2019 American Control Conference (ACC)
Philadelphia, PA, USA, July 10-12, 2019

A Probabilistic Approach to Design Switching Attacks against Interconnected Systems

Rajasekhar Anguluri, Vaibhav Katewa, and Fabio Pasqualetti

Abstract— **In this paper we study an attack design problem for interconnected systems where the attacker compromises a subsystem at each time, which is selected randomly based on a pre-computed probabilistic rule. The objective of the attacker is to degrade the system performance, which is measured based on a quadratic function of the system state, while remaining undetected from a centralized detector. First, we derive an explicit expression for the detection probability, analyze its properties, and compute an upper bound. Then, we use our upper bound to formulate and numerically solve a non-convex optimization problem for the computation of optimal attack strategies. Finally, we validate our results and show that our probabilistic attack strategy outperforms a deterministic attack strategy that compromises a fixed subsystem at each time.**

I. Introduction

Modern technological systems are large and inevitably comprise different subsystems. Each subsystem may be responsible for one or more functions and, upon interconnection, they realize the complex functions of the whole system [1]. The applications of these systems are far reaching, ranging from power and water networks, to telecommunication and transportation systems [2]. The performance of these systems is determined not only by the performance of the individual subsystems, but also by their interconnection dynamics. Importantly, large scale interconnected systems are prone to attacks at the subsystem and interconnection levels, thereby making their operation even more fragile [3].

In this paper, we study a security problem for interconnected systems, where the objective of the attacker is to degrade the performance of the interconnected system by compromising subsystems, while maintaining undetectability. In particular, we develop a probabilistic rule to randomly select an attacked subsystem over time, and optimize over the switching probabilities to maximize the degradation and maintain undetectability from a centralized detector, which uses a chi-squared test. Overall, our results show that the ability to selectively compromise different parts of a system over time greatly increases the severity of the attacks, thereby motivating the development of advanced detection schemes for interconnected system [4].

Related work In the last few years, with security emerging as a major concern for real time dynamical systems, different attack models and possible remedial frameworks have been studied by researchers to a great extent [5], [6], [7], [8]. Although, these works provide deep insights into the attackers capabilities in compromising systems, several of these works mainly restrict their attention to the attacks that target fixed subparts or the overall system, thereby undermining the vulnerabilities posed by the interconnected systems, at various subsystem and interconnection levels.

Only recently, researchers started to study attack models considering the challenges posed by the interconnected systems. A few notable works in this direction are as follows: Exploiting the sparsity structure in deterministic systems, authors in [9] proposed dynamic decoders to estimate the initial state accurately. Instead, for the stochastic systems, few authors proposed robust state estimation techniques exploiting the tools from hidden mode switching systems [10], [11]. Using the variable structure systems theory, authors in [12] demonstrated switching attacks that can disrupt the operation of the power grid within a short interval of time. Instead, authors in [13] considered a game-theoretic approach based power system stabilizers to counter attack switching attacks in smart grids. Further, few authors studied the detrimental effects due to coordinated attacks in cyber-physical systems [14], [4]. In our work we consider switching attacks using a probabilistic framework and argue for the need of studying them in the context of interconnected systems.

Contribution: The contribution of this paper is three-fold. First, we develop an attack model which randomly, through some pre-assigned probabilistic rule, compromise a subsystem. Second, we characterize the detection probability of a centralized detector, with respect to these attacks, and, derive upper bounds on the detection probabilities, both in the finite and asymptotic cases. Third, we formulate and numerically solve an optimization problem for computing optimal probabilistic rules with constraints on the detection probability. Finally, we demonstrate the superiority of using our optimal probabilistic strategy against attacking fixed subsystem strategy using a numerical example.

Paper organization: The remainder part of the paper is organized as follows. In Section II we introduce our interconnected system model, attack model, and pose attacker's objectives in an optimization problem. In Section III we illustrate a detection procedure using hypothesis testing framework, and characterize its detection probability. Section IV contains our attack design strategy followed by a numerical example. In Section V we conclude the paper.

Mathematical notation: $\mathrm{Tr}(\cdot)$ and $\mathrm{diag}()$ denote the trace and a vector of diagonal elements of a matrix, respectively. $\mathrm{blkdiag}(A_1, A_2, \ldots, A_n)$ denotes a block diagonal matrix

This material is based upon work supported in part by ARO award 71603NSYIP, and in part by NSF award ECCS1405330. Rajasekhar Anguluri, Vaibhav Katewa, Fabio Pasqualetti are with the Department of Mechanical Engineering, University of California at Riverside, {ranguluri,vkatewa,fabiopas}@engr.ucr.edu.

978-1-5386-7926-5/$31.00 ©2019 AACC

whose block diagonal entires are $A_1, A_2, \ldots, A_N$. A zero mean normally distributed random variable Y is denoted by $Y \sim \mathcal{N}(0, \Sigma)$, where Σ is the covariance of Y. If Y follows a noncentral chi-squared distribution, we denote it by $Y \sim \chi^2(m, \lambda)$, where p is the degrees of freedom and λ is the non-centrality parameter. Instead, if $\lambda = 0$, we denote it as $Y \sim \chi^2(m)$. CDF denotes the cumulative distributive function of a random variable. $\mathbf{1}$ denotes the all ones vector and, for $x \in \mathbb{R}^n$, $x \geq 0$ denotes the element wise inequality.

