2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Periodic Coordinated Attacks Against Cyber-Physical Systems: Detectability and Performance Bounds

Rajasekhar Anguluri, Vijay Gupta and Fabio Pasqualetti

Abstract— **Cyber-physical systems are vulnerable to attacks across different and, possibly, independent attack channels. In this paper we consider the situation where an attacker orchestrates periodic integrity and denial of service attacks against sensors and actuators. By switching between different attack modalities, the attacker is able to lower the probability of being detected while compromising the system to a greater extent. For single-input and single-output systems driven by noise, we frame the detection of periodic coordinated attacks as an hypothesis testing problem, and we characterize the detection time as a function of the system dynamics, noise statistics, and attack parameters. Our bounds allow us to design optimal attacks, and to highlight fundamental tradeoffs between the dynamics of the system and its resilience to attacks.**

I. Introduction

Cyber-physical systems require advanced protection mechanisms to secure all implementation layers and communication interfaces. In contrast with legacy control systems, typically isolated from the outer world, the cyber and physical components of modern cyber-physical systems are interconnected via local data networks, and connected to the outer world via the Internet. This poses significant risks to personal privacy, economic security, and critical infrastructure.

In cyber-physical systems, security implies not only data protection and authorized operation, but also satisfactory performance of the control system in the face of failure and sabotage. While different methods ensuring cyber-physical security have been proposed (see related work), existing studies mainly consider single attack modalities, thereby underestimating the possibility of coordinated independent attacks, thereby neglecting stability and other issues of the individual components of the system. In this paper, instead, we investigate the design and resilience against coordinated, concurrent and independent attack modes, namely integrity and denial of service attacks against, respectively, input and output transmission channels.

Related work 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. From early works in the 1980s [1], [2], computer scientists and information theorists have developed fundamental intrusion detection and security mechanisms for purely cyber systems [3], [4]. In the same years, control theorists have addressed fault detection and isolation problems for purely dynamic control systems [5], [6]. Motivated by the advent of cyber-physical systems, cyber-physical security has emerged as a separate and rather interdisciplinary research field [7], [8]. While early works focus on static representations [9], [10], game-theoretic [11], information theoretic [12], [13], and control-theoretic methods [14], [15], [16], [17] have been proposed for dynamic models and attacks. To the best of our knowledge, these work study detection, identification, and resilience for single attack modalities, such as integrity and denial of service [18], [19]. Yet, intuitively, an attacker capable of switching between independent attack modalities may delay detection and may affect the system performance to a greater extent. In this work we begin the investigation of this case by studying design and detection of combined integrity and denial of service attacks against stochastic control systems.

Contribution The contribution of this paper is twofold. First, we motivate and introduce the study of concurrent and independent attacks against stochastic cyber-physical systems, where an attacker periodically alternates between integrity and denial of service attacks. As shown in our numerical studies, combined attacks achieve a greater effect while undetected than single mode attacks. Second, for each attack modality, we characterize the detection performance of a defender using a Sequential Probability Ratio Test procedure as a function of the system dynamics and attack parameters. Additionally, we provide a quantitative design procedure for an attacker to remain undetected for a prespecified duration. We focus on periodic attacks against single-input single-output stochastic control systems, and provide results both theoretical and numerical for single and multiple period attacks.

Paper organization The rest of the paper is organized as follows. Section II contains our setup and some preliminary notions. Sections III and IV contain our detectability analysis for single and multi-modality attack respectively. Section V contains our numerical studies of combined, periodic and independent attacks. Finally, Section VI concludes the paper.

II. Problem setup and preliminary results

In this section we detail the considered system dynamics, attack model, and statistical detection mechanism. Our setup is illustrated in Fig. 1.

A. System model

We consider the single-input single-output stochastic linear time-invariant system governed by

$$x_{k+1} = ax_k + u_k + w_k, \qquad y_k = x_k + v_k, \qquad (1)$$

where $a \in \mathbb{R}$, with $|a| < 1$, and the random variables w_k, v_k are process and measurement noise realizations, respectively.

This material is based upon work supported in part by ONR award #N00014-14-1-0816. Rajasekhar Anguluri and Fabio Pasqualetti are with the Department of Mechanical Engineering, University of California, Riverside, CA 92521, rajasekhar.anguluri@ieee.org, fabiopas@engr.ucr.edu. Vijay Gupta is with the Department of Electrical Engineering, University of Notre Dame, IN 46556, vgupta2@nd.edu.

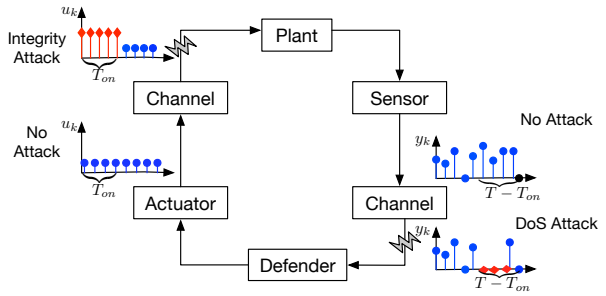

Fig. 1. Illustration of the coordinated attacks acting periodically on independent channels for a fixed amount of time. Blue (circled arrows) region indicates the nominal behavior of the inputs/measurements and the red region(pointed arrows) reflects influence of attacker on the same.

