2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Detectability of Intermittent Zero-Dynamics Attack in Networked Control Systems

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

Abstract— **This paper analyzes stealthy attacks, particularly the zero-dynamics attack (ZDA) in networked control systems. ZDA hides the attack signal in the null-space of the state-space representation of the control system and hence it cannot be detected via conventional detection methods. A natural defense strategy builds on changing the null-space via switching through a set of topologies. In this paper, we propose a realistic ZDA variation where the attacker is aware of this topology-switching strategy, and hence employs the policy to avoid detection: "pause (update and resume) attack" before (after) topology switching to evade detection. We first systematically study the proposed ZDA variation, and then develop defense strategies under the realistic assumptions. Particularly, we characterize conditions for detectability of the proposed ZDA variation, 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. We provide numerical results that demonstrate our theoretical findings.**

I. INTRODUCTION

Security concerns regarding the networked control systems pose significant challenges to their use and wide deployment, as highlighted by the recent incidents including distributed denial-of-service (DDOS) attack on Estonian web sites [1], Maroochy water breach [2] and cyber attacks on smart grids [3]. Particularly, a special class of "stealthy" attacks, namely the "zero-dynamics attack" (ZDA), poses a formidable security challenge [4]–[6]. The main idea behind the machinery of ZDA is to hide the attack signal in the null-space of the state-space representation of the networked control system so that the attack can evade conventional detection methods that are based on the observation signal (hence, the name "stealthy" attack). The objective of such a stealthy attack can vary from manipulating the controller to accept false data which would yield the system towards a desired (e.g., unstable) state to altering the system trajectory by maliciously changing the system dynamics (e.g., a stealthy topology attack).

Recent research efforts in this area have mainly focused on two directions: i) novel ZDA variations for particular systems [7], and ii) defense strategies [8]–[12]. For example, Park et al. [7] designed a robust ZDA for a particular stochastic cyber-physical system, for which the attack-detection signal can be guaranteed to stay below a threshold over a finite horizon. Jafarnejadsani et al. [5], [13] proposed a multi-rate $\mathcal{L}_1$ adaptive controller which can detect ZDA in the sampled-data control systems, by removing certain unstable zeros of discrete-time systems [4], [6].

Most of the prior work on defense strategies for the original ZDA require limiting assumptions regarding the connectivity of network topology and the number of the misbehaving agents (i.e., the agents under attack) [8]–[11]. For example, the detection approach in [12] works only for the scenario of single misbehaving agent in second-order systems [10], i.e., if there are multiple agents compromised by the attacker, the defense strategy cannot detect ZDA. Teixeira et al. [14] showed that the strategic changes in system dynamics could be utilized by defender to detect ZDA, but the approach taken in [14] requires the attack-starting times to be the initial time and known to the defender. In other words, the defense strategy can fail to work if the defender does not know the attack starting time, as is practically the case for most stealthy attack scenarios.

Towards designing a practical ZDA defense strategy, strategic topology switching is proposed in [15], [16]. The approach here is based on changing the topology through a carefully crafted set of topologies so that the ZDA can be detected. Controlling topology in this manner is physically feasible thanks to recent developments in mobile computing, wireless communication and sensing [17], [18]. We note, in passing, that the idea of using the changes in the state-space dynamics to detect ZDA first appeared in [14], albeit a realistic mechanism (e.g., changing the system topology) to achieve that objective was only very recently studied in [15], [16]. However, the defense strategy in [15], [16] still relies on the critical assumption of a naive attacker that does not take the topology switching strategy of the defender into account.

In this paper, we systematically address the following problem: can an attack comprising intermittently pausing, updating and resuming ZDA based on the knowledge of the sequence of topologies that the system goes through, be successful? We study this attack strategy, which we refer to as *intermittent ZDA*, in detail. We next analyze possible optimal defense strategies, beyond switching the topology, against the proposed intermittent ZDA. We demonstrate our theoretical findings via numerical simulation results.

II. PRELIMINARIES

A. Notations

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, and

Y. Mao and E. Akyol are with the Department of Electrical and Computer Engineering, Binghamton University–SUNY, Binghamton, NY, 13902 USA (e-mail: {ymao3, eakyol}@binghamton.edu).

