Hybrid Attack Monitor Design to Detect Recurrent Attacks in a Class of Cyber-Physical Systems

Sean Phillips, Alessandra Duz, Fabio Pasqualetti and Ricardo G. Sanfelice

***Abstract*— In this paper, we study a security problem for attack detection in a class of cyber-physical systems consisting of discrete computerized components interacting with continuous agents. We consider an attacker that may inject recurring signals on both the physical dynamics of the agents and the discrete interactions. We model these attacks as additive unknown inputs with appropriate input signatures and timing characteristics. Using hybrid systems modeling tools, we design a novel hybrid attack monitor and, under reasonable assumptions, show that it is able to detect the considered class of recurrent attacks. Finally, we illustrate the general hybrid attack monitor using a specific finite time convergent observer and show its effectiveness on a simplified model of a cloud-connected network of autonomous vehicles.**

I. Introduction

Cyber-physical systems are at the core of a myriad of smart, innovative, and human-centric applications. Naturally, cyber-physical systems integrate discrete computerized components, communicational interfaces, and physical systems obeying some continuous-time dynamics. These systems offer open and physically accessible interfaces on both their cyber side (e.g., network interfaces and control algorithms) and their physical side (e.g., sensors and actuators). These interfaces can be exploited by adversaries to deny control, disable alarms, manipulate sensors, and initiate actions to adversely affect the outputs, or cause physical damage. Numerous examples show that it is likely unfeasible to secure all components from attack, and that the emphasis of defense must be on careful design, anomaly detection and localization, and reactive controls; see, e.g. [1], [2], [3].

Typical security methods for cyber-physical security rely on purely cyber mechanisms, such as data protection and authentication, or on anomaly detection techniques based on simple representations of the physical dynamics. While these methods have revealed important tradeoffs and limitations, the advent of sophisticated distributed systems and networks will necessitate new security theories for systems with complex structures and dynamics. Such coordinated attacks can degrade the performance of the system and hinder recovery while remaining undetected for long periods of time.

With security emerging as a major concern for cyber-physical systems, different modeling frameworks and protection schemes have been proposed for a variety of systems and attacks. While early works focus on static representations [4], [5], game-theoretic [6], [7], information theoretic [8], [9], and control-theoretic methods [10], [11] have been developed for dynamic models and attacks. These approaches represent a step towards addressing dynamic security features, and form the threshold for the new fundamental approach proposed here. To the best of our knowledge, most works study detection, identification, and resilience for systems with linear dynamics and attacks compromising integrity or availability of resources [12]. Yet, as systems evolve and become more complex, security methods based on simple dynamic models will likely be inapplicable or ineffective in practical scenarios. New security methods will have to be developed for systems with coupled cyber and physical dynamics, and constraints on the utilization of resources and timing. Despite notable developments in the theory of hybrid systems [15], security for hybrid systems remains a potentially transformative yet unexplored area.

The main contributions of this paper are as follows. We consider a class of cyber-physical systems modeled by linear continuous-time dynamics which may, at isolated time instances, discretely update their states. We focus our study on a class of recurrent attacks, modeled as additive inputs into both the continuous and discrete system dynamics, where there is sufficient allowed time between attack activity. We propose a generic hybrid attack monitor that detects the considered class of attacks and, when the attack has ceased, has an estimate that converges to the state of the cyber-physical system in finite time. Finally, we consider a specific form of the hybrid attack monitor and exemplify its use in a numerical example featuring a simplified model of cloud-connected networks for the surveillance of urban environments.

The rest of the paper is organized as follows. Section II contains a concise review of the hybrid systems framework employed and some preliminary definitions. Section III defines the model of the cyber-physical system, the attacker, the monitor, and the problem addressed in this paper. Section IV contains our hybrid monitor to detect attacks and our main results concerning the detection of such recurrent attacks in cyber-physical systems. In Section V, we illustrate our results numerically in an application involving cloud-connected aerial vehicles.

Sean Phillips and Ricardo G. Sanfelice are with the Computer Engineering Department, University of California at Santa Cruz, {seaphill, ricardo}@ucsc.edu. Their research has been partially supported by the National Science Foundation under CAREER Grant no. ECS-1450484 and Grant no. CNS-1544396, the Air Force Office of Scientific Research under Grant no. FA9550-16-1-0015, CITRIS, and the Banatao Institute at the University of California.

Alessandra Duz and Fabio Pasqualetti are with the Mechanical Engineering Department, University of California at Riverside, {alessd, fabiopas}@engr.ucr.edu. Their research has been supported in part by awards ARO 71603NSYIP and NSF ECCS1405330, and in part by CITRIS, the Banatao Institute at the University of California and the University of California at Riverside.

Notation: The set of real and natural numbers are denoted as $\mathbb{R}$ and $\mathbb{N}$, respectively. Given two vectors $u, v \in \mathbb{R}^n$, $|u| := \sqrt{u^\top u}$ and notation $[u^\top \ v^\top]^\top$ is equivalent to (u, v). Given a function $t \mapsto z(t)$, $z(t^+) = \lim_{s \searrow t^+} z(s)$.