For all $k > 0$, we assume these random variables to be independent and identically distributed (i.i.d) Gaussian processes with $w_k \sim \mathcal{N}(0, \sigma_w^2)$, $v_k \sim \mathcal{N}(0, \sigma_v^2)$. A steady state Kalman filter is used to compute the Minimum-Mean-Squared-Error (MMSE) estimate $\hat{x}_{k+1}$ of x_{k+1} from the measurements $y^k = [y_0, y_1, \ldots, y_k]$. The estimates read as

$$\hat{x}_{k+1} = a\hat{x}_k + u_k + K(y_k - \hat{x}_k) \tag{2}$$

where K is the steady state Kalman gain, and $\hat{x}_0 = \mathbb{E}[x_0] = 0$. As the system is observable, the Kalman filter converges in the mean square sense, that is, the error covariance satisfies $\lim_{k\to\infty} P_k = P$, where $P_k \triangleq \mathbb{E}[(\hat{x}_k - x_k)^2]$ and P is the solution to an algebraic Riccati equation [20]. The innovation is computed as $z_k \triangleq y_k - \hat{x}_k$. Due to the assumption of steady state Kalman filtering, the innovation sequence is an i.i.d Gaussian process with $z_k \sim \mathcal{N}(0, P + \sigma_v^2)$. To simplify the analysis and without affecting generality, we will consider the nominal input to be zero at all times. Additionally, we will also assume that the system operates at steady state.

We assume the wireless communication channel between actuator/sensor and the plant to be unencrypted, so that an attacker can, for instance, arbitrarily replace the content of the actuation signal (*integrity attack*). Instead, we assume the communication between the sensor and the defender to be encrypted, and that the attacker can interfere with this communication by jamming the communication channel (*denial of service attack (DoS)*), without altering the content of the output signal. In the remainder of this section we introduce our model of attacker and detection mechanism.

B. Attack model

We assume that an attacker can independently and concurrently cast integrity attacks on the input packets by accessing the actuator-plant channel, and denial of service attacks by jamming the sensor-defender channel. We focus on periodic attacks, which we model as follows. Let the total attack time be mT, where m is the total number of periods and T is the duration of each period. We assume that the attacker knows the system parameters, let $T_{\text{on}} < T$, and model integrity and denial of service attacks in each period as follows:

$$u_k = \begin{cases} u_k^a, & \text{if } 0 \le k < T_{\text{on}}, \\ 0, & \text{otherwise}, \end{cases}$$

$$y_k = \begin{cases} \tilde{y}_k, & \text{if } T_{\text{on}} \le k < T \\ y_k, & \text{otherwise.} \end{cases}$$

Thus, for $0 \le k \le T_{\text{on}}$, the attacker performs an integrity attack by injecting an arbitrary control value instead of the nominal control input (0 in this case). Instead, for $T_{\text{on}} \le k < T$, the attacker performs a denial of service attack, so that the defender receives some predetermined value $\tilde{y}_k$ instead of the measurement y_k. We restrict our attention to constant integrity attacks, where $u_k^a = u$ ($\in \mathbb{R}$) for all times k.

C. Defender mechanism

To check if the system is operating normally, the defender tests for two anomalies in the system, that is, whether there is any disruption in the input channel or measurement channel. As the defender has no direct access to monitor the input channel, the innovation sequence is used instead to check for integrity attacks. Instead, the measurements are used to reveal the presence of denial of service attacks. The attack detection problem can be cast as a hypothesis testing problem, where the four hypotheses are:

$$\begin{cases} \mathcal{H}_{1,0}: & \text{no integrity attack;} \\ \mathcal{H}_{1,1}: & \text{integrity attack in progress;} \\ \mathcal{H}_{2,0}: & \text{no DoS attack;} \\ \mathcal{H}_{2,1}: & \text{DoS attack in progress.} \end{cases} \tag{3}$$

Notice that, while $\mathcal{H}_{1,0}$ and $\mathcal{H}_{1,1}$ (resp. $\mathcal{H}_{2,0}$ and $\mathcal{H}_{2,1}$) are mutually exclusive, $\mathcal{H}_{1,j}$ and $\mathcal{H}_{2,j}$ can be concurrently true, for all $j \in \{0, 1\}$. The rationale behind the hypothesis testing framework is as follows. When $\mathcal{H}_{1,0}$ is true, the system is operating normally and the innovation sequence (z_k) satisfies $\mathbb{E}_{\mathcal{H}_{1,0}}[z_k] = 0$. When an integrity attack is in progress, then the Kalman filters operates with the incorrect system input, so that the innovation sequence is expected to violate its nominal statistics leading to the detection of the attack. Similarly, when $\mathcal{H}_{2,0}$ is true, the defender expects to receive measurements from the sensor with erasure probability p_e. On the other hand, when $\mathcal{H}_{2,1}$ is true, the erasure probability should differ leading to the detection of the DoS attack. Although the mechanism leading to detection of the two attacks is slightly different, we will employ the same statistical detection tool, which is next described.