H. Jafarnejadsani, 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: {jafarne2, panzhao2, nhovakim}@illinois.edu).

978-1-7281-1398-2/19/$31.00 ©2019 IEEE

let $\mathbb{N}$ represent 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. $\ker(Q) \triangleq \{y : Qy = \mathbf{0}_n, Q \in \mathbb{R}^{n\times n}\}$, $A^{-1}\mathbb{F} \triangleq \{y : Ay \in \mathbb{F}\}$. 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. For a matrix $W \in \mathbb{R}^{n\times n}$, $[W]_{i,j}$ denotes the element in row i and column j.

B. Zero-Dynamics Attack

Consider the following system, where $\breve{g}(t)$ represents an attack signal:

$$\dot{\breve{z}}(t) = A\breve{z}(t) + B\breve{g}(t), \tag{1a}$$
$$\breve{y}(t) = C\breve{z}(t) + D\breve{g}(t), \tag{1b}$$

and $\breve{z}(t) \in \mathbb{R}^{\bar{n}}$, $\breve{y}(t) \in \mathbb{R}^{\bar{m}}$, $\breve{g}(t) \in \mathbb{R}^{\bar{o}}$, $A \in \mathbb{R}^{\bar{n}\times\bar{n}}$, $B \in \mathbb{R}^{\bar{n}\times\bar{o}}$, $C \in \mathbb{R}^{\bar{m}\times\bar{n}}$, $D \in \mathbb{R}^{\bar{m}\times\bar{o}}$.

Definition 1: [19] The attack signal $\breve{g}(t) = ge^{\eta t}$ is a ZDA, if there exist scalar $\eta \in \mathbb{C}$, and nonzero vectors $\mathbf{z}_0 \in \mathbb{R}^{\bar{n}}$ and $g \in \mathbb{R}^{\bar{o}}$, that satisfy

$$\left[\begin{array}{c} \mathbf{z}_0 \\ \hdashline -g \end{array}\right] \in \ker\left(\left[\begin{array}{c:c} \eta\mathbf{1}_{\bar{n}\times\bar{n}} - A & B \\ \hdashline -C & D \end{array}\right]\right). \tag{2}$$

III. System Description

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

$$\dot{x}_i(t) = v_i(t), \tag{3a}$$
$$\dot{v}_i(t) = u_i(t), \quad i = 1, \ldots, n, \tag{3b}$$

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[1].

The interaction among the 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 the 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 the 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 $i \neq j$.

Definition 2: [27] The agents in the system (3) are said to achieve the asymptotic consensus with final zero common velocity if for any initial condition:

$$\lim_{t\to\infty} |x_i(t) - x_j(t)| = 0 \text{ and } \lim_{t\to\infty} |v_i(t)| = 0, \forall i, j \in \mathbb{V}. \tag{4}$$

[1]Several real-world networked systems that can be represented by (3) including the second-order consensus [20], flocking [21], swarming [22], velocity synchronization and regulation of relative distances [23], and their applications in the decentralized formation control of mobile robots [24] and spacecrafts [25], and distributed continuous-time optimization [26], etc.

A. Control Protocol

In this paper, we will propose a defense strategy that is based on topology switching. We now borrow a control protocol that involves topology switching from [27], [28] to achieve consensus according to (4) for the agents in system (3):

$$u_i(t) = -v_i(t) + \sum_{j\in\mathbb{V}} a_{ij}^{\sigma(t)}(x_j(t) - x_i(t)), i \in \mathbb{V}, \tag{5}$$

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 the communication graph. For a system with finite agents, the topology set $\mathbb{S}$ has finite elements as well.