II. Preliminaries

A hybrid system with inputs has data $\mathcal{H} = (C, f, D, G)$ and is defined by

$$\mathcal{H}: \qquad z \in \mathbb{R}^n \begin{cases} \dot{z} &= f(z,u) & (z,u) \in C \\ z^+ &\in G(z,u) & (z,u) \in D \\ r &= h(z) \end{cases} \tag{1}$$

where $z \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the input, f defines the flow map capturing the continuous dynamics, and C defines the flow set on which f is effective. The set-valued map G defines the jump map and models the discrete behavior, while D defines the jump set, the set of points where jumps are allowed. Solutions to $\mathcal{H}$ are given on *hybrid time domains*.[1] We define an *open* (and *closed*) shifted hybrid time domain, which is defined on a standard hybrid time domain starting from the hybrid time instant $(\underline{t}_i, \underline{j}_i)$ and ending at the time instant $(\overline{t}_i, \overline{j}_i)$, with open (closed) to the left leftmost subinterval and open (closed) to the right rightmost subinterval. We denote such a open shifted hybrid time domain by $((\underline{t}_i, \underline{j}_i), (\overline{t}_i, \overline{j}_i))$. Similarly, we denote a closed shifted hybrid time domain as $[(\underline{t}_i, \underline{j}_i), (\overline{t}_i, \overline{j}_i)]$, wherein the left leftmost subinterval and the right rightmost subinterval are closed. Note that, if the open shifted hybrid time domain has constant jump component, then the t component of the shifted hybrid time domain is an open interval, i.e., $((t_1, j), (t_2, j))$ is equivalent to $(t_1, t_2) \times \{j\}$.[2] A solution is a function defined on $\operatorname{dom}(\phi, u)$ $(= \operatorname{dom}\phi = \operatorname{dom} u)$ that satisfies the dynamics of $\mathcal{H}$ with the property that, for each $j \in \mathbb{N}$, $t \mapsto \phi(t, j)$ is absolutely continuous and $t \mapsto u(t, j)$ is Lebesgue measurable and locally essentially bounded on $\{t : (t, j) \in \operatorname{dom}(\phi, u)\}$.

In the next definition, we introduce the notion of δ-time detectable hybrid inputs, which will be utilized to define the class of recurring attacks that attackers can use; see Section III.

Definition 2.1 (δ-time detectable hybrid input): Given a positive constant δ, a hybrid input $(t, j) \mapsto (u_c(t, j), u_d(t, j))$ such that

- the set

$$\Gamma_c \ := \ \{(t, j) \in \operatorname{dom}(u_c, u_d) : u_c(t, j) \neq 0\} \tag{2}$$

is given by a collection of (maximally defined) shifted hybrid time domains $\mathcal{I}_i = [(\underline{t}_i^c, \underline{j}_i^c), (\overline{t}_i^c, \overline{j}_i^c))$ for $i \in \{1, 2, \ldots, N_c\}$, $N_c \in \mathbb{N} \cup \infty$.
- the set

$$\Gamma_d := \{(t, j) \in \operatorname{dom}(u_c, u_d) : u_d(t, j) \neq 0, (t, j+1) \in \operatorname{dom}(u_c, u_d)\}$$

is given by the collection of hybrid time instances

$$\{(t_i^d, j_i^d)\}_{i=1}^{N_d}, \qquad N_d \in \mathbb{N} \cup \infty;$$

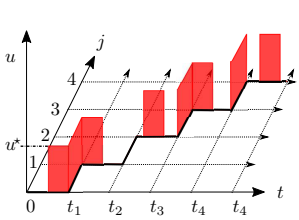
is said to be *δ-time detectable* if, for each $(t^*, j^*) \in \Gamma_c \cup \Gamma_d =: \Gamma$, there exist hybrid times $(t_1', j_1'), (t_2', j_2') \in \operatorname{dom}(u_c, u_d)$ such that $t_2' + j_2' \geq t_1' + j_1' + \delta$ satisfying the following properties:

1) the shifted hybrid time domain $[(t_1', j_1'), (t_2', j_2')) \subset \operatorname{dom}(u_c, u_d)$ is nonempty and has null intersection with Γ;
2) if $(t^*, j^*) \in \Gamma_c$

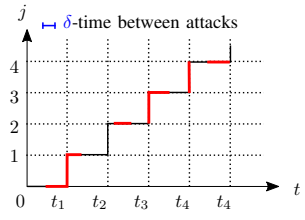
$$\begin{aligned} t_1' + j_1' &\geq \overline{t}_{i^*-1}^c + \overline{j}_{i^*-1}^c \qquad \text{if } i^* > 1 \\ t_2' + j_2' &\leq \underline{t}_{i^*}^c + \underline{j}_{i^*}^c \end{aligned}$$

where i^* is such that $(t^*, j^*) \in \mathcal{I}_{i^*}$;
3) if $(t^*, j^*) \in \Gamma_d \setminus \Gamma_c$, then $t_1' + j_1' \geq t_{i^*}^d + j_{i^*}^d$ or $t_2' + j_2' \leq t_{i^*}^d + j_{i^*}^d$ where i^* is such that $(t_{i^*}^d, j_{i^*}^d) \in \Gamma_c$.