II. Problem setup and preliminary notions

A. Nominal system model

We consider an interconnected system composed of N interacting subsystems whose dynamics are as follows:

$$x_i(k+1) = A_{ii}x_i(k) + B_i u_i(k) + \sum_{j \neq i}^{N} A_{ij}x_j(k) + w_i(k),$$
$$y_i(k) = C_i x_i(k) + v_i(k),$$

where $x_i \in \mathbb{R}^{n_i}$ is the system state, $y_i \in \mathbb{R}^{m_i}$ is the measurement, and $u_i \in \mathbb{R}^{q_i}$ is the known input of the i-th subsystem. Further, $w_i \sim \mathcal{N}(0, W_i)$, $v_i \sim \mathcal{N}(0, V_i)$ are the process and measurement noise affecting the i-th subsystem dynamics. Since the input u_i is known, its contribution to the output $y_i(k)$ is also known and, therefore, $u_i(k)$ can be ignored. In vector form, the system dynamics read as

$$\begin{aligned} x(k+1) &= Ax(k) + w(k), \\ y(k) &= Cx(k) + v(k), \end{aligned} \tag{1}$$

where the state x, measurements y, and the noise vectors w and v of the interconnected system are given by $x = \begin{bmatrix} x_1^\mathsf{T} & \ldots & x_N^\mathsf{T} \end{bmatrix}^\mathsf{T}$, $y = \begin{bmatrix} y_1^\mathsf{T} & \cdots & y_N^\mathsf{T} \end{bmatrix}^\mathsf{T}$, $w = \begin{bmatrix} w_1^\mathsf{T} & \ldots & w_N^\mathsf{T} \end{bmatrix}^\mathsf{T} \in \mathbb{R}^n$, $v = \begin{bmatrix} v_1^\mathsf{T} & \ldots & v_N^\mathsf{T} \end{bmatrix}^\mathsf{T} \in \mathbb{R}^m$, $n = \sum_{i=1}^N n_i$, and $m = \sum_{i=1}^N m_i$. Furthermore, we have

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{bmatrix} \text{ and } B = \begin{bmatrix} C_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_N \end{bmatrix}.$$

The initial state $x(0) \sim \mathcal{N}(0, \Sigma_0)$, the noises $w \sim \mathcal{N}(0, W)$ and $v \sim \mathcal{N}(0, V)$ are uncorrelated, for all $k \in \mathbb{N}$, where the noise covariance matrices are $W \triangleq \text{blkdiag}(W_1, \cdots, W_N)$ and $V \triangleq \text{blkdiag}(V_1, \cdots, V_N)$.

We assume that (1) is operating in steady state. We allow for the presence of attackers that compromise the dynamics of the subsystems, and we model such attacks as exogenous unknown inputs (see Section II-B). We equip system (1) with a detector whose role is to trigger an alarm, based on the innovations signals generated by a Kalman filter. Assuming (A, C) is observable and (A, W) is controllable, a steady state Kalman filter employs the following recursion:

$$\begin{aligned} \widehat{x}(k) &= A\widehat{x}(k-1) + Kz_k, \\ z(k) &= y(k) - CA\widehat{x}(k-1), \end{aligned} \tag{2}$$

where $\widehat{x}(k) \triangleq \mathbb{E}[x(k)|y(0), \ldots, y(k)]$ is the *minimum mean squared error* (MMSE) estimate of $x(k)$, and $K = PC^T[CPC^T + V]^{-1}$ and $P = A(I - KC)PA^T + W$. Further, the innovations $z(k) \sim \mathcal{N}(0, \Sigma)$ forms an i.i.d sequence with covariance $\Sigma = CPC^T + V$.

B. Objectives of attacker and attacked system model

We assume that the main objective of an attacker is to inject malicious inputs into the system (1) such that,

(i) at any given time, one subsystem is selected with a probabilistic rule to inject a malicious inputs and
(ii) the rule used in (i) should maximize a quadratic cost of the state in (1) with minimum detection probability.

First, we model our randomized policy to select subsystem at every time instant. Then, we develop an optimization framework to select an optimal policy. Let $\{a_k\}_{k=0}^{\infty}$ be a scalar valued i.i.d stochastic process, taking value in the finite set $\{1, \ldots, N\}$, at every time $k \in \mathbb{N}$, with probability $\mathbb{P}[a_k = i] \triangleq p_i$, for all $i \in \{1, \ldots, N\}$, such that $\sum_{i=1}^N p_i = 1$. Let $p \triangleq [p_1, \ldots, p_N]^\mathsf{T}$, and note that p denotes the probabilities of selecting subsystems. Thus, by specifying p, the attack process $\{a_k\}_{k=0}^{\infty}$, realizes a subsystem index, for any given time k. Let $\delta_i(a_k)$ be a indicator random variable of a_k, i.e., $\delta_i(a_k) = 1$ if $a_k = i$, else $\delta_i(a_k) = 0$ otherwise. Let $\delta(a_k) \triangleq [\delta_1(a_k), \ldots, \delta_N(a_k)]^\mathsf{T}$. Then, the attacked system dynamics can be modeled as

$$\begin{aligned} x^e(k+1) &= Ax^e(k) + w(k) + \Pi(k)\delta(a_k), \\ y^e(k) &= Cx^e(k) + v(k) + \Psi(k)\delta(a_k), \end{aligned} \tag{3}$$

where $x^e(k)$ and $y^e(k)$ denote the state and the measurement of the interconnected system under attack. The attack matrices are $\Pi \triangleq \text{blkdiag}(\Pi_1 u_1, \ldots, \Pi_N u_N)$ and $\Psi \triangleq \text{blkdiag}(\Psi_1 \tilde{u}_1, \ldots, \Psi_N \tilde{u}_N)$, respectively, where $\Pi_i u_i(k)$ and $\Psi_i \tilde{u}(k)$ are the malicious inputs that an attacker wants to inject into the i-th subsystem dynamics at time k. Further, we assume that the process $\{a_k\}_{k=0}^{\infty}$ is independent of $w(k)$ and $v(k)$. Thus, the random variables $\delta(a_k)$, $w(k)$, and $v(k)$ are mutually independent, for all $k \in \mathbb{N}$.