B. System in the Presence of Attacks

We let $\mathbb{K} \subseteq \mathbb{V}$ denote the set of misbehaving agents, i.e., the agents under attack. We let the increasingly ordered set $\mathbb{M} \triangleq \{1, 2, \ldots\} \subseteq \mathbb{V}$ denote the set of monitored agents. Under time-dependent switching topology, the multi-agent system in (3), with the control input given by (5) and the outputs of monitoring agents in $\mathbb{M}$ subject to the ZDA signal $\breve{g}_i(t)$, can be written as

$$\dot{\breve{x}}_i(t) = \breve{v}_i(t) \tag{6a}$$
$$\dot{\breve{v}}_i(t) = -\breve{v}_i(t) + \sum_{i\in\mathbb{V}} a_{ij}^{\sigma(t)}(\breve{x}_j(t) - \breve{x}_i(t)) + \begin{cases} \breve{g}_i(t), i \in \mathbb{K} \\ 0, \ i \in \mathbb{V}\backslash\mathbb{K} \end{cases} \tag{6b}$$
$$\breve{y}_i(t) = c_{i1}\breve{x}_i(t) + c_{i2}\breve{v}_i(t) + d_i\breve{g}_i(t), i \in \mathbb{M}, \tag{6c}$$

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

Remark 1: If $d_i \neq 0$ for some $i \in \mathbb{M}$, (6c) means that the sensor outputs are under attack. However, the attack form in (1) with the attack policy computation (2) indicate that the ZDA signals for sensor outputs and control inputs are not independent of each other.

The system in (6) can be equivalently rewritten in the form of a switched system under attack:

$$\dot{\breve{z}}(t) = A_{\sigma(t)}\breve{z}(t) + \breve{g}(t) \tag{7a}$$
$$\breve{y}(t) = C\breve{z}(t) + D\breve{g}(t), \tag{7b}$$

where

$$\breve{z}(t) \triangleq \left[\breve{x}_1(t) \ \vdots \ \ldots \ \vdots \ \breve{x}_{|\mathbb{V}|}(t) \ \vdots \ \breve{v}_1(t) \ \vdots \ \ldots \ \vdots \ \breve{v}_{|\mathbb{V}|}(t)\right]^\top, \tag{8a}$$
$$A_{\sigma(t)} \triangleq \left[\begin{array}{c:c} \mathbf{0}_{|\mathbb{V}|\times|\mathbb{V}|} & \mathbf{1}_{|\mathbb{V}|\times|\mathbb{V}|} \\ \hdashline -\mathcal{L}_{\sigma(t)} & -\mathbf{1}_{|\mathbb{V}|\times|\mathbb{V}|} \end{array}\right], \tag{8b}$$
$$C \triangleq \left[\begin{array}{c:c} C_1 & C_2 \end{array}\right], \tag{8c}$$
$$C_j \triangleq \left[\mathrm{diag}\{c_{1j}, \ldots, c_{|\mathbb{M}|j}\} \ \vdots \ \mathbf{0}_{|\mathbb{M}|\times(|\mathbb{V}|-|\mathbb{M}|)}\right], j = 1, 2 \tag{8d}$$
$$D \triangleq \left[\mathbf{0}_{|\mathbb{M}|\times|\mathbb{V}|} \ \vdots \ \mathrm{diag}\{d_1, \ldots, d_{|\mathbb{M}|}\} \ \vdots \ \mathbf{0}_{|\mathbb{M}|\times(|\mathbb{V}|-|\mathbb{M}|)}\right], \tag{8e}$$
$$\breve{g}(t) \triangleq \left[\mathbf{0}_{|\mathbb{V}|}^\top \ \vdots \ \bar{g}^\top(t)\right]^\top, \tag{8f}$$
$$[\bar{g}(t)]_i \triangleq \begin{cases} \breve{g}_i(t), i \in \mathbb{K} \\ 0, \quad i \in \mathbb{V}\backslash\mathbb{K}. \end{cases} \tag{8g}$$

In addition, we consider the system (7) in the absence of attacks:

$$\dot{z}(t) = A_{\sigma(t)} z(t), \tag{9a}$$
$$y(t) = Cz(t). \tag{9b}$$

IV. Problem Formulation

We make the following assumptions on the attacker and the defender.

Assumption 1: The attacker

- is aware that the changes in system dynamics are used by the defender (system operator);
- knows the output matrix, the initial topology and the switching times before the first topology switching;
- needs a non-negligible time to compute and update the attack policy and identify the newly activated topology.
- can record the newly obtained knowledge of network topology into her memory.