The definition of a δ-time detectable hybrid input insures that the input remains zero for at least δ amount of hybrid time as soon as it becomes zero. The sets Γ_c and Γ_d collect the sets of hybrid time instances at which u_c and u_d are nonzero, respectively. The set Γ_c results in a set of intervals of hybrid time (aptly referred to as shifted hybrid time) upon which the input is nonzero. Due to item 2), a δ-time detectable hybrid input $u_c(t, j)$ cannot be zero at hybrid time $(t, j) = (0, 0)$. Due to the discrete properties of the jumps, the set Γ_d contains only isolated points of the hybrid time domain of the input. Note that it is not necessary for the input to be continuous or even differentiable, but it is required that if either u_c or u_d become zero at some (t_1', j_1') then there exists another hybrid time (t_2', j_2') in the domain of u such that: $t_2' + j_2' \geq \delta + t_1' + j_1'$, and for all (t, j) such that $t_2' + j_2' \geq t + j \geq \delta + t_1' + j_1'$, the input satisfies $u(t, j) = 0$; see Figure 1 for an example of a δ-time detectable hybrid input with constant magnitude when the input is nonzero.



(a) A hybrid arc for an input u. The transparent red surfaces indicate the points in hybrid time when the input is nonzero.



(b) The domain of the input u. The black line shows the domain of a hybrid arc. The red lines indicate the shifted hybrid time domains when $u \neq 0$.

Fig. 1. A δ-time detectable hybrid input $u \in \{0, u^\star\}$, where $u^\star$ is some fixed positive value. Note that when the input becomes zero, then it remains zero for at least a δ amount of time.

[1] A solution pair (ϕ, u) to $\mathcal{H}$ is called *maximal* if there does not exist a pair (ϕ', u') such that $\operatorname{dom}(\phi, u)$ is a proper subset of $\operatorname{dom}(\phi', u')$. The pair (ϕ, u) is called *complete* if its domain $\operatorname{dom}(\phi, u)$ is unbounded. A solution is Zeno if it is complete and its domain is bounded in the t direction. Hybrid time domains are denoted by $\operatorname{dom}(\phi, u)$, and are subsets of $\mathbb{R}_{\geq 0} \times \mathbb{N}$ with the following structure: for each $(T, J) \in \operatorname{dom}\phi$, the set $\operatorname{dom}\phi \bigcap([0, T] \times \{0, 1, \ldots, J\})$ can be written as $\mathcal{E} = \bigcup_{j=0}^{J}(I_j \times \{j\})$, where $I_j := [t_j, t_{j+1}]$ for a time sequence $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_{J+1}$.

[2] Note that a shifted hybrid time domain may be open on the right rightmost subinterval and closed on the left leftmost subinterval, or vice versa; for example, $((t_{a,i}, j_{a,i}), (t_{b,i}, j_{b,i})]$ or $[(t_{a,i}, j_{a,i}), (t_{b,i}, j_{b,i}))$, respectively.

III. Problem Motivation and Formulation

A. Model of Cyber-physical System

In this paper, we consider the case of a cyber-physical system which has continuous dynamics and updates its state discretely at some unknown times given by the sequence of times $\{t_s\}_{s=1}^{\infty}$ such that, after the first event, the length of the interval of time between each subsequent event is bounded below by a constant, i.e., for some $T > 0$, the sequence of times $\{t_s\}_{s=1}^{\infty}$ is such that

$$t_{s+1} - t_s \geq T \tag{3}$$

for each $s \in \mathbb{Z}_{\geq 1}$. When $t \notin \{t_s\}_{s=1}^{\infty}$ the cyber-physical system operates under the continuous-time dynamics given by

$$\dot{x}(t) = A_c x(t), \qquad y(t) = H_c x(t) \tag{4}$$

where $x \in \mathbb{R}^n$ is the state, $y \in \mathbb{R}^p$ is the output, $A_c \in \mathbb{R}^{n\times n}$ is the physical system matrix, and $H_c \in \mathbb{R}^{n\times p}$ is the output matrix during flows. When $t \in \{t_s\}_{s=1}^{\infty}$, an event occurs and the states of the agents are subjected to an impulsive change according to

$$x(t^+) = A_d x(t), \qquad y(t) = H_d x(t) \tag{5}$$

where $x(t^+)$ indicates an impulsive change in x at time t and generates the output y, $A_d \in \mathbb{R}^{n\times n}$ and $H_d \in \mathbb{R}^{n\times p}$ are the cyber system and output matrices, respectively. This framework can be used to model many cyber-physical systems, such as those in [5], [13].

B. Model of Attackers

We are interested in the detection of attacks on the cyber-physical system in (4)-(5). In general, such attackers may have limited or partial information about the system and its dynamics. However, in this article, we consider a *worst case* attacker acting under the following assumptions.

Assumption 3.1: The attacker has full knowledge of the system matrices A_c and A_d, the output matrices H_c and H_d, and the communication event times $\{t_s\}_{s=1}^{\infty}$, which satisfy (3). The attacker is only allowed to generate a hybrid signal $(t, j) \mapsto (u_c(t, j), u_d(t, j))$ that is a δ-time detectable input.

Remark 3.2: The attack model adopted in this work is in line with the model of Byzantine [14] attacks, where the attacker is assumed to have complete knowledge of the system and infinite computational power to design its strategy. However, the attack model considered in this paper differs from most works on cyber-physical security, as we allow for attacks to occur as we allow for recurrent attacks that remain active over disjoint time intervals.