D. Sequential probability ratio test (SPRT)

We assume that the defender employs SPRT to solve the hypothesis testing problem (3). Our choice is motivated by the fact that innovations and measurements are iteratively received by the defender. Moroever, this approach is known to be optimal to minimize the average sample size required to decide on a hypothesis. We define Λ_k[1] to be the *log-likelihood* ratio (LLR) of some random sequence ω_k that are governed by either of the hypothesis $\mathcal{H}_{1,0}$ and $\mathcal{H}_{1,1}$.

$$\Lambda_k = \ln \frac{f(\omega_1, \omega_2, \ldots, \omega_k | \mathcal{H}_{1,1})}{f(\omega_1, \omega_2, \ldots, \omega_k | \mathcal{H}_{1,0})} \tag{4}$$

where $f(\cdot|\mathcal{H}_{1,0})$ and $f(\cdot|\mathcal{H}_{1,1})$ are known probability density functions. For the SPRT to decide among the hypothesis, two

[1]The SPRT setup for $\mathcal{H}_{2,0}$ and $\mathcal{H}_{2,1}$ is identical, and it is omitted here.

thresholds A and B are defined, with $0 < B < A < \infty$. In particular, if $\Lambda_k \leq B$, then $\mathcal{H}_{1,0}$ is accepted, whereas $\mathcal{H}_{1,1}$ is accepted if $\Lambda_k \geq A$. No decision is taken if $B < \Lambda_k < A$, and the test iterates to include further evidence. The following thresholds are typically used in SPRT:

$$A = \ln\frac{\beta}{1-\alpha}, \text{ and } B = \ln\frac{1-\beta}{\alpha},$$

where α and β are type (i) and type (ii) probability errors (i.e., the probabilities to accept a given hypothesis wrongly when it is true), respectively. If the random variables ω_k are independent, it can be shown that $\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\gamma]$ can take the boundary values A and B with probabilities β and $1-\beta$, respectively [21], and it is given by

$$\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\gamma] = (1-\beta)\ln\left(\frac{1-\beta}{\alpha}\right) - \beta\ln\left(\frac{1-\alpha}{\beta}\right), \quad (5)$$

where γ is the time when SPRT decides for Hypothesis $\mathcal{H}_{1,1}$. Similar expression for $\mathbb{E}_{\mathcal{H}_{1,0}}[\Lambda_\gamma]$ can be found in [21]. It should be noticed that from now onwards emphasis will be mostly on $\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\gamma]$ as we are interested in charactering the detection time of the attacker.

III. Characterization of Single Period Attacks

In this section we now quantify the detection time as a function of the system dynamics, noise statistics, and attack parameters, for periodic integrity and DoS attacks. Our approach uses information-theoretic notions, which we now introduce. We refer the interested reader to [22].

Definition 1: ***(Kullback-Leibler divergence)*** Let ω_k be a random variable with probability density functions either $f(\omega_k|\mathcal{H}_{1,0})$ or $f(\omega_k|\mathcal{H}_{1,1})$. Then, the Kullback-Leibler (KL) divergence between $f(\omega_k|\mathcal{H}_{1,1})$ and $f(\omega_k|\mathcal{H}_{1,0})$ is

$$\mathbb{E}_{\mathcal{H}_{1,1}}\left[\ln\frac{f(\omega_k|\mathcal{H}_{1,1})}{f(\omega_k|\mathcal{H}_{1,0})}\right] = \int_{\mathbb{R}} f(\omega_k|\mathcal{H}_{1,1})\ln\frac{f(\omega_k|\mathcal{H}_{1,1})}{f(\omega_k|\mathcal{H}_{1,0})}d\omega_k. \quad (6)$$

With slight abuse of terminology we shall use the notation $D_{\omega_{[k]}}$ to refer (6). The KL divergence is a non-negative measure that measures the distance between two probability density functions.