Let $P^D(k)$ be the detection probability of a detector. Then, the attacker's goal can be cast as an optimization problem:

$$\textbf{(P.1)} \quad \underset{p \geq 0}{\arg\max} \quad \mathbb{E}\left[\sum_{k=0}^{T-1} x^e(k+1)^\mathsf{T} x^e(k+1)\right],$$
$$\text{subject to} \quad \mathbf{1}^\mathsf{T} p = 1 \tag{4}$$
$$P^D(k) \leq \zeta \quad \forall k \in \{0, \ldots, T-1\}, \tag{5}$$

where the expectation, $\mathbb{E}[\cdot]$, is taken over the noise variables and the process $\{a_k\}_{k=0}^{T-1}$. To solve (p.1), from an attacker's stand point of view, we make the following assumption.

Assumption 2.1: The attacker has full knowledge about the matrices of the system (1) and of the Kalman filter (2).

C. Relation between nominal and attacked system

In this section we characterize the bias accumulated in the interconnected system dynamics (1) and the Kalman filter dynamics (2) due to the attacks. Let $\gamma(k)$ and $\beta(k)$ denote the bias in the state and the measurements of (1), respectively. Then, $x^e(k) = x(k) + \gamma(k)$ and $y^e(k) = y(k) + \beta(k)$, where

$$\begin{aligned} \gamma(k+1) &= A\gamma(k) + \Pi(k)\delta(a_k), \\ \beta(k) &= C\gamma(k) + \Psi(k)\delta(a_k). \end{aligned} \tag{6}$$

Now, consider the following filter under attacks

$$\begin{aligned} \widehat{x}^e(k) &= A\widehat{x}^e(k-1) + Kz^e(k), \\ z^e(k) &= y^e(k) - CA\widehat{x}^e(k-1), \end{aligned} \tag{7}$$

where $\widehat{x}_k^e$ and z_k^e are analogous to the state estimate and the innovations defined in (2). Let $\alpha(k)$ and $\epsilon(k)$ are biases accumulated in the MMSE estimate and innovations due to the attack. Then, it is easy to see that $\widehat{x}^e(k) = \widehat{x}(k) + \alpha(k+1)$ and $z^e(k) = z(k) + \epsilon(k)$, respectively. By using (6) and (7), $\alpha(k)$ and $\epsilon(k)$ can be obtained recursively as

$$\begin{aligned} \alpha(k+1) &= (I - KC)A\alpha(k) + K\beta(k), \\ \epsilon(k) &= C\left[\gamma(k) - A\alpha(k)\right] + \Psi(k)\delta(a_k). \end{aligned} \tag{8}$$

Notice that in the absence of attacks, the bias satisfy $\epsilon(k) = 0$, and $z^e(k) = z(k)$ for all $k \in \mathbb{N}$. Instead, in the presence of attacks, $\epsilon(k) \neq 0$ and $z^e(k) \neq z(k)$ (at least for one k).

III. Detection framework

Let H_0 and H_1 be the null and the alternative hypothesis corresponding to the presence and absence of attacks, respectively. We assume that the detector uses a chi-squared test statistic, to compare with a threshold (τ) and decide against the attacks [15], [16]. Formally, we have following test

$$\Lambda(k) \triangleq z^e(k)^\mathsf{T}\Sigma^{-1}z^e(k) \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \quad \forall k \in \mathbb{N}. \tag{9}$$

The false alarm probability (P^F) and the detection probability (P^D) of the test (9) are defined in the following way:

$$\begin{aligned} &P^F(k) \triangleq \mathbb{P}\left[\Lambda(k) \geq \tau | H_0\right], \text{ and} \\ &P^D(k) \triangleq \mathbb{P}\left[\Lambda(k) \geq \tau | H_1\right]. \end{aligned}$$

We assume that $P^F(k) = P^F$ is identical for all $k \in \mathbb{N}$. By recalling the fact that under H_0 (no attack) the bias $\epsilon(k) = 0$, we have $z^e(k) \sim \mathcal{N}(0, \Sigma)$. It now follows that, under H_0, $\Lambda(k) \sim \chi^2(m)$, where m is the degrees of freedom. Further, we assume that P^F is predetermined and the threshold τ is computed by the inverse CDF of $\chi^2(m)$.

A. Characterization of the detection probability

Notice that, in order to inject attacks that can evade the detector, i.e., bypass the test (9), the attacker needs to know $P^D(k)$. Thus, in this section we derive an expression for $P^D(k)$. We now state a proposition that expresses the bias $\epsilon(k)$ in the terms of attack input matrices $\Pi(k)$ and $\Psi(k)$, respectively.

Proposition 3.1: Let $\mathcal{A} = A(I - KC)$ and $\mathcal{B}(k) = \Pi(k) - AK\Psi(k)$. Then,

$$\epsilon(k) = \underbrace{\left[C\mathcal{A}^{k-1}\mathcal{B}(0) \ \ldots \ C\mathcal{B}(k-1) \ \ \Psi(k)\right]}_{\triangleq \mathcal{E}_k}\delta(a_{0:k}), \tag{10}$$

where $\delta(a_{0:k}) = \left[\delta(a_0)^\mathsf{T} \quad \ldots \quad \delta(a_k)^\mathsf{T}\right]^\mathsf{T}$.