Assumption 2: The defender

- designs the switching times (when to switch) and the switching topologies (what topology to switch to);
- chooses candidate agents to monitor, i.e., the monitoring agent set $\mathbb{M}$, for attack detection;
- knows that the states of (3) in the absence of attack are continuous with respect to time, i.e., $x(t^-) = x(t) = x(t^+)$ and $v(t^-) = v(t) = v(t^+)$;
- has no knowledge of the attack starting, pausing and resuming times, and the misbehaving agents.

A. Attack-Starting Time

We now use the following to describe the system (1) in the absence of attacks:

$$\dot{z}(t) = Az(t), \tag{10a}$$
$$y(t) = Cz(t). \tag{10b}$$

Let us first recall the properties of ZDA to review the prior defense strategies.

Lemma 1: [19] Consider the systems (1) and (10). Under the ZDA policy (2), the outputs and system states satisfy

$$y(t) = \breve{y}(t), t \geq 0 \tag{11}$$
$$\breve{z}(t) = z(t) + \mathbf{z}_0 e^{\eta t}. \tag{12}$$

If the attack-starting time, denoted by κ, is not the initial time, the ZDA signal in (1) and the state (12) would intuitively update as

$$\breve{y}(t) = \begin{cases} g e^{\eta(t-\kappa)}, & t \geq \kappa \\ \mathbf{0}_{\bar{\mathbb{o}}}, & \text{otherwise}, \end{cases}. \tag{13}$$

$$\breve{z}(t) = \begin{cases} z(t), & t \in [0, \kappa) \\ z(t) + \mathbf{z}_0 e^{\eta(t-\kappa)}, & t \in [\kappa, \infty). \end{cases}. \tag{14}$$

We note that the system (1) is not subject to attack when $t < \kappa$, which results in the first term in (14). Consequently, $\breve{z}(\kappa) = z(\kappa^-)$. Moreover, after the attacker launches attack at $t = \kappa$, according to the second term in (14) we have $\breve{z}(\kappa) = z(\kappa) + \mathbf{z}_0$. Since $\mathbf{z}_0$ is a non-zero vector, we have $\breve{z}(\kappa) \neq z(\kappa) = \breve{z}(\kappa^-)$, i.e., if $\kappa \neq 0$, the attack causes "jump" of the system state $\breve{z}(\kappa)$, which contradicts with the continuity of $\breve{z}(\cdot)$ w.r.t. time. Therefore, we conclude here that the defense strategy against ZDA – according to Definition 1 – implicitly assumes that the attack-starting time is the initial time.

B. Naive Attacker

Strategically changing system dynamics has been demonstrated to be an effective approach to detect system-based stealthy attacks, see e.g., ZDA [14] and $\mathcal{C}_k$/$\mathcal{C}$ stealthy attacks [29]. The core idea behind this defense strategy is the intentional mismatch between the models of the attacker and the defender. Specifically, the attacker uses the original system dynamics to make the stealthy attack decision before the system starts to operate, while the defender strategically changes the system dynamics at some operating point in time. However, this defense strategy assumes that the attacker has no capability of inferring the altered system dynamics. Otherwise, the attacker can make a synchronous attack decision according to the inferred changed dynamics to evade the detection.

To remove these unrealistic assumptions, [15] first investigated the attack policy for the attack signal (13), subject to the properties (11) and (14). A corresponding defense strategy of topology switching was then proposed to detect this ZDA variation in the multi-agent systems like (3). However, the attack signal (13) in [15] assumes that once the attacker launches the attack, she never changes the attack strategy. This further indicates the attacker's assumption that the defender follows the defense strategy proposed in [14], [29] to switch the topology only several times after the system starts the operation. In other words, the attacker assumes that without the knowledge of the attack-starting time the defender cannot switch the topology infinitely many times over infinite time to detect her. Otherwise, the attacker can try to infer the topologies online and update the stealthy attack decision accordingly.