Lemma* 3.1:** ***(KL divergence of innovations) The KL divergence of innovations z_k at particular instant k is directly proportional to square of the expected value of z_k, and inversely related to variance of the z_k at that instant.

$$D_{z_{[k]}} = \frac{\mathbb{E}_{\mathcal{H}_{1,1}}[z_k]^2}{2\sigma^2}.$$

Proof: As z_k follows normal distribution, it's probability density function can be computed for both the hypothesis with parameters $\mathbb{E}_{\mathcal{H}_{1,j}}[z_k]$ ($j \in \{0,1\}$) and $\sigma^2 = P + \sigma_v^2$. Then divergence of innovations at the k-th instant can be obtained using equations (6) as

$$\begin{aligned} D_{z_{[k]}} &= \mathbb{E}_{\mathcal{H}_{1,1}}\left[\ln\left(\frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(z_k-\mathbb{E}_{\mathcal{H}_{1,1}}[z_k])^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(z_k-\mathbb{E}_{\mathcal{H}_{1,0}}[z_k])^2}{2\sigma^2}\right)}\right)\right] \\ &\overset{(a)}{=} \mathbb{E}_{\mathcal{H}_{1,1}}\left[\ln\left(\exp\frac{-\mathbb{E}_{\mathcal{H}_{1,1}}[z_k]^2 + 2z_k\mathbb{E}_{\mathcal{H}_{1,1}}[z_k]}{2\sigma^2}\right)\right] \\ &\overset{(b)}{=} \frac{\mathbb{E}_{\mathcal{H}_{1,1}}[z_k]^2}{2\sigma^2} \end{aligned} \quad (7)$$

(a) follows from the fact that $\mathbb{E}_{\mathcal{H}_{1,0}}[z_k] = 0$ and (b) from the linear operation of expectation operator. ■

It can be inferred from Lemma 3.1 that if the expected value of innovations is greater than zero at a particular instant, the divergence between probability density functions governed by the corresponding hypothesis increases (and vice versa), and the innovation random variable is governed by the distribution of hypothesis $\mathcal{H}_{1,1}$. This key lemma also plays a crucial role in the upcoming theorems.

A. Integrity attack with single period

In this section we study detectability of integrity attacks with a single execution period. Notice that, when an integrity attack is in progress, the system is driven by the attack inputs while the Kalman filter uses the zero input. Consequently, the expected value (with respect to $\mathcal{H}_{1,1}$) of innovations can be characterized as

$$\mathbb{E}_{\mathcal{H}_{1,1}}[z_k] = \begin{cases} u\left[\frac{1-\Gamma^k}{1-\Gamma}\right], & k \in \{1,\ldots,T_{\text{on}}\}, \\ \Gamma^{(k-T_{\text{on}})}\mathbb{E}_{\mathcal{H}_{1,1}}\left[z_{T_{\text{on}}}\right], & k \geq T_{\text{on}}+1, \end{cases} \quad (8)$$

where $\Gamma = a - K$ (recall that a and K denote system dynamics and Kalman gain, respectively), u is the integrity attack input acting until T_{on}. It should be observed that, for $k \geq T_{\text{on}} + 1$, the expected value of the innovations are multiples of the expected value of innovations at time T_{on}, which is due to the fact that the attack input becomes zero at time T_{on}. From Lemma 3.1 and (8) we obtain a recursive formula to compute the divergence of innovations:

$$D_{z_{[k]}} = \begin{cases} \frac{u^2\left[\frac{1-\Gamma^k}{1-\Gamma}\right]^2}{2\sigma^2}, & \text{if } k \in \{1,\ldots,T_{\text{on}}\}, \\ \Gamma^{2(k-T_{\text{on}})}D_{z_{[T_{\text{on}}]}}, & \text{if } k \geq T_{\text{on}}+1. \end{cases} \quad (9)$$

Further, by exploiting the fact that the defender tests the expected value of innovations assuming that they are uncorrelated, the LLR of the innovations defined in (4) simplifies to

$$\Lambda_k = \ln\prod_{i=1}^{k}\frac{f(z_i|\mathcal{H}_{1,1})}{f(z_i|\mathcal{H}_{1,0})} = \sum_{i=1}^{k}\ln\frac{f(z_i|\mathcal{H}_{1,1})}{f(z_i|\mathcal{H}_{1,0})}. \quad (10)$$

Finally, from the definition of KL divergence, the expected value of Λ_k is given by

$$\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_k] = \sum_{i=1}^{k} D_{z_{[k]}}. \quad (11)$$

We are now ready to characterize the relation between expected value of detection time as a function of system

dynamics, attack inputs (u and T_{on}) and SPRT parameters, which is formalized in the following theorem.

