}) = v_j(t_{k+1}), \forall i \neq j \in \mathbb{V}$. It is straightforward to verify from the dynamics (66) that the second-order consensus is achieved at some time $t < \infty$ if and only if the individual initial conditions are identical, i.e., $x_i(t_0) = x_j(t_0)$ and $v_i(t_0) = v_j(t_0)$. Hence, the cooperative ZDA is undetectable only in the case of identical initial condition that corresponds to the steady state.

APPENDIX G: PROOF OF THEOREM 3

We define $\mathbf{e}_x(t) \triangleq q(t) - \breve{x}(t)$ and $\mathbf{e}_v(t) \triangleq w(t) - \breve{v}(t)$. The dynamics of tracking errors in the presence of the attack obtained from (54) and (14) are given as:

$$\dot{\mathbf{e}}_{x_i}(t) = \mathbf{e}_{v_i}(t), \tag{115a}$$

$$\dot{\mathbf{e}}_{v_i}(t) = -\mathbf{e}_{v_i}(t) + \sum_{i\in\mathbb{V}} a_{ij}^{\sigma(t)}\big(\mathbf{e}_{x_j}(t) - \mathbf{e}_{x_i}(t)\big) - \begin{cases}\breve{g}_i(t), i\in\mathbb{K}\\ 0,\ i\in\mathbb{V}\backslash\mathbb{K}\end{cases} - \begin{cases} r_i(t), & c_{i1}\neq 0, i\in\mathbb{M}\\ \int_{t_0}^{t} r_i(b)\mathrm{d}b, & c_{i1}=0, i\in\mathbb{M}\\ 0, & i\in\mathbb{V}\backslash\mathbb{M}\end{cases} \tag{115b}$$

$$r_i(t) = c_{i1}\mathbf{e}_{x_i}(t) + c_{i2}\mathbf{e}_{v_i}(t) - d_i\breve{g}_i(t), i\in\mathbb{M}. \tag{115c}$$

The attack is not detected by the observer (54) means that $r_i(t) = 0, i\in\mathbb{M}$, for any $t \geq t_0$. Substituting it into the above equation results in

$$\dot{\mathbf{e}}_{x_i}(t) = \mathbf{e}_{v_i}(t)$$

$$\dot{\mathbf{e}}_{v_i}(t) = -\mathbf{e}_{v_i}(t) + \sum_{i\in\mathbb{V}} a_{ij}^{\sigma(t)}\big(\mathbf{e}_{x_j}(t) - \mathbf{e}_{x_i}(t)\big) - \begin{cases}\breve{g}_i(t), i\in\mathbb{K}\\ 0,\ i\in\mathbb{V}\backslash\mathbb{K}\end{cases}$$

$$r_i(t) = c_{i1}\mathbf{e}_{x_i}(t) + c_{i2}\mathbf{e}_{v_i}(t) - d_i\breve{g}_i(t), i\in\mathbb{M}$$

which has the same form of dynamics as that of (14). Therefore, the analysis of ZDA variations in the observer (54) follows the same analysis of the system (14). Moreover, the required condition (52) implies that the monitored agents output full observations of velocity or partial observations: either (48) or (50) implies (19). Hence, the topology attacker cannot infer the real-time full states of the non-monitored agents, and the topology attacker has to consider the scope of the target connections implied by (43). Therefore, the proof of the first statement follows from Theorems 1 and 2.

In the absence of attacks, the system matrix of system (115) is $\widehat{\mathcal{A}}_{\sigma(t)}$ defined in (55). Since the condition (46) implies that all of the switching topologies provided to Algorithm 1 are connected graphs and condition (52) implies (57), the matrix $\widehat{\mathcal{A}}_{\sigma(t)}$ is Hurwitz by Lemma 3. Thus, there exists a $P > 0$ such that both (59) and (58) hold. Hence, the proof of the second statement follows from Proposition 1 and Lemma 3.

REFERENCES

[1] Y. Mao, H. Jafarnejadsani, P. Zhao, E. Akyol, and N. Hovakimyan, "Detectability of intermittent zero-dynamics attack in networked control systems," in *Proceedings of the 58th IEEE Conference on Decision and Control*, pp. 5605–5610, 2019.

[2] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[3] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1465–1476, 2004.

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

[5] L.-Y. Lu and C.-C. Chu, "Consensus-based droop control synthesis for multiple dics in isolated micro-grids," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2243–2256, 2015.