The signal $(t, j) \mapsto (u_c(t, j), u_d(t, j))$ generated by the attacker may affect both the continuous and discrete evolution of the state of (4)-(5). This is modeled by injecting $B_c u_c$ to the continuous dynamics in (4) and $B_d u_d$ to the discrete dynamics in (5). The system in (4)-(5) with the addition of the attacker becomes

$$\left.\begin{aligned} \dot{x} &= A_c x + B_c u_c \\ y &= H_c x \end{aligned}\right\} \quad \text{when } t \notin \{t_s\}_{s=1}^{\infty}, \qquad \left.\begin{aligned} x^+ &= A_d x + B_d u_d \\ y &= H_d x \end{aligned}\right\} \quad \text{when } t \in \{t_s\}_{s=1}^{\infty}. \tag{6}$$

C. Model of the Hybrid Attack Monitor

To detect attacks on the system in (6), we utilize a monitor with hybrid dynamics and local state $\zeta = (\hat{x}, \chi) \in \mathbb{R}^n \times \mathbb{R}^m$, where $\hat{x}$ is the state estimate of the cyber-physical system from the measured output y in (6), and χ contains auxiliary states necessary for the detection of the attacks. The monitor takes the output information and the knowledge of the discrete events, and generates a residual signal r to indicate when an attack has occurred. We say that when the residual is identically zero, then no attack is occurring; when its value is non-zero and an attack is initiated, then such an attack has been detected. We will exploit this property of the residual to report when attacks occur on the cyber-physical system.

The state of the monitor is allowed to continuously evolve according to $\dot{\zeta} = f_m(\zeta, y)$ whenever the state ζ belongs to the set C_m. Due to the hybrid nature of the monitor, it may be designed to have a self-induced jump, or to jump when a discrete event occurs in the cyber-physical system. Specifically, ζ is allowed to jump impulsively according to the difference inclusion $\zeta^+ \in g_m^1(\zeta, y) \cup g_m^2(\zeta, y)$, where $g_m^1 : \mathbb{R}^{n+m} \times \mathbb{R}^p \to \mathbb{R}^{n+m}$ and $g_m^2 : \mathbb{R}^{n+m} \times \mathbb{R}^p \to \mathbb{R}^{n+m}$ are to be designed. Namely, g_m^1 is active when events of (6) occur (namely, when $t \in \{t_s\}_{s=1}^{\infty}$), and g_m^2 is active when the internal events when the monitor occur. The closed-loop model resulting from connecting the hybrid monitor with the cyber-physical system is presented in Section IV-A.

D. Formulation of the Attack Detection Problem

Our objective is to design a class of hybrid monitors, as outlined in Section III-C to detect the class of attacks in Assumption 3.1. The following statement summarizes our problem.

Problem 1: Let A_c, B_c, H_c, A_d, B_d, H_d, $T > 0$, and attacks satisfying Assumption 3.1 be given; namely, for any δ-time detectable hybrid input $(t, j) \mapsto (u_c(t, j), u_d(t, j))$ satisfying Definition 2.1, where $\Gamma_c = \cup_{i=1}^{N_c} \mathcal{I}_i$, $N_c \in \mathbb{N} \cup \infty$ with shifted hybrid time domains are given by $\mathcal{I}_i := [(\underline{t}_i^c, \underline{j}_i^c), (\overline{t}_i^c, \overline{j}_i^c))$ for each $i \in \{1, 2, \ldots, N_c\}$, and $\Gamma_d := \{(t_i^d, j_i^d)\}_{i=1}^{N_d}$, $N_d \in \mathbb{N} \cup \infty$, satisfy $\min\{\underline{t}_1^c + \underline{j}_1^c, t_1^d + j_1^d\} > \delta$. The problem is to design the data of the hybrid attack monitor $(C_m, f_m, D_m, g_m^1 \cup g_m^2)$ such that, for any initial condition of the cyber-physical system in (6) and of the estimate of the state $\hat{x}$, the hybrid attack monitor *detects every unique attack*, namely, it determines the set of points

$$\{(\underline{t}_i^c, \underline{j}_i^c)\}_{i=1}^{N_c} \tag{7}$$

and

$$\{(t_i^d, j_i^d)\}_{i=1, i\notin\mathcal{J}}^{N_d}. \tag{8}$$

where $\mathcal{J} \subset \{1, 2, \ldots, N_d\}$ such that $\Gamma_c \cap \Gamma_d = \{(t_i^d, j_i^d)\}_{i\in\mathcal{J}}$.

IV. Closed-loop Hybrid Model for Attack Detection and Main Results

A. Hybrid Modeling

The three subsystems introduced in Section III, namely, the attacker, the cyber-physical system, and the monitor lead to a closed-loop hybrid system. Note that the monitor and the cyber-physical system have both continuous and discrete dynamics. We utilize the hybrid systems framework in [15], which is summarized in Section II, to model their interconnection. The resulting hybrid system, denoted as $\mathcal{H}$, has state $z = (x, \tau_p, \zeta) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0} \times \mathbb{R}^{n+m} =: \mathcal{X}$, where x is the state of the cyber-physical system in (6), τ_p is a decreasing timer which triggers events by reaching zero and reseting to a value in the interval $[T, \infty)$, and ζ is the state of the monitor, which contains the estimate $\hat{x}$ and the internal state χ. The dynamics of $\mathcal{H}$ are