Theorem* 3.2:** ***(Expected detection time for integrity attack with single period) Let T_{on} be the total time of the integrity attack and τ be the time at which SPRT triggers an alarm indicating the attack. Then, the expected detection time of the attack is characterized by the following expression.

$$\mathbb{E}_{\mathcal{H}_{1,1}}[\tau] = T_{\text{on}} - 1 + \frac{\ln\left(g - \frac{\kappa \mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\gamma]}{D_{z_{[T_{\text{on}}]}}}\right)}{2\ln(\Gamma)}, \tag{12}$$

where $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau]$ is the expected value of τ and g is a well defined function of system dynamics Γ, and the attack time T_{on}, which is given by,

$$g = \frac{\kappa T_{\text{on}}}{(1-\Gamma^{T_{\text{on}}})^2} - \frac{2\Gamma}{(1-\Gamma^{T_{\text{on}}})}.$$

Proof: Let $T_{on} < \tau$ be the time for which attacker operates on the input channel, from (11) we observe that

$$\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\tau] = \sum_{k=1}^{T_{\text{on}}} D_{z_{[k]}} + \sum_{k=T_{\text{on}}+1}^{\tau} D_{z_{[k]}}. \tag{13}$$

Using Lemma 3.1 and the characterization of expected value of innovations defined in (8) or (9) we have,

$$\sum_{k=1}^{T_{\text{on}}} D_{z_{[k]}} = \frac{u^2/2\sigma^2}{(1-\Gamma)^2}\left[T_{\text{on}} - \frac{2\Gamma(1-\Gamma^{T_{\text{on}}})}{(1-\Gamma)} + \frac{\Gamma^2(1-\Gamma^{2T_{\text{on}}})}{(1-\Gamma^2)}\right]$$

$$\sum_{k=T_{\text{on}}+1}^{\tau} D_{z_{[k]}} = \frac{u^2}{2\sigma^2}\left[\frac{\Gamma^2\left(1-\Gamma^{T_{\text{on}}}\right)^2}{(1-\Gamma^2)(1-\Gamma)^2}\right]\left(1-\Gamma^{2(\tau-T_{\text{on}})}\right). \tag{14}$$

By substituting (14) in Eq (13), we have,

$$\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\tau] = \frac{D_{z_{[T_{\text{on}}]}}}{\kappa}\left[g - \Gamma^{2(\tau-T_{\text{on}}+1)}\right]. \tag{15}$$

where g is defined in the Theorem 3.2 and $\kappa = 1-\Gamma^2$. From (5), for the SPRT to decide in expectation $\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\tau] = \mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\gamma]$ (see Section II-D for explicit characterization). Substituting Eq (15) in the above equality and there by rearranging terms (assuming $D_{z_{[T_{\text{on}}]}} \neq 0$)and taking expectations we have the desired expression for $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau]$.

Moreover if $\mathcal{T}$ is the detection time horizion then,

(i) if $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau] < \mathcal{T}$, the attacker gets detected.
(ii) if $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau] \geq \mathcal{T}$, the attacker remains undetected.

■

Theorem 3.2 highlights several tradeoffs between the system and attack parameters and the detection time.

Effect of attack magnitude on the detection time: as one may expect, reducing the attack magnitude results in a longer detection time. In fact, when the integrity attack time T_{on}, the dynamics $\Gamma = a - K$ and the noise statistics are fixed, then from (9), $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau]$ becomes function of the attack input. By observing that $\ln(\Gamma) < 0$, if the attack magnitude is reduced we see that $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau]$ increases because the quantity $\left(g - \frac{\kappa \mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\gamma]}{D_{z_{[T_{\text{on}}]}}}\right)$ inside the logarithm is well defined.

Effect of the system dynamics on the detection time: For fixed T_{on}, attack magnitude u and noise statistics, it is easy to see that $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau]$ is inversely related to $\ln(\Gamma)$. Hence, if $a << 1$ then, $\ln(\Gamma)$ ($\Gamma < 0$) decreases and the attack is detected quickly. This is because the system has faster dynamics and any change in the inputs are reflected in the innovations instantaneously.

Effect of the measurement noise on the detection time: For fixed T_{on}, dynamics $\Gamma = a - K$, and attack u, $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau]$ is a function of $\sigma^2 = P + \sigma_v^2$ (since, $D_{z_{[T_{\text{on}}]}}$ is inversely related to σ^2; see 9). As P is constant, if the noise level i.e., σ_v^2 increases, the logarithm expression in (12) contributes positively to the $\mathbb{E}_{\mathcal{H}_{1,1}}$ and the attack is detected quickly.

B. DoS attack with single period

As discussed earlier, the attacker implements a DoS attack when the integrity attack is not in progress, that is, from $T_{\text{on}}+1$ to T. Recall that, when a DoS attack is in progress, the defender does not receive any measurement, and it uses a predetermined value to update its current LLR statistic for integrity attacks. We assume this value to be 0, although other choices could be of interest.

Let θ_k be a random variable, such that $\theta_k = 0$ with probability p, and $\theta_k = 1$ with probability $1-p$. Let $\mathcal{H}_{2,0}$ be the hypothesis where the defender does not receive packets with probability $p = p_e$ (nominal erasure probability of the channel), and $\mathcal{H}_{2,1}$ the hypothesis with $p = \tilde{p}_e$, for some $\tilde{p}_e$ unknown to the defender. The probability density function $h(\theta_k)$ is governed in either hypothesis by

$$h(\theta_k = j|\mathcal{H}_i) = p^{1-j}(1-p)^j,$$

where $i, j \in \{0, 1\}$.