We note that the recently developed inference algorithms can use the data of the agents' states in a *non-negligible* time interval to exactly infer either undirected [30] or directed network topology [31], [32]. If the attacker has such inference ability and is aware of the topology-switching defense strategy, as made in Assumption 1, she can avoid detection via the following strategy:

- pauses the attack before the incoming switching times;
- resumes, and if necessary, updates the attack, after the newly switched/activated topology is inferred.

In this paper, we first systematically study the behavior of such realistic ZDA variation that is referred to as intermittent ZDA. We then investigate its detectability. The two problems are stated formally as follows:

Problem I: Following which attack policy, the attacker can make stealthy attack decision?

Problem II: Following which defense strategy, the defender can detect intermittent ZDA?

V. Problem I: Attack Policy

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 the monitored output of the system (6) with intermittent pausing and resuming behaviors are described as:

$$\breve{g}_i(t) = \begin{cases} g_i^{\sigma(t_k)} e^{\eta_{\sigma(t_k)}(t-\xi_k)}, & t \in [\xi_k, \zeta_k) \subseteq [t_k, t_{k+1}) \\ 0, & \text{otherwise.} \end{cases} . \quad (15)$$

To analyze this ZDA, we review the monitored output (6c) at the first "pausing" time ζ_0:

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

which implies that $\breve{y}_i\left(\zeta_0^-\right) = \breve{y}_i\left(\zeta_0\right)$ if and only if $d_i\breve{g}_i\left(\zeta_0^-\right) = d_i\breve{g}_i\left(\zeta_0\right)$, 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 defender knows that 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 (**??**), the attacker cannot completely pause the attack, i.e., whenever the attacker pauses injecting ZDA signals to control inputs at pausing time ζ_k, she must continue to inject the same attack signals to monitored outputs (6c):

$$\breve{y}(t) = C\breve{z}(t) + D\sum_{m=0}^{k}\breve{g}\left(\zeta_m^-\right), t \in [\zeta_k, \xi_{k+1}). \quad (16)$$

Based on the above analysis, for the ZDA policy consisting of "pausing attack" and "resuming attack" behaviors to remain stealthy, it should satisfy (16) and

$$\mathbf{z}(t_0) \in \widehat{\mathbf{N}}_1^k \bigcap \widetilde{\mathbf{N}}_1^k, \quad (17a)$$

$$\left[\begin{array}{c} \mathbf{z}(\xi_k) \\ \hdashline -\breve{g}(\xi_k) \end{array}\right] \in \ker(\mathcal{P}_r), \forall \sigma(\xi_k) \in \mathbb{T} \quad (17b)$$

, where

$$\widehat{\mathbf{N}}_k^k = \ker(\mathcal{O}_k), \quad (18)$$

$$\widehat{\mathbf{N}}_q^k = \ker(\mathcal{O}_q) \bigcap e^{-A_{\sigma(t_q)}(\tau_q-(\zeta_q-\xi_q))}\mathbf{N}_{q+1}^m, 1 \le q \le m-1 \quad (19)$$

$$\widetilde{\mathbf{N}}_k^k = \ker(\widetilde{\mathcal{O}}_k), \quad (20)$$

$$\widetilde{\mathbf{N}}_q^k = \ker(\widetilde{\mathcal{O}}_q) \bigcap e^{-A_{\sigma(t_q)}(\tau_q-(\zeta_q-\xi_q))}\mathbf{N}_{q+1}^m, 1 \le q \le m-1 \quad (21)$$

$$\mathcal{O}_r \triangleq \left[(C)^\top \;\vdots\; \left(CA_r^2\right)^\top \;\vdots\; \dots \;\vdots\; \left(CA_r^{2|\mathbb{V}|-1}\right)^\top\right]^\top, \quad (22)$$

$$\widetilde{\mathcal{O}}_r \triangleq \left[(CA_r)^\top \;\vdots\; \left(CA_r^2\right)^\top \;\vdots\; \dots \;\vdots\; \left(CA_r^{2|\mathbb{V}|}\right)^\top\right]^\top, \quad (23)$$

$$\mathcal{P}_r \triangleq \left[\begin{array}{c:c} \eta_r\mathbf{1}_{2|\mathbb{V}|\times 2|\mathbb{V}|} - A_r & \mathbf{1}_{2|\mathbb{V}|\times 2|\mathbb{V}|} \\ -C & D \end{array}\right], \quad (24)$$

$$\mathbf{z} = \left[\mathbf{x}^\top \;\vdots\; \mathbf{v}^\top\right]^\top \triangleq \breve{z} - z = \left[\breve{x}^\top - x^\top \;\vdots\; \breve{v}^\top - v^\top\right]^\top, \quad (25)$$

with τ_q denoting the dwell time of the activated topology, i.e., $\tau_q = t_q - t_{q-1}$.