$$\begin{aligned} \dot{z} = f(z) := \begin{bmatrix} A_c x + B_c u_c \\ -1 \\ f_m(\zeta, H_c x) \end{bmatrix} \\ z \in C := \{z \in \mathcal{X} : \zeta \in C_m\}, \\ z^+ \in G(z) := \{G_i(z) : z \in D_i, i \in \{1,2\}\} \\ z \in D := D_1 \cup D_2 \end{aligned} \tag{9}$$

where, for each $i \in \{1, 2\}$, the map G_i and the set D_i are given by

$$G_1(z) = \begin{bmatrix} A_d x + B_d u_d \\ [T, \infty) \\ g_m^1(\zeta, H_d x) \end{bmatrix}, \qquad G_2(z) = \begin{bmatrix} x \\ \tau_p \\ g_m^2(\zeta, H_c x) \end{bmatrix},$$
$$D_1 = \{z \in \mathcal{X} : \tau_p = 0\}, \quad D_2 = \{z \in \mathcal{X} : \zeta \in D_m\}.$$

The map G_1 captures the jumps in the monitor and G_2 captures the jumps in the cyber-physical system. Moreover, note that when $z \in D_1 \cap D_2$, the jump map G is set valued and leads to nonunique solutions.

To detect attacks, the output of the hybrid system $\mathcal{H}$ utilizes the residual signal between its estimated state and the output of the cyber-physical system. Namely, given a hybrid signal $(t, j) \mapsto z(t, j)$, we have that

$$r(z(t,j)) = \begin{cases} H_d x(t,j) - H_d \hat{x}(t,j) & \text{if } (t,j) \in \mathcal{K}, \\ H_c x(t,j) - H_c \hat{x}(t,j) & \text{otherwise} \end{cases} \tag{10}$$

where $\mathcal{K}$, which can be defined for any given solution, is the set of hybrid times where (6) jumps.

To detect attacks under Assumption 3.1, we propose the following hybrid monitor with estimates that converge in finite time when no attack is occurring.

Assumption 4.1: Given A_c, B_c, A_d, B_d, H_c, H_d, $T > 0$, and $\delta > 0$, there exist f_m, g_m^1, g_m^2, C_m, and D_m that satisfy the following:

- B1) For some $\gamma \in (0, \delta)$, every solution pair $(\phi, u) \in \mathcal{S}_\mathcal{H}$, $\phi = (\phi_x, \phi_\tau, \phi_\zeta)$, and each $(t, j) \in \operatorname{dom}(\phi, u)$ such that $u(t', j') = 0$ for all $(t', j') \in \operatorname{dom}(\phi, u)$ from $t + j \leq t' + j' \leq t + j + \gamma$, we have that there exists $(\tilde{t}, \tilde{j}) \in \operatorname{dom}(\phi, u)$ satisfying $t + j \leq \tilde{t} + \tilde{j} \leq t + j + \gamma$ and $\phi_x(t^*, j^*) = \phi_{\hat{x}}(t^*, j^*)$ for each $(t^*, j^*) \in \operatorname{dom}(\phi, u)$ such that $\tilde{t} + \tilde{j} \leq t^* + j^* \leq t + j + \gamma$.
- B2) For every solution pair $(\phi, u) \in \mathcal{S}_\mathcal{H}$, if there exists $(t, j) \in \operatorname{dom}(\phi, u)$ such that $|u(t, j)| > 0$, then $|r(\phi(t, j))| > \varepsilon$ for some $\varepsilon > 0$.
- B3) If $u \equiv 0$, the set $\{z \in \mathcal{X} : x = \hat{x}\}$ is forward invariant for $\mathcal{H}$. Namely, for every $(\phi, u) \in \mathcal{S}_\mathcal{H}$ such that $u(t, j) = 0$ for all $(t, j) \in \operatorname{dom}(\phi, u)$ and $\phi_x(0, 0) = \phi_{\hat{x}}(0, 0)$ then $\phi_x(t, j) = \phi_{\hat{x}}(t, j)$ for all $(t, j) \in \operatorname{dom}(\phi, u)$.

Assumption B1) requires that, if $u(t, j) = 0$ for windows of hybrid time of length γ, then the estimate generated by the hybrid monitor will converge to the state in finite time within that window. Sufficient conditions for finite time convergence for hybrid systems can be found in [16], including sufficient conditions given in terms of Lyapunov functions. Assumption B2) guarantees that when an attack is occurring the residual is larger than a positive constant ε for each solution to $\mathcal{H}$. Assumption B3) requires that when there is no attack and the estimate generated by the monitor is identically equal to the state, then it remains equal for all hybrid time.

B. Main Results

The following result establishes that, when Assumption 4.1 holds, then Problem 1 is solved.

Theorem 4.2: Given $T > 0$, $\delta > 0$ and the hybrid system $\mathcal{H}$ as in (9) *with attacks satisfying Assumption 3.1, if the hybrid attack monitor satisfies Assumption 4.1, then the monitor solves Problem 1.*

Remark 4.3: In the literature, a typical assumption for attack detection is the requirement that the initial conditions are known to the monitor, e.g., see [17]. If we consider that same assumption in our setting, we can relax the need for the initial attacks to occur after a δ amount of time.

Recall that the times at which u is nonzero are given by (7) and (12). We have the following result.