[6] Q. Li and D. Rus, "Global clock synchronization in sensor networks," *IEEE Transactions on Computers*, vol. 55, no. 2, pp. 214–226, 2006.

[7] W. Ren and E. Atkins, "Distributed multi-vehicle coordinated control via local information exchange," *International Journal of Robust and Nonlinear Control*, vol. 17, no. 10-11, pp. 1002–1033, 2007.

[8] A. Abdessameud and A. Tayebi, "Attitude synchronization of a group of spacecraft without velocity measurements," *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2642–2648, 2009.

[9] B. B. Johnson, S. V. Dhople, A. O. Hamadeh, and P. T. Krein, "Synchronization of nonlinear oscillators in an LTI electrical power network," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 3, pp. 834–844, 2014.

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

[11] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in *International Conference on Critical Infrastructure Protection*, pp. 73–82, 2007.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] J. Chen, J. Wei, W. Chen, H. Sandberg, K. H. Johansson, and J. Chen, "Protecting positive and second-order systems against undetectable attacks," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8373–8378, 2017.

[22] 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.

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

[24] Y. Mao, E. Akyol, and Z. Zhang, "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.

[25] H. Hartenstein, K. P. Laberteaux *et al.*, "A tutorial survey on vehicular ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164–171, 2008.

[26] S. K. Mazumder, *Wireless networking based control*. Springer, 2011.

[27] M. Xue, W. Wang, and S. Roy, "Security concepts for the dynamics of autonomous vehicle networks," *Automatica*, vol. 50, no. 3, pp. 852–857, 2014.

[28] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1294–1305, 2013.

[29] R. Skowyra, L. Xu, G. Gu, V. Dedhia, T. Hobson, H. Okhravi, and J. Landry, "Effective topology tampering attacks and defenses in software-defined networks," in *48th IEEE/IFIP International Conference on Dependable Systems and Networks*, pp. 374–385, 2018.

[30] 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.

[31] S. Rahili and W. Ren, "Distributed continuous-time convex optimization with time-varying cost functions," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1590–1605, 2017.

[32] 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.

[33] J. R. Lawton, R. W. Beard, and B. J. Young, "A decentralized approach to formation maneuvers," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 6, pp. 933–941, 2003.

[34] 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.

[35] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st international conference on High Confidence Networked Systems*, pp. 55–64, 2012.

[36] G. Xie and L. Wang, "Consensus control for a class of networks of dynamic agents: switching topology," in *Annual American Control Conference*, pp. 1382–1387, 2006.

[37] T. Smith, "Hacker jailed for revenge sewage attacks," https://www.theregister.co.uk/2001/10/31/hacker_jailed_for_revenge_sewage/, accessed 2001-10-31.

[38] A. Teixeira, G. Dán, H. Sandberg, R. Berthier, R. B. Bobba, and A. Valdes, "Security of smart distribution grids: Data integrity attacks on integrated volt/var control and countermeasures," in *Annual American Control Conference*, pp. 4372–4378, 2014.

[39] H. J. van Waarde, P. Tesi, and M. K. Camlibel, "Topology reconstruction of dynamical networks via constrained Lyapunov equations," *IEEE Transactions on Automatic Control*, vol. 64, no. 10, pp. 4300–4306, 2019.

[40] Y. Mao and E. Akyol, "On inference of network topology and confirmation bias in cyber-social networks," *to appear in IEEE Transactions on Signal and Information Processing over Networks (Special Issue on Network Topology Inference), arXiv:1908.09472.*

[41] A. D. Ames and S. Sastry, "Characterization of Zeno behavior in hybrid systems using homological methods," in *Annual American Control Conference*, pp. 1160–1165, 2005.

[42] J. Weimer, S. Kar, and K. H. Johansson, "Distributed detection and isolation of topology attacks in power networks," in *Proceedings of the 1st international conference on High Confidence Networked Systems*, pp. 65–72, 2012.

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

[44] A. Tanwani, H. Shim, and D. Liberzon, "Observability for switched linear systems: characterization and observer design," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 891–904, 2013.

[45] J. Zhang and L. Sankar, "Implementation of unobservable state-preserving topology attacks," in *North American Power Symposium*, pp. 1–6, 2015.

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

[47] 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.

[48] R. A. Horn and C. R. Johnson, "Topics in matrix analysis," *Cambridge UP, New York*, 1991.