As $\tilde{p}_e$ is not predetermined, to test the hypothesis using SPRT, the defender fixes an arbitrary value (which is unknown to the attacker). Let the attacker casts DoS attack in a such way that the porbability $\tilde{p}_e$ is less than the fixed value of defender. Then the LLR of DoS attack will always be negative and hence, undetected (see Section II-D) by the SPRT. Instead, if $\tilde{p}_e$ is greater than the fixed value the attack gets detected. The following lemma provides an estimate for the attacker to choose p_e value so that the DoS attack remains undetected.

Lemma* 3.3:** ***(Maximum erasure probability of channel under attack) Let $\tilde{p}_{max}$ be the largest erasure probability that can be chosen by the attacker while remaining undetected. Then, $\tilde{p}_{max}$ satisfies

$$\left(\tilde{p}_{max}\ln\frac{\tilde{p}_{max}}{p_e} + (1-\tilde{p}_{max})\ln\frac{1-\tilde{p}_{max}}{1-p_e}\right) = \frac{\mathbb{E}[\Lambda_\gamma]}{\mathcal{T}},$$

where $\mathbb{E}[\Lambda_\gamma]$ is defined in (5).

Proof: By observing that packet losses follows an i.i.d distribution, the proof of lemma directly follow from expected number samples of SPRT in the case of binomial distribution [21]. Hence, the details are omitted. ■

Lemma 3.3 says that to remain undetected in $\mathcal{T}$ time steps, attacker can jam arbitrary number of packets , provided that the $\tilde{p}_e < \tilde{p}_{max}$. Similar results for the expected detection time (as in the case of integrity attacks) of DoS attacks can derived, if the attacker has access to the erasure probability used by the defender to perform the SPRT.

IV. Extension to multiple period attacks

In this section we consider the scenario where the attacker performs integrity and DoS attacks over multiple periods with a constant duty cycle $Dc = \frac{T_{\text{on}}}{T}$. Specifically, the attacker uses integrity attacks during the T_{on} component of each period, and DoS during the $T - T_{\text{on}}$ component of each period. Let m be the number of periods. In what follows, we assume that $Dc = 0.5$, although the analysis can be done also for different values of the duty cycle.

From Lemma 3.1 and equation (8) we obtain the following expressions for the KL divergence of the innovations at the instants T_{on} and T of the j-th period ($j > 1$).

$$D_{z_{[jT_{\text{on}}]}} = D_{z_{[T_{\text{on}}]}} \left(\frac{1 - \Gamma^{T_{on}}}{1 - \Gamma} \right)^2, \qquad D_{z_{[jT]}} = \Gamma^T D_{z_{[jT_{\text{on}}]}}. \tag{16}$$

We now provide characterization for the expected value of detection during an multiple period integrity attack as a function of system dynamics, attacker inputs and SPRT parameters in the following theorem.

Theorem* 4.1:** ***(Expected detection time for integrity attack with multiple periods) Let mT be the periodic attack time with m peirods, such that the integrity attack is active in the T_{on} component of each period. Then, the expected detection time of the attack is characterized by the following expression.

$$\mathbb{E}_{\mathcal{H}_{1,1}}[\tau_m] = mT + \frac{\ln\left(1 + \left[\frac{D_{on} + D_{off} - \mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\gamma]}{\kappa D_{z_{[mT]}}}\right]\right)}{2\ln(\Gamma)} \tag{17}$$

where D_{on} and D_{off} represent the divergence[2] of innovations during T_{on} and $T - T_{on}$ for the attack time mT.

Proof: By proceeding similarly as in the case of integrity attack with single period, if $mT < \tau_m$, the expected value of Λ_{τ_m} can be expressed as

$$\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_{\tau_m}] = \sum_{J=1}^{m} \sum_{i=1}^{T_{\text{on}}} \left[D_{z_{[i]}}^{J'} + D_{z_{[i]}}^{J''} \right] + \sum_{i=mT+1}^{\tau_m} D_{z_{[i]}} \tag{18}$$

where $D_{z_{[i]}}^{J'}$ and $D_{z_{[i]}}^{J''}$ are divergences of J^{th} period i^{th} instant during T_{on} component and during $T - T_{on}$ component respectively. If we set $\mathbb{E}_{\mathcal{H}_{1,1}}[z_0] = 0$, then for each J^{th} period, from Lemma 3.1, (8) and (16) we have,

$$\sum_{i=1}^{T_{\text{on}}} D_{z_i}^{J'} = \sum_{i=1}^{T_{\text{on}}} D_{z_{[i]}} + \Gamma^{2i} D_{(J-1)T} + \frac{2\Gamma^i \mathbb{E}[z_i]\mathbb{E}[z_{(J-1)T}]}{2\sigma^2}$$

$$\sum_{i=1}^{T_{\text{on}}} D_{z_i}^{J''} = \sum_{i=1}^{T_{\text{on}}} \Gamma^{2i} D_{z_{[(J-1)T]}}$$

$$\sum_{i=mT+1}^{\tau_m} D_{z_{[i]}} = \Gamma^2 \left(\frac{1 - \Gamma^{2(\tau_m - mT)}}{1 - \Gamma^2} \right) D_{z_{[mT]}} \tag{19}$$