Corollary 4.4: Given $T > 0$, the hybrid system $\mathcal{H}$ as in (9) *from $\{z = (x, \tau_p, \zeta) \in \mathcal{X} : x = \hat{x}, \zeta = (\hat{x}, \chi)\}$ with attacks satisfying Assumption 3.1, the sets Γ_c and Γ_d satisfying $\min\{\underline{t}_1^c + \underline{j}_1^c, t_1^d + j_1^d\} > 0$, then the the hybrid attack monitor satisfying Assumption 4.1 can determine the set of points*

$$\{(\underline{t}_i^c, \underline{j}_i^c)\}_{i=1}^{N_c} \tag{11}$$

and

$$\{(t_i^d, j_i^d)\}_{i=1, i \notin \mathcal{J}}^{N_d} \tag{12}$$

where $\mathcal{J} \subset \{1, 2, \ldots, N_d\}$ such that $\Gamma_c \cap \Gamma_d = \{(t_i^d, j_i^d)\}_{i \in \mathcal{J}}$.

Remark 4.5: In light of Assumption 4.1 and the fact that there is no attack occurring when the residual is zero, we can construct an estimate of the duration of an attack. More specifically, given $\phi, u \in \mathcal{S}_\mathcal{H}$, we can generate the set of estimated attack times, given by $\Gamma_r := \{(t, j) \in \operatorname{dom} \phi : r(t, j) \neq 0\}$, which consists of shifted hybrid times domains such that $\Gamma \subset \Gamma_r$.

C. A Particular Construction of the Hybrid Attack Monitor

In [18] and [19], a finite-time convergent observer is proposed to estimate the state of systems with continuous-time dynamics given by $\dot{x} = A_c x$ with output $y = H_c x$. Therein, the authors consider an observer of the following form:

$$\dot{\hat{x}}_i(t) = A_c\hat{x}_i(t) - L_i(H_c\hat{x}_i(t) - y(t)) \quad \forall t \neq k\gamma, k \in \mathbb{N},$$
$$\hat{x}_i(t^+) = \tilde{K}_{1,i}(k)\hat{x}_1 - \tilde{K}_{2,i}(k)\hat{x}_2 \qquad \forall t = k\gamma, k \in \mathbb{N},$$

for each $i \in \{1,2\}$; $\hat{x}_1, \hat{x}_2 \in \mathbb{R}^n$; $d > 0$; $F_1 = A_c - L_1H_c$, $F_2 = A_c - L_2H_c$, and $L_1, L_2 \in \mathbb{R}^{n\times p}$; $\tilde{K}_{2,i}(1) = (I - \exp(F_2\gamma)\exp(-F_1\gamma))^{-1}$; and $\tilde{K}_{1,i}(1) = I - \tilde{K}_{2,i}(1)$; $\tilde{K}_{1,1}(k) = I$, $\tilde{K}_{2,1}(k) = 0$ for each $k \in \mathbb{N}\setminus\{1\}$; $\tilde{K}_{1,2}(k) = 0$, $\tilde{K}_{2,2}(k) = I$ for each $k \in \mathbb{N}\setminus\{1\}$; see [18], [19] for more details. Let the estimation error be given by $e_i = \hat{x}_i - x$. The parameter $\gamma > 0$ defines the time when $e_i^+(\gamma) = 0$ for each $i \in \{1,2\}$. Based on [18], finite time convergence occurs (at $t = \gamma$) when $\hat{x}_1(0) = \hat{x}_2(0)$ if $\tilde{K}_{1,i}(1)$, $\tilde{K}_{2,i}(1)$ are well defined, which is guaranteed when L_1, L_2, and γ are chosen to satisfy the following conditions:

Assumption 4.6: Given $T > 0$ and $\delta > 0$, the parameters $L_1, L_2 \in \mathbb{R}^{n\times p}$ and $\gamma \in (0, \min\{T, \delta\})$ satisfy

C1) $F_i = A_c - L_iH_c$ is Hurwitz for each $i \in \{1,2\}$;
C2) $I - \exp(F_2\gamma)\exp(-F_1\gamma)$ is invertible.

We adapt this observer scheme into the hybrid attack monitor with data $(C_m, f_m, D_m, g_m^1 \cup g_m^2)$ defined in the closed loop system $\mathcal{H}$ in (9) with internal states $\zeta = (\hat{x}, \eta, \tau_m)$, where $\hat{x}$ is the estimated state, η is an auxiliary state, and τ_m is a timer that triggers the jumps of the hybrid monitor. The set $C_m := \{\zeta \in \mathcal{X}_m : \tau_m \in [0,\gamma]\}$ and the map f_m is given by

$$f_m(\zeta, y_c) = \begin{bmatrix} F_1\hat{x} + L_1y_c \\ F_2\eta + L_2y_c \\ 1 \end{bmatrix}. \tag{13}$$

The jump map of the monitor due to resets of the timer τ_m, (i.e., when $\tau_m = \gamma$) is

$$g_m^2(\zeta) = \begin{bmatrix} K_1\hat{x} + K_2\eta \\ K_1\hat{x} + K_2\eta \\ 0 \end{bmatrix} \tag{14}$$

and due to a reset of the cyber-physical system is

$$g_m^1(\zeta, y_d) = \begin{bmatrix} A_d\hat{x} - E(H_d\hat{x} - y_d) \\ A_d\hat{x} - E(H_d\hat{x} - y_d) \\ 0 \end{bmatrix}. \tag{15}$$

Note that g_m^1 resets both $\hat{x}$ and η to the same point and reinitializes the timer state to zero when the timer of the cyber-physical system τ_p expires. For the hybrid attack monitor defined in (13) – (15), it can be shown that, when $u \equiv 0$, the residual of the monitor in (10) converges to zero in a tunable amount of time γ.