Remark 2: By the proof of Theorem in [33], $\widehat{\mathbf{N}}_q^k$ is obtained via recursive computation consisting of (18) and (19), which holds for $\widetilde{\mathbf{N}}_q^k$ as well. The recursive computations indicate that the attacker also needs the knowledge of inferred switched topologies recorded in her memory to make attack decision.

Proposition 1: [34] Under the stealthy attack policy consisting of (16) and (17), the states and monitored outputs of the systems (9) and (7) in the presence of attack signal (15) satisfy

$$\breve{y}(t) = y(t), t \in [t_0, t_{k+1}), \quad (26)$$

$$\breve{z}(t) = z(t) + e^{\eta_{\sigma(t_k)}(t-\xi_k)}\mathbf{z}(\xi_k), t \in [\xi_k, \zeta_k). \quad (27)$$

VI. Problem II: Detectability

The following theorem presents the detectability of intermittent ZDA.

Theorem 1: [34] Consider the system (7) in the presence of attack signals (15). Under the defense strategy:

$$\mathcal{L}_r \text{ has distinct eigenvalues for } \forall r \in \mathbb{S} \quad (28a)$$

$$\exists i \in \mathbb{M} : [Q_r]_{i,j} \neq 0, \forall j \in \mathbb{V}, \forall r \in \mathbb{S} \quad (28b)$$

- 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

$$\mathbf{N}_1^\infty = \left\{\mathbf{0}_{2|\mathbb{V}|}, \left[\mathbf{1}_{|\mathbb{V}|}^\top \;\vdots\; \mathbf{0}_{|\mathbb{V}|}^\top\right]^\top\right\}; \quad (29)$$

- 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

$$\mathbf{N}_1^\infty = \left\{\mathbf{0}_{2|\mathbb{V}|}\right\}; \quad (30)$$

- 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

$$\mathbf{N}_1^\infty = \left\{\mathbf{0}_{2|\mathbb{V}|}, \left[\mathbf{1}_{|\mathbb{V}|}^\top \;\vdots\; -\mathbf{1}_{|\mathbb{V}|}^\top\right]^\top\right\}; \quad (31)$$

and the intermittent ZDA is detectable if

$$\xi_0 > t_0 \text{ or } D = \mathbf{0}_{|\mathbb{M}|\times 2|\mathbb{V}|}, \quad (32)$$

where Q_r is the orthogonal matrix of $\mathcal{L}_r$, and $\mathbf{N}_1^\infty$ is recursively computed by

$$\mathbf{N}_m^m = \ker(\mathcal{O}_m) \quad (33a)$$

$$\mathbf{N}_q^m = \ker(\mathcal{O}_q) \bigcap e^{-A_{\sigma(t_q)}\tau_q}\mathbf{N}_{q+1}^m, 1 \le q \le m-1 \quad (33b)$$

with $\mathcal{O}_q$ given by (22).

Remark 3: Under the defense strategy (28), the result (30) means if the monitored agents output full observations of positions, the system (9) is observable at any time $t > t_0$.

As a result, using the available data of sensor outputs (9b), the attacker can infer the global system state and the global initial condition. While the results (29) and (31) show that if the monitored agents output full observations of velocity or partial observations, the privacy of full states of non-monitored agents are preserved, which would be useful in defending ZDA in cooperation with topology attack, since the stealthy topology needs the data of real-time states to decide target links to the attack [34]. Therefore, for the purpose of privacy preserving of non-monitored agents' states, consequently, restricting the scope of target links of the stealthy topology attack, the defender (system operator) has to abandon full observation of the position.