[2]See (9) for equivalent notions of divergence in single period

By substituting (19) and denoting $\kappa = \frac{\Gamma^2}{1-\Gamma^2}$ in (18) we have,

$$\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_{\tau_m}] = \underbrace{m \sum_{i=1}^{T_{\text{on}}} D_{z_{[i]}} + \sum_{J=1}^{m} \sum_{i=1}^{T_{\text{on}}} 2\Gamma^i \sqrt{D_{z_{[i]}} D_{z_{[(J-1)T]}}}}_{D_{on}}$$
$$+ \underbrace{\sum_{J=1}^{m} \kappa \left[D_{z_{[JT_{\text{on}}]}} + D_{z_{[(J-1)T]}} \right] \left(1 - \Gamma^{2T_{\text{on}}}\right)}_{D_{off}}$$
$$+ \kappa \left(1 - \Gamma^{2(\tau_m - mT)}\right) D_{z_{[mT]}} \tag{20}$$

By equating $\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_{\tau_m}]$=$\mathbb{E}_{\mathcal{H}_{1,1}}[\Lambda_\gamma]$ (see proof of Theorem 3.2 for similar argument) and there by rearranging terms, followed by taking expectations (with respect to $\mathcal{H}_{1,1}$) on both sides of the equality we have the desired result for the expected detection time in multiple periods setting. ■

Hence, for the attacker to be undetected during multiple periodic attacks he should choose his inputs such that $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau_m] < \mathcal{T}$. In the case of DoS attack, the attacker should jam packets in each $T - T_{on}$ component of m periods in such a way that $\tilde{p}_e < \tilde{p}_{max}$ for the detection horizon time $\mathcal{T}$, Lemma 3.3 can be used to calculate $\tilde{p}_{max}$.

V. An illustrative example

In this section we validate our analysis for the single-period attack case. We consider the following parameters for the dynamical system: $a = 0.5$, the nominal input u_k is given by the LQG controller, process and measurement noise statistics are $\sigma_w^2 = 0.5$, $\sigma_w^2 = 1$, and the measurement channel erasure probability is $p_e = 0.2$. The type (i) error α and type (ii) error β for testing SPRT are set to 0.05 and 0.80. Finally, we set the detection horizon time to $\mathcal{T} = 50$ samples. We consider the following attacks.

Case 1 (*Integrity attack with no DoS attack*): In this scenario we inject malicious inputs ($u_k^a = 5u_k$) for a time period $T_{\text{on}} = 15$ samples and remain inactive for rest of the time. The expected time ($\mathbb{E}_{\mathcal{H}_{1,1}}[\tau]$) for which the attacker remains undetected turns to be greater than 18 time samples (see (12)). To see if this result is valid, we performed simulations for five different realizations (see Fig. 2) and observe that the estimates are indeed correct.

Case 2 (*Integrity attack with DoS attack*): In this section we consider the coordinated attack where the attacker in addition to the integrity also casts DoS attack on the output sensor for a time period $T_{\text{on}} = T - T_{\text{on}} = 15$ samples. Furthermore, the erasure probability $\tilde{p}_e$ for DoS attack has been chosen be less than $\tilde{p}_{max}$. Fig. 3 shows the behavior of LLR for the integrity attacks for five different realizations. We Observe that the average detection time in this type of attack is more than the case of individual attacks (i.e., $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau]$=18). This is because of the fact that there is no increase in LLR whenever the defender receives no measurement (see Section III-B).

Case 3 (*Effect of $\tilde{p}_e$ on DoS attack*): Finally, we study the effect of $\tilde{p}_e$ (attacked channel erasure probability) on the LLR of DoS attacks. Fig. 4 shows the LLR statistic for DoS attack and it is clear that if the attacker chooses $\tilde{p}_e$ value to be less than $\tilde{p}_{max}$ (0.30 in this case) the attack remains undetected, which supports the claim of Lemma 3.3. This

shows the importance of selecting $\tilde{p}_e$ so that the attacker can cause maximum degradation to system without being detected in DoS attacks and also increasing detection time duration in integrity attacks (see case 2).