V. Detecting Attacks in Cloud-Connected Cooperative networks

In this section, we numerically illustrate our results using a model of a multi-agent system over a network to survey an urban environment. To coordinate the agents, we assume that nearby agents can establish ad-hoc communication links, and a subset of agents in the network can connect to the cloud and interact over a longer distance. It has been shown that cloud-based cooperation in autonomous networks not only improves the agents' communication range, but it also increases their computational capabilities and contextual awareness; see, e.g, [20], [21]. Let the agent state be given by $x = (x_c, x_d)$ where x_c contains the physical positions of the agents, while x_d represent the virtual information extracted from the cloud to guide the agents. When the physical and cloud cooperation algorithms are linear, the nominal network dynamics read as

$$\begin{bmatrix} \dot{x}_c \\ \dot{x}_d \end{bmatrix} = \begin{bmatrix} A_{cc} & A_{cd} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ x_d \end{bmatrix} \tag{16}$$

for all times $t \notin \{t_s\}_{s=1}^{\infty}$, and

$$\begin{bmatrix} x_c^+ \\ x_d^+ \end{bmatrix} = \begin{bmatrix} I & 0 \\ A_{dc} & A_{dd} \end{bmatrix} \begin{bmatrix} x_c \\ x_d \end{bmatrix}, \tag{17}$$

for all times $t \in \{t_s\}_{s=1}^{\infty}$ satisfying (3).

Consider the network with five agents where the state vector $x = (x_c, x_d)$, where $x_c = (x_1, x_2, x_3, x_4, x_5)$ contains the positions of each agent, and $x_d = (\vartheta_1, \vartheta_2, \vartheta_3)$ contains three variables representing virtual limits (white circles in the figure), which are updated sporadically by the cloud with a lower bound $T = 0.3$. In this example, the agents are split into two groups (x_1, x_2, x_3) and (x_4, x_5), where the agents in each group are coupled together and continuously communicate their positions to their neighbor. The agents must rely on the cloud for instructions on how to orient themselves in space, namely, the cloud provides the values of the states ϑ_i, where ϑ_i indicates the boundaries on a line that the agents are to cover. More specifically, ϑ_1 is the lower boundary, ϑ_3 is the upper boundary and ϑ_3 is the boundary separating the two groups. The goal of the agents is to cover a line based on the values of ϑ_i. Therefore, we define the system matrices in (16) and (17) as follows:

$$A_{cc} = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}, A_{cd} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

$$A_{dc} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad A_{dd} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{6} & \frac{2}{6} & \frac{1}{6} \\ 0 & 0 & 1 \end{bmatrix}.$$

In this example, we consider two types of attacks; namely, a malware attack (where a malicious agent injects an authorized service to the cloud to control the information received by the cloud users) on the second virtual state ϑ_2, and a man in the middle attack (where the malicious entity interpose itself between cloud and users, and arbitrarily compromises data exchanged in both directions) against the first agent x_1, represented by u_d and u_c, respectively. More information on specific forms of attacks can be found in [22], [23]. This attack structure leads us to define the input matrices in (9) as $B_c = e_1$ and $B_d = e_7$, where e_i is the i-th canonical vector. For this example, we consider the case when $u_c(t,j) = 3$ for all $(t,j) \in \operatorname{dom} u$ such that $t \in [1, 1.5]$ and u_c is zero otherwise. Moreover, $u_d(t,j) = 5$

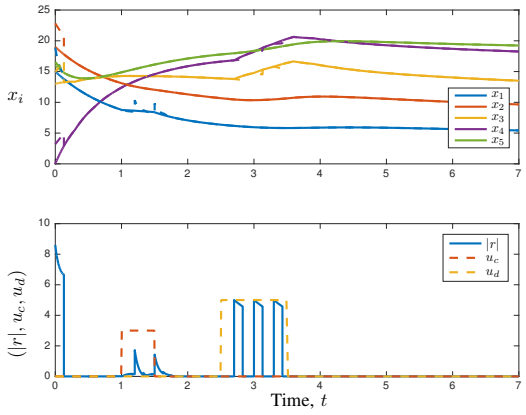


Fig. 2. Numerical solution of the cloud-cooperation model. The upper plot has both the positions of the agents and the estimates (dashed line of the same color). The bottom plot shows the residual and the attacks. Note that, after the attacks have ceased, the residual decreases to zero in finite-time.

for each $(t, j), (t, j+1) \in \operatorname{dom} u$ such that $t \in [2.5, 3.5]$ and $\tau_p(t, j) = 0$ and $u(t, j) = 0$ elsewhere. Namely, on this interval of flow time, u_d is nonzero when communication between the agents and the cloud occurs. With these attacks above, the δ-time detectable hybrid input has $\delta = 1$. Then, let $\gamma = 0.13$, the matrices L_1 and L_2 satisfying Assumption 4.6 given by

$$L_1 = \begin{bmatrix} 1.1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 3.3 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 14 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -5.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2.7 \end{bmatrix}$$

$$L_2 = \begin{bmatrix} 4.2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1.3 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 6.2 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 6 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 5.3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -8.3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 11 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3.7 \end{bmatrix}.$$