Proof: See the Appendix. ■

Consider the truncation $\{a_j\}_{j=0}^k$ from the original process $\{a_j\}_{j=0}^\infty$. Let $\mathcal{S}_k$ denote the set of all possible realizations of $\{a_j\}_{j=0}^k$, and $\pi_k \in \mathcal{S}_k$. The components of π_k can be enumerated as $[\pi_k^0, \pi_k^1 \ldots, \pi_k^k]$. With slight abuse of notation define $\delta(\pi_k) \triangleq [\delta_1(\pi_k^0)^\mathsf{T}, \ldots, \delta_N(\pi_k^k)^\mathsf{T}]^\mathsf{T}$. We emphasize that $\delta(a_{0:k})$ is a random vector but $\delta(\pi_k)$ is a deterministic vector.

Lemma* 3.2: *(Detection probability) The detection probability of the test (9) is given by

$$P^D(k) = \sum_{\pi_k \in \mathcal{S}_k} Q(\tau; m, \lambda(\pi_k))p_{\pi_k^0}p_{\pi_k^1} \cdots p_{\pi_k^k}, \tag{11}$$

where $Q(\tau; r, \lambda(\pi_k))$ is the complementary CDF of $\chi^2(\tau, \lambda(\pi_k))$ and $\lambda(\pi_k) \triangleq \widetilde{\delta}(\pi_k)^\mathsf{T}\mathcal{E}_k^\mathsf{T}\Sigma^{-1}\mathcal{E}_k\widetilde{\delta}(\pi_k)$.

Proof: See the Appendix. ■

For the attacks that randomly select a subsystem, Lemma 3.2 states that the $P^D(k)$ is a weighted sum of detection probability, $Q(\tau; m, \lambda(\pi_k))$, associated with all possible ways of selecting the locations. Also, notice that the expression (11) depends on the matrices of the interconnected system and the KF through the impulse response $\mathcal{E}_k$ in (10) of $\lambda(\pi_k) = \widetilde{\delta}(\pi_k)^\mathsf{T}\mathcal{E}_k^\mathsf{T}\Sigma^{-1}\mathcal{E}_k\widetilde{\delta}(\pi_k)$. Finally, by assumption 2.1, we note the attacker has the capability to compute $P^D(k)$.

B. Upper bound on the detection probability

Although the formula of $P^D(k)$ we obtained in Lemma 3.2 is exact, the number of summands in (11) increases exponentially with time k. Hence, for practical purposes, computing the detection probability using (11) is not efficient. In this section we provide an upper bound on $P^D(k)$ using Markov's inequality. We now define the following matrices that will be helpful in expressing our bound compactly:

$$\mathcal{E}_k^\mathsf{T}\Sigma^{-1}\mathcal{E}_k \triangleq \begin{bmatrix} L_k(0,0) & \cdots & L_k(0,k) \\ \vdots & \ddots & \vdots \\ L_k(k,0) & \ldots & L_k(k,k) \end{bmatrix}, \tag{12}$$

where $L_k(i,j)$, $0 \leq i,j \leq k$ is obtained by performing block wise multiplication of matrices in $\mathcal{E}_k^\mathsf{T}$ with those in $\Sigma^{-1}\mathcal{E}_k$. Moreover, this construction results in $L_k(i,j) = L_k(j,i)^\mathsf{T}$, for all i,j. Further, define $\overline{L}_k$ and $\widehat{L}_k$ as

$$\overline{L}_k \triangleq \sum_{i=j} L_k(i,j) \text{ and } \widehat{L}_k \triangleq \sum_{i \neq j} L_k(i,j), \tag{13}$$

respectively. It now follows that $\overline{L}_k$ is a positive semi definite matrix, while $\widehat{L}_k$ is only a symmetric matrix.

Lemma* 3.3: *(Upper bound of the detection probability) Let $p = [p_1, p_2, \ldots, p_N]^\mathsf{T}$ be the vector of probabilities with p_i denoting the probability of attacking the i-th subsystem, $\forall i \in \{1, \ldots, N\}$. Then, for all $k \in \mathbb{N}$, it holds that

$$P^D(k) \leq \underbrace{\frac{m + \text{diag}(\overline{L}_k)^\mathsf{T}p + p^\mathsf{T}\widehat{L}_k p}{\tau}}_{\overline{P}^D(k)}. \tag{14}$$

Proof: See the Appendix. ■

Notice that, unlike the expression in (11), the upper bound $\overline{P}^D(k)$ is a quadratic expression in the probability vector p. Further, $\overline{P}^D(k)$ does not depend on the Q function, which is an infinite series. Rather it depends on the impulse response (10) through the matrices $\overline{L}_k$ and $\widehat{L}_k$. Finally, note that the bound becomes loose if τ is not sufficiently large.

C. Asymptotic upper bound

To characterize an asymptotic expression for the bound $\overline{P}^D(k)$ when $k \to \infty$, we make the following assumption:

Assumption 3.4: The attack matrices are constant all times, i.e., $\Pi(k) \triangleq \Pi$ and $\Psi(k) \triangleq \Psi$ for all $k \in \mathbb{N}$.