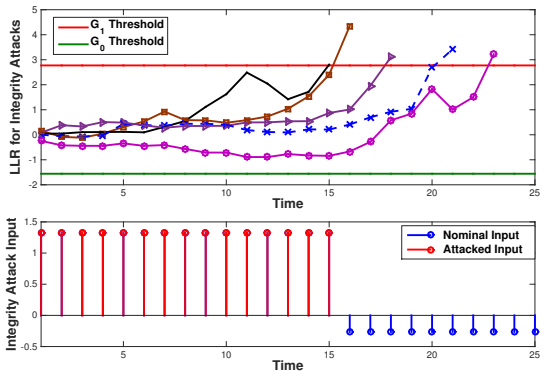

Fig. 2. In the presence of integrity attacks the average detection time for the five realizations is close to 18 which is consitent with $\mathbb{E}_{\mathcal{H}_{1,1}}[\tau]$

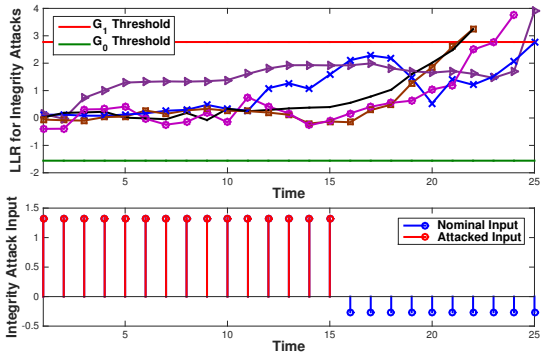

Fig. 3. Due to the coordination of DoS with integrity attack, detection time of integrity attack in this scenario is more than the detection time of individual attack (see Fig 2 for comparison).

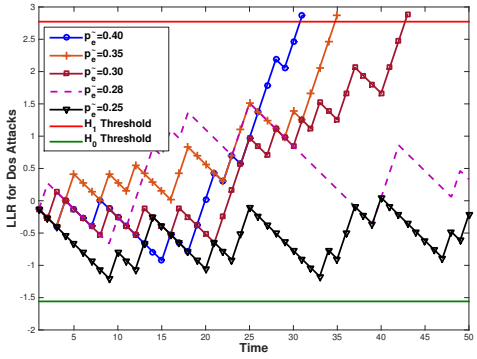

Fig. 4. As the DoS attacked channel erasure probability $\tilde{p}_e$ increases, the detection time decreases. This result is consistent with the Lemma 3.3, which provides upper bound for $\tilde{p}_e$ (0.30 in this case) to remain undetected.

VI. Conclusion

In this paper we study combined attacks against stochastic cyber-physical systems, that is, we investigate the case where the attacker is capable of compromising the system through independent, and concurrent, attack modalities. By doing so, the attacker avoids detection for a longer time, while compromising the system to a greater extent. We derive expressions for the detectability of combined attacks for periodic (constant) attacks, and assuming that the defender employs SPRT for detection. Our results show how the system dynamics, noise statistics and attack parameters influence the expected detection time. Several aspects are left for future investigation, including a generalization to non-periodic attacks, and the study of combined, simultaneous, and time-varying attacks.

References

[1] L. Lamport, R. Shostak, and M. Pease, "The Byzantine generals problem," *ACM Transactions on Programming Languages and Systems*, vol. 4, no. 3, pp. 382–401, 1982.

[2] D. Dolev, "The Byzantine generals strike again," *Journal of Algorithms*, vol. 3, pp. 14–30, 1982.

[3] M. Bishop, "What is computer security?" *Security & Privacy, IEEE*, vol. 1, no. 1, pp. 67–69, 2003.

[4] C. P. Pfleeger and S. L. Pfleeger, *Security in computing*. Prentice Hall Professional, 2003.

[5] R. Patton, P. Frank, and R. Clark, *Fault Diagnosis in Dynamic Systems: Theory and Applications*. Prentice Hall, 1989.

[6] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, 1993.

[7] 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*, Berkeley, CA, USA, 2008, pp. 6:1–6:6.

[8] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 1–15, 2012.

[9] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *ACM Conference on Computer and Communications Security*, Chicago, IL, USA, Nov. 2009, pp. 21–32.

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

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

[12] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[13] C.-Z. Bai, F. Pasqualetti, and V. Gupta, "Security in stochastic control systems: Fundamental limitations and performance bounds," in *American Control Conference*, Chicago, Il, Jul. 2015, pp. 195–200.

[14] 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.

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

[16] S. Sundaram and C. 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.

[17] R. Smith, "A decoupled feedback structure for covertly appropriating network control systems," in *IFAC World Congress*, Milan, Italy, Aug. 2011, pp. 90–95.

[18] G. K. Befekadu, V. Gupta, and P. J. Antsaklis, "Risk-sensitive control under markov modulated denial-of-service (dos) attack strategies," *IEEE Transactions on Automatic Control*, vol. 60, no. 12, pp. 3299–3304, Dec 2015.

[19] C. D. Persis and P. Tesi, "On resilient control of nonlinear systems under denial-of-service," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, Dec 2014, pp. 5254–5259.

[20] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. Prentice Hall Upper Saddle River, NJ, 2000.

[21] A. Wald, *Sequential Analysis*. New York, Wiley, 1947.

[22] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.