Remark 4: The compact defense strategy also includes a strategy on switching times. The building block of our defense strategy is the time-dependent topology switching that does not need a central unit to trigger links to switch. The critical reason that we do not consider state-dependent topology switching is that the attack signals injected into control input may generate Zeno behaviors [35], such that the control protocol (5) becomes infeasible. For this part of work, we refer the readers to [34], [36].

VII. Simulation

We consider a system with $n = 16$ agents. The initial position and velocity conditions are chosen randomly as $x(t_0) = [2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 4, 4, 4, 4, 4, 4]^\top$ and $v(t_0) = [6, 6, 6, 6, 6, 6, 6, 6, 8, 8, 8, 8, 8, 8, 8, 8]^\top$. The considered network topologies are given in Fig. 1, where the agents 1, 2 and 3 are the monitored agents, and the coupling weights are uniformly set as ones. We denote $r_i(t) = \breve{y}_i(t) - y_i(t), i \in \mathbb{M} = \{1, 2, 3\}$, as the attack-detection signals.

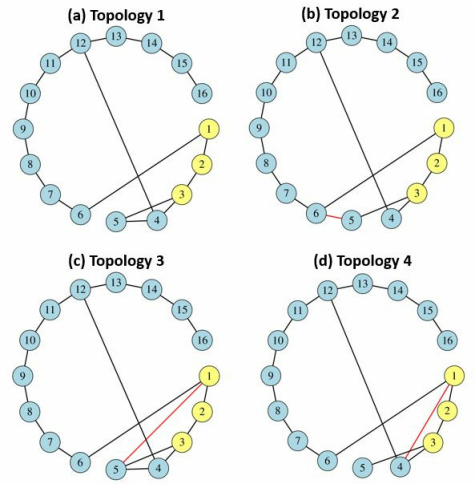

Fig. 1. Network topologies for intermittent ZDA, where agents 1, 2 and 3 are monitored agents.

We first consider the periodic topology switching sequence: $1 \to 2 \to 1 \to 2 \to \ldots$ with the dwell times $\tau_1 = \tau_2 = 2$. It can be verified that neither Topology 1 nor 2 in Fig. 1 satisfies the defense strategy (28). Therefore, the attacker can design undetectable intermittent ZDA by:

- inject false data $\mathbf{z}(t_0) = [0, 0, 0, -1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -0.5, 0.5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^\top$ to the data of initial condition;
- inject ZDA signals $\breve{g}_4(t) = -1.75e^{0.5t}$ and $\breve{g}_5(t) = 1.75e^{0.5t}$ to the local control inputs of agents 4 and 5 for the initial Topology 1;
- pause the ZDA before the incoming new topology, e.g. Topology 2, since the new topology has not been inferred yet;
- update the attack policy if necessary and resume the feasible attack after switching to the new topology finishes;
- iterate the last two steps.

The trajectories of some agents' positions and the attack-detection signals in Fig. 2 show that the designed intermittent ZDA is not detected, and the stealthy attack destabilizes the system.

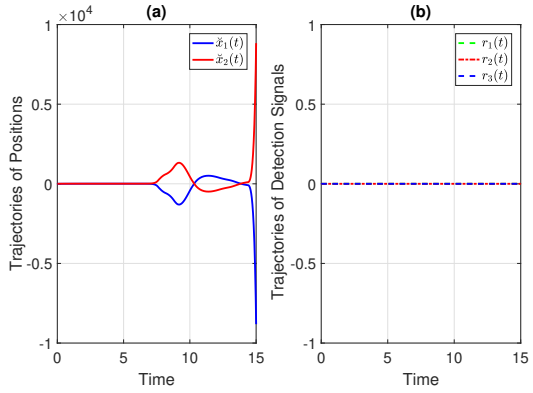

Fig. 2. Trajectories of individual positions and attack-detection signals: undetectable attack under switching Topologies 1 and 2.