***Lemma* 3.5: (Asymptotic upper bound of $P^D(k)$)** Let $\overline{P}^D(k)$ be as in (14). Then,

$$\overline{P}^D_\infty \triangleq \lim_{k\to\infty} P^D(k) = \frac{m + \operatorname{diag}(\overline{L}_\infty)^\mathsf{T} p + p^\mathsf{T} \widehat{L}_\infty p}{\tau}.$$

where

$$\begin{aligned}
\overline{L}_\infty &= \mathcal{B}^\mathsf{T} \mathcal{O} \mathcal{B} + \Psi^\mathsf{T} \Sigma^{-1} \Psi,\\
\widehat{L}_\infty &= \mathcal{B}^\mathsf{T} \left[\mathcal{O} - \mathcal{M}\right] \mathcal{B} - \Psi \Sigma^{-1} \Psi,\\
\mathcal{B} &= \Pi - AK\Psi,\\
\mathcal{O} &\triangleq \sum_{j=0}^{\infty} (\mathcal{A}^j)^\mathsf{T} C^\mathsf{T} \Sigma^{-1} C \mathcal{A}^j, \text{and}\\
\mathcal{M} &\triangleq (I - \mathcal{A})^{-\mathsf{T}} C^\mathsf{T} C (I - \mathcal{A})^{-1}.
\end{aligned}$$

Proof: See the Appendix. ■

As $P^D(k) \le \overline{P}^D(k)$, for all $k \in \mathbb{N}$, from Lemma 3.5, we note that for large k, $\overline{P}^D(k)$ is constant. Intuitively, if the attacker is not detected during the transient of the filter (2) dynamics, since the beginning of attacks, then it is unlikely for an attacker to be detected once the filter (2) reaches the steady state. Finally, if the matrices Π and Ψ are chosen such that the constraint (16), i.e., $P^D(k) \le \zeta$ for $k \in \{0, \ldots, T_0 - 1\}$, where T is sufficiently large, Lemma 3.5 guarantees that $P^D(k) \le \zeta$, for all times $k \in \mathbb{N}$. This type of asymptotic analysis helps the attacker to select attack matrices that yields minimum detection probability $P^D(k)$.

IV. Design of an optimal probabilistic strategy

In this section we solve the optimization problem (P.1) described in Section II with the help of numerical optimization techniques. First, we rewrite the cost function of (P.1) in such way that it depends explicitly on the variable p. Under Assumption 3.4, consider the following impulse response matrices associated with the system (6):

$$\mathcal{H}_{T-1} = \begin{bmatrix} \Pi & \cdots & 0 \\ \vdots & \ddots & \vdots \\ A^{T-1}\Pi & \ldots & \Pi \end{bmatrix} \text{ and,}$$

$$\mathcal{H}_{T-1}^\mathsf{T} \mathcal{H}_{T-1} = \begin{bmatrix} G(0,0) & \cdots & G(0,T-1) \\ \vdots & \ddots & \vdots \\ G(T-1,0) & \ldots & G(T-1,T-1) \end{bmatrix}. \quad (15)$$

Also, let $\overline{G}_{T-1} = \sum_{i=j} G(i,j)$ and $\widehat{G}_{T-1} = \sum_{i\neq j} G(i,j)$. It is straightforward to see that $\overline{G}_{T-1}$ is a positive definite matrix, while $\widehat{G}_{T-1}$ is a symmetric matrix. The following proposition expresses the cost function of (P.1) in terms of the matrices $\overline{G}_{T-1}$ and $\widehat{G}_{T-1}$.

Proposition 4.1: The cost function of (P.1) can be equivalently replaced by $\operatorname{diag}(\overline{G}_{T-1})^\mathsf{T} p + p^\mathsf{T} \widehat{G}_{T-1} p$.

Proof: See the Appendix. ■

As $P^D(k)$ is inefficient for computational purposes we relax the constraint (5) of (P.1) by replacing it with constraint on the upper bound $\overline{P}^D(k)$. By incorporating the aforementioned changes in (P.1) we now have the following quadratically constrained quadratic programming type problem, whose solution yields a sub-optimal probabilistic attack strategy with respect to the original problem (P.1).

$$\begin{aligned}
\textbf{(P.2)} \quad & \arg\max_{p\ge 0} && \operatorname{diag}(\overline{G}_{T-1})^\mathsf{T} p + p^\mathsf{T} \widehat{G}_{T-1} p,\\
& \text{subject to} && \mathbf{1}^\mathsf{T} p = 1,\\
& && \operatorname{diag}(\overline{L}_k)^\mathsf{T} p + p^\mathsf{T} \widehat{L}_k p \le \tau\, \zeta - m\\
& && \qquad \forall k \in \{0, \ldots, T-1\} \quad (16)
\end{aligned}$$

Notice that (P.2) is a non-convex optimization problem, since the matrices $\widehat{L}_k$, for all $k \in 0, \ldots, T-1$, and $\widehat{G}_{T-1}$ are only symmetric matrices. Thus, the standard convex optimization techniques/analysis are not applicable. Hence, to obtain a feasible solution to the maximization problem (P.2) we use standard numerical solvers. We also note that this optimal solution might not be a global maximum.

A. Numerical Example

We consider a chemical reactor consisting of two continuous stirred-tank reactors [17]. The discretized system matrices, with sampling time $T_s = 1$sec, are given by

$$A_{11} = \begin{bmatrix} 0.2603 & -0.1862 \\ 0.1862 & 0.2603 \end{bmatrix}, A_{12} = \begin{bmatrix} -0.0188 & -0.0230 \\ 0.0232 & -0.00188 \end{bmatrix},$$
$$A_{21} = \begin{bmatrix} -0.0215 & -0.0266 \\ 0.0263 & -0.0215 \end{bmatrix}, A_{22} = \begin{bmatrix} -0.3120 & 0.2713 \\ -0.2713 & -0.3120 \end{bmatrix}.$$

We consider the state and measurement attack matrices as

$$\Pi = \Psi = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}^\mathsf{T}. \quad (17)$$

Our results are illustrated in Fig. 1 and Fig. 2. For the probabilistic rule $p = [0.5, 0.5]^\mathsf{T}$ and $P^F = 0.01$, in Fig. 1 we report the actual detection probability $P^D(k)$ (11) and the upper bound $\overline{P}^D(k)$. As discussed in Section III, we can now see that the bound (14) converges to a constant when T increases.