Yanbing Mao received the B.S. degree in Electronic Information Science and Technology from Liaocheng University, Shandong, China, in 2010, and the M.E. degree in Circuits & Systems from University of Electronic Science and Technology of China, Sichuan, China, in 2013. He received the Ph.D. degree in Electrical and Computer Engineering from State University of New York at Binghamton, NY, USA, in 2019. He is currently a postdoctoral research associate in the Department of Mechanical Science and Engineering at University of Illinois at Urbana–Champaign. His research interests include self-driving vehicles, security and privacy of networked cyber-physical systems, social networks and social learning.

Hamidreza Jafarnejadsani is an Assistant Professor in the Department of Mechanical Engineering at Stevens Institute of Technology. Before joining the faculty at Stevens, he was a postdoctoral research associate in the Department of Computer Science at the University of Illinois at Urbana-Champaign (UIUC). Hamid received his Ph.D. degree in Mechanical Engineering from UIUC in 2018, where he worked in the Advanced Controls Research Laboratory. He got his B.S. and M.S. degrees both in Mechanical Engineering from the University of Tehran in 2011 and the University of Calgary in 2013, respectively. His research interests include control, optimization, robotics, machine learning, and cyber-physical security. In particular, he is interested in resilient control of intelligent autonomous systems in uncertain and adversarial environments.

Pan Zhao is a Postdoc Researcher in the Department of Mechanical Science and Engineering at the University of Illinois at Urbana–Champaign. He received the B.S. and M.S. degrees from Beihang University, Beijing, China, in 2009 and 2012, respectively, and the Ph.D. degree from the Department of Mechanical Engineering at the University of British Columbia, Vancouver, BC, Canada, in 2018. During his Ph.D. study, he received the Vanier Canada Graduate Scholarship from Natural Sciences and Engineering Research Council of Canada. From 2012 to 2013, he was a modeling & simulation engineer in Hirain Technologies, Beijing, China. His current research interests include robust adaptive control, gain-scheduled control and reinforcement learning.

Emrah Akyol is an Assistant Professor of Electrical and Computer Engineering at the State University of New York at Binghamton. He received the Ph.D. degree in 2011 from the University of California at Santa Barbara. From 2006 to 2007, he held positions at Hewlett-Packard Laboratories and NTT Docomo Laboratories, both in Palo Alto, CA where he worked on topics in image and video compression. From 2013 to 2014, Dr. Akyol was a postdoctoral researcher in the Electrical Engineering Department at University of Southern California, and between 2014 and 2017, in the Coordinated Science Laboratory at University of Illinois at Urbana-Champaign. His current research focuses on information processing challenges associated with socio-cyber-physical systems. He is a senior member of IEEE.

Naira Hovakimyan received her MS degree in Theoretical Mechanics and Applied Mathematics in 1988 from Yerevan State University in Armenia. She got her Ph.D. in Physics and Mathematics in 1992 from the Institute of Applied Mathematics of Russian Academy of Sciences in Moscow, majoring in optimal control and differential games. Before joining the faculty of UIUC in 2008, she spent time as a research scientist at Stuttgart University in Germany, French Institute for Research in Computer Science and Automation (INRIA) in France, Georgia Institute of Technology, and she was on faculty of Aerospace and Ocean Engineering of Virginia Tech during 2003-2008. She is currently a W. Grafton and Lillian B. Wilkins Professor of Mechanical Science and Engineering at UIUC. In 2015 she was named inaugural director for Intelligent Robotics Lab of Coordinated Science Laboratory at UIUC. She was the recipient of the SICE International scholarship for the best paper of a young investigator in the VII ISDG Symposium (Japan, 1996), the 2011 recipient of AIAA Mechanics and Control of Flight Award, the 2015 recipient of SWE Achievement Award, the 2017 recipient of IEEE CSS Award for Technical Excellence in Aerospace Controls, and the 2019 recipient of AIAA Pendray Aerospace Literature Award. In 2014 she was awarded the Humboldt prize for her lifetime achievements. In 2015 she was awarded the UIUC Engineering Council Award for Excellence in Advising. She is Fellow and life member of AIAA, a Fellow of IEEE, and a member of SIAM, AMS, SWE, ASME and ISDG. She is cofounder and chief scientist of IntelinAir. Her research interests are in control and optimization, autonomous systems, neural networks, game theory and their applications in aerospace, robotics, mechanical, agricultural, electrical, petroleum, biomedical engineering and elderly care.