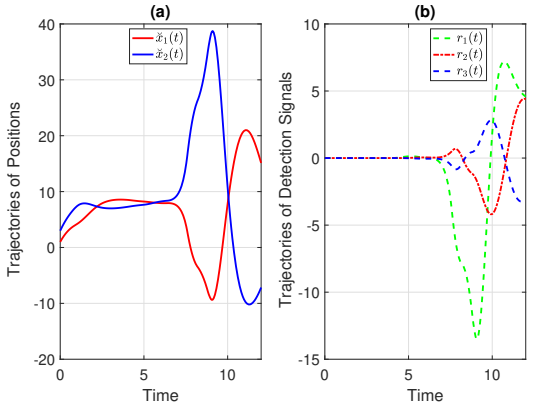

Fig. 3. Trajectories of individual positions and attack-detection signals: detectable attack under switching Topologies 3 and 4.

It is straightforward to show that both Topologies 3 and 4 in Fig. 1 satisfy the defense strategy (28). Hence, we can turn to the following periodic topology switching sequence at some time to detect the stealthy attack: $3 \to 4 \to 3 \to 4 \to \ldots$ with the dwell times $\tau_3 = \tau_4 = 2$. Under the periodic sequence, the trajectories of attack-detection

signals in Fig. 3 show that the stealthy intermittent ZDA is successfully detected.

VIII. Conclusion

This paper introduces one ZDA variant for a scenario where the attacker is informed about the switching strategy of the defender: intermittent ZDA where the attacker pauses, and updates and resumes ZDA in conjunction with the knowledge of switching topology and dwell times. A defense strategy without requiring any knowledge of the set of misbehaving agents or the start, pause and resume times of the attack is proposed to detect the intermittent ZDA.

IX. Acknowledgment

This work has been 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.

References

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

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

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

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

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

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

[7] 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 *IEEE 55th Conference on Decision and Control*, pp. 5085–5090, 2016.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] H. Jafarnejadsani, H. Lee, N. Hovakimyan, and P. Voulgaris, "Dual-rate $\mathcal{L}_1$ adaptive controller for cyber-physical sampled-data systems," in *IEEE 56th Annual Conference on Decision and Control*, pp. 6259–6264, 2017.

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

[15] Y. Mao, E. Akyol, and Z. Zhang, "Strategic topology switching for security-Part II: detection & switching topologies," *https://arxiv.org/abs/1711.11181*.

[16] Y. Mao and E. Akyol, "Detectability of cooperative zero-dynamics attack," in *56th Annual Allerton Conference on Communication, Control, and Computing*, pp. 227–234, 2018.

[17] H. Hartenstein, K. P. Laberteaux *et al.*, "A tutorial survey on vehicular ad hoc networks," *IEEE Communications magazine*, vol. 46, no. 6, p. 164, 2008.

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

[19] 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.

[20] 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.

[21] 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.

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

[23] 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.

[24] 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.

[25] W. Ren and R. Beard, "Decentralized scheme for spacecraft formation flying via the virtual structure approach," *Journal of Guidance, Control, and Dynamics*, vol. 27, no. 1, pp. 73–82, 2004.

[26] 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.

[27] 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.

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

[29] 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 *2014 American Control Conference*, pp. 4372–4378, 2014.

[30] H. J. van Waarde, P. Tesi, and M. K. Camlibel, "Topology reconstruction of dynamical networks via constrained lyapunov equations," *IEEE Transactions on Automatic Control, DOI: 10.1109/TAC.2019.2894585*.

[31] Y. Mao and E. Akyol, "On inference of network topology and confirmation bias in cyber-social networks," *https://arxiv.org/abs/1908.09472*.

[32] ——, "On network topology inference of social networks," *to appear in 57th Annual Allerton Conference on Communication, Control, and Computing*.

[33] 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, 2012.

[34] Y. Mao, H. Jafarnejadsani, P. Zhao, E. Akyol, and N. Hovakimyan, "Novel stealthy attack and defense strategies for networked control systems," *https://arxiv.org/abs/1908.09466*.

[35] A. D. Ames, A. Abate, and S. Sastry, "Sufficient conditions for the existence of zeno behavior," in *44th IEEE Conference on Decision and Control*, pp. 696–701, 2005.

[36] Y. Mao, E. Akyol, and Z. Zhang, "Strategic topology switching for security-Part I: Consensus & switching times," *https://arxiv.org/abs/1711.11183*.