In Fig. 2 we report the values of the cost function (P.2) for the optimal probabilistic rule $p = p*$ and the fixed location rule, i.e., the degenerate probability vectors $p = [1, 0]^\mathsf{T}$ and $p = [0, 1]^\mathsf{T}$, respectively. From Fig. 2, and as expected, the optimal rule results in higher degradation of the system performance. Thus, this work shows that the use of probabilistic rule for switching location attacks benefits the attacker, as opposed to attacking fixed locations.

V. Conclusions

This paper studies a security problem for interconnected systems, where the attacker objective is to randomly compromise subsystems such that the performance degradation of interconnected system is maximum. We developed a probabilistic rule for attacking subsystems and characterized the bias accumulated in the system due to these attacks. We also characterized the detection probability of a centralized

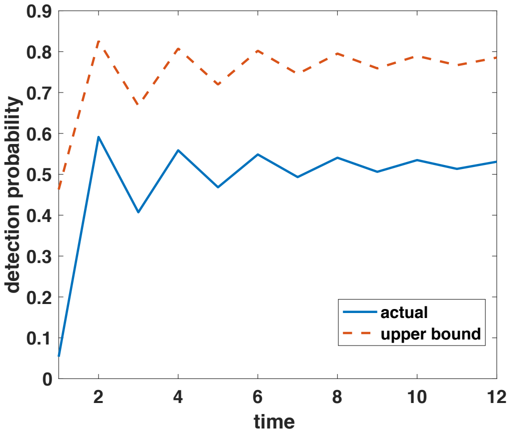

Fig. 1. This figure shows the detection probability $P^D(k)$ (blue solid line) and its corresponding upper bound $\overline{P}^D(k)$ (orange dashed line) as a function of time, which are computed using expressions in (11) and (14), respectively. For the parametric values $P^F = 0.01$, $m = 4$, and the matrices Π and Ψ in (17) we notice that, although there is an initial transience, due to the dynamics of Kalman filter, as discussed in Section III, the actual value and the bound converges to a constant.

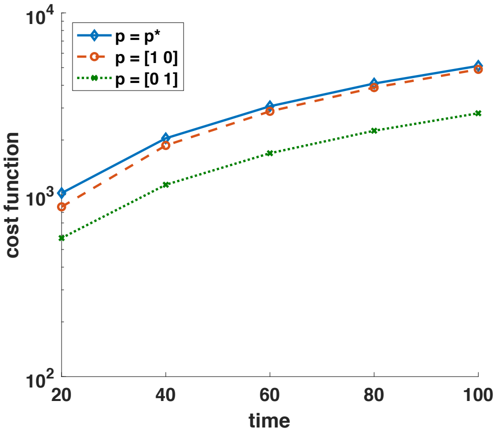

Fig. 2. This figure shows the performance degradation of interconnected systems, evaluated by the cost function value of (P.2), for various switching rules. The blue solid line correspond to the optimal probabilistic rule, that was obtained by solving (P.2) using numerical solver. The dashed orange (resp. dotted green) is obtained by using fixed attack locations. As expected, the cost value for all the rules increased with time. In particular, the optimal probabilistic rule resulted in worst performance degradation than the rest.

detector, and formulated an optimization problem to find an optimum probabilistic rule that maximizes system degradation, while maintaining minimum detection probability.

References

[1] M.S Mahmoud. *Decentralized Control and Filtering in Interconnected Dynamical Systems*. CRC Press (Taylor and Francis Group), 2011.

[2] R. Baheti and H.Gill. Cyber-physical systems. *The Impact of Control Technology*, pages 161–166, 2011.

[3] A.A Cardenas, S. Amin, B. Sinopoli, A. Giani, P. Adrian, and S. Sastry. Challenges for securing cyber physical systems. In *IEEE Workshop Future Directions Cyber-Phys.Syst. Security*, July 2009. Available at https://ptolemy.berkeley.edu/projects/chess/pubs/601.html.

[4] R. Anguluri, V. Gupta, and F. Pasqualetti. Periodic coordinated attacks against cyber-physical systems: Detectability and performance bounds. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 5079–5084, Dec 2016.

[5] Y. Z. Lun, A. DInnocenzo, F. Smarra, I. Malavolta, and M. D. Di Benedetto. State of the art of cyber-physical systems security: An automatic control perspective. *Journal of Systems and Software*, 149:174 – 216, 2019.

[6] 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.

[7] 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.

[8] Y. Mo, S. Weerakkody, and B. Sinopoli. Physical authentication of control systems. *IEEE Control Sys. Magazine*, 35(1):93–109, 2015.

[9] C. Liu, J. Wu, C. Long, and Y. Wang. Dynamic state recovery for cyber-physical systems under switching location attacks. *IEEE Transactions on Control of Network Systems*, 4(1):14–22, March 2017.

[10] S. Z. Yong, M. Zhu, and E. Frazzoli. Resilient state estimation against switching attacks on stochastic cyber-physical systems. In *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015.