Figure 2 shows a numerical solution from $x_c(0,0) = (15, 19, 13, 0, 16)$, $x_d(0,0) = (1, 10, 20)$, $\hat{x}(0,0) = (15.2, 19.4, 16.3, 2.8, 17.3, 3.8, 10.1, 21.1)$, where $\hat{x}(0,0) = \eta(0,0)$, necessarily[3]. In this figure, it can be seen that the residual is pushed away from zero whenever an attack occurs. This indicates that the hybrid attack monitor designed in Section IV-C is capable of detecting such recurring attacks.

VI. Conclusion

In this work, we studied a security problem for a class of cyber-physical systems with linear continuous and discrete dynamics. We considered an attacker that injects potentially recurring signals into the dynamics of the cyber-physical system. Using hybrid system modeling tools, we designed a general hybrid attack monitor and, under reasonable assumptions, showed that is able to detect the initial time of the recurring attacks. We illustrated the hybrid attack monitor using a specific finite-time convergent observer model and validated our results in a cloud-connected network of autonomous vehicles scenario.

[3]Code at `github.com/HybridSystemsLab/CPSRecurrentAttackMonitor`

References

[1] D. D. Kushner. The real story of stuxnet. *IEEE Spectrum*, 3(50):48–53, 2013.

[2] S. Karnouskos. Stuxnet worm impact on industrial cyber-physical system security. In *Annual Conference on IEEE Industrial Electronics Society*, pages 4490–4494. 2011.

[3] N. Shachtman. Computer virus hits us drone fleet. *cnn.com*, 10, 2011.

[4] Y. Liu, M. K. Reiter, and P. Ning. False data injection attacks against state estimation in electric power grids. In *ACM: Conf. on Computer and Comm. Sec.*, pages 21–32, Chicago, IL, USA, Nov. 2009.

[5] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S.S. Sastry. Cyber security analysis of state estimators in electric power systems. In *IEEE Conf. on Decision and Control*, pages 5991–5998, Atlanta, GA, USA, December 2010.

[6] S. Bhattacharya and T. Başar. Differential game-theoretic approach to a spatial jamming problem. In *Advances in Dynamic Games*, pages 245–268. Springer, 2013.

[7] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Başar. Dependable demand response management in the smart grid: A stackelberg game approach. *IEEE Trans. Smart Grid*, 4(1):120–132, 2013.

[8] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, 2014.

[9] C. Bai, F. Pasqualetti, and V. Gupta. Security in stochastic control systems: Fundamental limitations and performance bounds. In *American Control Conference*, pages 195–200, Chicago, Il, July 2015.

[10] F. Pasqualetti, A. Bicchi, and F. Bullo. Consensus computation in unreliable networks: A system theoretic approach. *IEEE Transactions on Automatic Control*, 56(12), 2011.

[11] F. Pasqualetti, F. Dörfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[12] A. A. Cárdenas, S. Amin, and S. S. Sastry. Research challenges for the security of control systems. In *Proceedings of the 3rd Conference on Hot Topics in Security*, pages 6:1–6:6, Berkeley, CA, USA, 2008.

[13] S. Phillips, Y. Li, and R. G. Sanfelice. On distributed intermittent consensus for first-order systems with robustness. In *Proc. of 10th IFAC Symposium on Nonlinear Control Systems*, pages 146–151, 2016.

[14] L. Lamport, R. Shostak, and M. Pease. The Byzantine generals problem. *ACM Trans. on Prog. Lang. and Sys.*, 4(3):382–401, 1982.

[15] R. Goebel, R. G. Sanfelice, and A. R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, New Jersey, 2012.

[16] Y. Li and R. G. Sanfelice. Results on finite time stability for a class of hybrid systems. In *Proc. of the American Control Conference*, pages 4263–4268, 2016.

[17] F. Pasqualetti, F. Dörfler, and F. Bullo. Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems. *IEEE Cont. Sys. Mag.*, 35(1):110–127, 2015.

[18] T. Raff and F. Allgower. An impulsive observer that estimates the exact state of a linear continuous-time system in predetermined finite time. In *2007 Medi. Conf. on Control Auto.*, pages 1–3, June 2007.

[19] R. Engel and G. Kreisselmeier. A continuous-time observer which converges in finite time. *IEEE Transactions on Automatic Control*, 47(7):1202–1204, Jul 2002.

[20] M. Gharibi, R. Boutaba, and S. L. Waslander. Internet of drones. *IEEE Access*, 4:1148–1162, 2016.

[21] S. Srinivasan, H. Latchman, J. Shea, T. Wong, and J. McNair. Airborne traffic surveillance systems: video surveillance of highway traffic. In *Workshop on video surv. & sensor networks*, pages 131–135, 2004.

[22] N. Gruschka. M. Jensen. Attack surfaces: A taxonomy for attacks on cloud services. In *3rd International Conf. on Cloud Computing*, 2010.

[23] M. Shrivastava A. Singh. Overview of attacks on cloud computing. *International Journal of Engineering and Innovative Tech.*, 1, 2012.