[11] D. Shi, R. J. Elliott, and T. Chen. On finite-state stochastic modeling and secure estimation of cyber-physical systems. *IEEE Transactions on Automatic Control*, 62(1):65–80, Jan 2017.

[12] S. Liu, S. Mashayekh, D. Kundur, T. Zourntos, and K. Butler-Purry. A framework for modeling cyber-physical switching attacks in smart grid. *IEEE Trans. on Emerging Topics in Comp.*, 1(2):273–285, 2013.

[13] A. K. Farraj, E. M. Hammad, A. A. Daoud, and D. Kundur. A game-theoretic control approach to mitigate cyber switching attacks in smart grid systems. In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 958–963, Nov 2014.

[14] G. Wu and J. Sun. Optimal switching integrity attacks in cyber-physical systems. In *2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 709–714, May 2017.

[15] Y. Mo and B. Sinopoli. Secure estimation in the presence of integrity attacks. *IEEE Transactions on Automatic Control*, 60(4):1145–1151, April 2015.

[16] Y. Chen, S. Kar, and J. M. F. Moura. Cyber-physical attacks with control objectives. *IEEE Transactions on Automatic Control*, 63(5):1418–1425, May 2018.

[17] L. Jian and W. Chun-Yu. Finite-time robust fault detection filter design for interconnected systems concerning with packet dropouts and changing structures. *Int. Journal of Control*, pages 1–12, 2018.

[18] N L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions, Volume 2*. Wiley & Sons, 1995.

Appendix

Proof of Proposition 3.1: Let $\theta(k) \triangleq \gamma(k) - A\alpha(k)$, then from (6) and (8) it follows that

$$\begin{aligned} \theta(k+1) &= \gamma(k+1) - A\alpha(k+1) \\ &= A\left[\gamma(k) - A\alpha(k)\right] - AKC\left[\gamma(k) - A\alpha(k)\right] \\ &\quad + \Pi(k)u(k) - AK\Psi(k)u(k) \\ &= \mathcal{A}\theta(k) + \mathcal{B}(k)\delta(a_k). \end{aligned}$$

From (8), $\epsilon(k)$ can be computed in the following way

$$\begin{aligned} \theta(k+1) &= \mathcal{A}\theta(k) + \mathcal{B}(k)\delta(a_k), \\ \epsilon(k) &= C\theta(k) + \Psi(k)\delta(a_k), \end{aligned}$$

by recursively expanding $\theta(k)$ and observing that $\theta(0) = 0$, since $\gamma(0) = 0$ and $\alpha(0) = 0$, the result follows. ■

Proof of Lemma 3.2: For any $k \in \mathbb{N}$, let $I_{\{\Lambda(k) \geq \tau\}}$ be the indicator of the event $\{\Lambda(k) \geq \tau\}$, and notice that

$$\begin{aligned} P^D(k) &= \mathbb{E}\left[I_{\{\Lambda(k) \geq \tau\}} | H_1\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[I_{\{\Lambda(k) \geq \tau\}} \mid H_1, \delta(a_{0:k})\right] \mid H_1\right], \end{aligned} \tag{18}$$

where the inner expectation is with respect to $\{a_j\}_{j=0}^k$. Let $\widetilde{\delta}(\pi_k)$ be a realization of $\delta(a_{0:k})$, where $\pi_k = [\pi_k^0, \dots, \pi_k^k]^\mathsf{T}$.

Then, under H_1, we note that $\epsilon(k) = \mathcal{E}_k\widetilde{\delta}(\pi_k)$ is a deterministic quantity, and further it follows that the distribution of $z^e(k)$ given H_1 and $\delta(a_{0:k}) = \widetilde{\delta}(\pi_k)$ is $\mathcal{N}(\epsilon(k), \Sigma)$. Since, $\Lambda(k)$ is a quadratic transformation of $z^e(k)$ we have [18],

$$\Lambda(k) \mid_{H_1, \delta(a_{0:k}) = \widetilde{\delta}(\pi_k)} \sim \chi^2(m, \lambda(\pi_k)).$$

Thus, from the above characterizations it follows that

$$\mathbb{E}\left[I_{\{\Lambda(k) \geq \tau\}} \mid H_1, \delta(a_{0:k}) = \widetilde{\delta}(\pi_k)\right] = Q(\tau; m, \lambda(\pi_k)).$$

Substituting above expression in (18) and taking the expectation over all possible realizations of $a_{0:k}$ we have

$$P^D(k) = \sum_{\pi_k \in \mathcal{S}_k} Q(\tau; m, \lambda(\pi_k))\, \mathbb{P}(a_{0:k} = \pi_k). \qquad (19)$$

Since the process $\{a_k\}_{k=0}^{\infty}$ is i.i.d we now have, $\mathbb{P}(a_{0:k} = \pi_k) = \prod_{j=0}^{k} \mathbb{P}(a_j = \pi_k^j) = \prod_{j=0}^{k} p_{\pi_k^j}$. By substituting above expression in (19) the result follows. ■

Proof of Lemma 3.3: By Markov's inequality we have

$$\underbrace{\mathbb{P}[(z^e(k))^{\mathsf{T}}\Sigma^{-1}z^e(k) \geq \tau | H_1]}_{\triangleq P^D(k)} \leq \underbrace{\frac{\mathbb{E}\left[z^e(k)^{\mathsf{T}}\Sigma^{-1}z^e(k)|H_1\right]}{\tau}}_{\triangleq \overline{P}^D(k)}.$$

Notice that under hypothesis H_1, $z^e(k) = z(k) + \epsilon(k)$ and, from Proposition 3.1, $\epsilon(k) = \mathcal{E}_k\delta(a_{0:k}) \neq 0$. As the attack process $\{a_k\}_{k=0}^{\infty}$ is independent of noise random variables, it follows that $z(k)$ and $\delta(a_{0:k})$ are independent as well. Thus

$$\begin{aligned}\mathbb{E}\left[(z^e(k))^{\mathsf{T}}\Sigma^{-1}z^e(k)|H_1\right] = \mathbb{E}[&z(k)^{\mathsf{T}}\Sigma^{-1}z(k) \\ &+ \epsilon(k)^{\mathsf{T}}\Sigma^{-1}\epsilon(k)], \qquad (20)\end{aligned}$$

where the equality follows from the fact that $z(k)$ is independent of $\epsilon(k)$ and $\mathbb{E}[z(k)] = 0$. By cyclic property of trace operator, we note that $\mathbb{E}[z(k)^{\mathsf{T}}\Sigma^{-1}z(k)] = \mathrm{Tr}\left(\Sigma^{-1}\Sigma\right) = m$. For simplifying the second term of (20) observe that

$$\begin{aligned}\mathbb{E}\left[\delta_{a_{0:k}}\delta_{a_{0:k}}^{\mathsf{T}}\right] &= \begin{bmatrix}\mathbb{E}[\delta(a_0)\delta(a_0)^{\mathsf{T}}] & \dots & \mathbb{E}[\delta(a_0)\delta(a_k)^{\mathsf{T}}] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[\delta(a_k)\delta(a_0)^{\mathsf{T}}] & \dots & \mathbb{E}[\delta(a_k)\delta(a_k)^{\mathsf{T}}]\end{bmatrix} \\ &= \begin{bmatrix}\mathrm{diag}(p) & pp^{\mathsf{T}} & \dots & pp^{\mathsf{T}} \\ pp^{\mathsf{T}} & \mathrm{diag}(p) & \dots & pp^{\mathsf{T}} \\ \vdots & \vdots & \ddots & \vdots \\ pp^{\mathsf{T}} & pp^{\mathsf{T}} & \dots & \mathrm{diag}(p)\end{bmatrix}, \qquad (21)\end{aligned}$$

where the second equality follows because $\{a_k\}_{k=0}^{\infty}$ is i.i.d and $\mathbb{E}[\delta(a_k)] = p$, for all $k \in \mathbb{N}$. Further, $p = [p_1, \dots, p_N]^{\mathsf{T}}$ and, $\mathrm{diag}(p)$ is the diagonal matrix where the diagonal entries are elements of p. Now consider, the following:

$$\begin{aligned}\mathbb{E}[\epsilon(k)^{\mathsf{T}}\Sigma^{-1}\epsilon(k)] &= \mathrm{Tr}\left(\Sigma^{-1}\mathbb{E}[\epsilon(k)\epsilon(k)^{\mathsf{T}}]\right) \\ &= \mathrm{Tr}\left(\Sigma^{-1}\mathbb{E}[\mathcal{E}_k\delta(a_{0:k})\delta(a_{0:k})^{\mathsf{T}}\mathcal{E}_k^{\mathsf{T}}]\right) \\ &= \mathrm{Tr}\left(\mathcal{E}_k^{\mathsf{T}}\Sigma^{-1}\mathcal{E}_k\mathbb{E}[\delta_{a_{0:k}}\delta_{a_{0:k}}^{\mathsf{T}}]\right).\end{aligned}$$

where, the second equality follows from3.1. By invoking (12) and (21), the above expression can be further simplified as

$$\begin{aligned}\mathbb{E}[\epsilon(k)^{\mathsf{T}}\Sigma^{-1}\epsilon(k)] &= \mathrm{Tr}\left(\overline{L}_k\mathrm{diag}(p)\right) + \mathrm{Tr}\left(\widehat{L}_k pp^{\mathsf{T}}\right) \\ &= p^{\mathsf{T}}\mathrm{diag}(\overline{L}_k) + p^{\mathsf{T}}\widehat{L}_k p \qquad (22)\end{aligned}$$

Substituting $\mathbb{E}[z(k)^{\mathsf{T}}\Sigma^{-1}z(k)] = m$ and (22) in (20), the statement of the lemma follows. ■

Proof of Lemma 3.5: From (14) we note the following:

$$\lim_{k\to\infty} \overline{P}^D(k) = \frac{m + \lim\limits_{k\to\infty} \mathrm{diag}(\overline{L}_k)^{\mathsf{T}}p + \lim\limits_{k\to\infty} p^{\mathsf{T}}\widehat{L}_k p}{\tau} \qquad (23)$$

Under Assumption 3.4 and from (13) it follows that

$$\begin{aligned}\overline{L}_k &= \sum_{j=0}^{k-1} \mathcal{B}(j)^{\mathsf{T}}(\mathcal{A}^j)^{\mathsf{T}}C^{\mathsf{T}}\Sigma^{-1}C\mathcal{A}^j\mathcal{B}(j) + \Psi(k)^{\mathsf{T}}\Sigma^{-1}\Psi(k), \\ &= \sum_{j=0}^{k-1} \mathcal{B}^{\mathsf{T}}(\mathcal{A}^j)^{\mathsf{T}}C^{\mathsf{T}}\Sigma^{-1}C\mathcal{A}^j\mathcal{B} + \Psi^{\mathsf{T}}\Sigma^{-1}\Psi,\end{aligned}$$

Assuming that $\mathcal{A} = A(I - KC)$ is stable, it is easy to see that $\lim\limits_{k\to\infty} \sum_{j=0}^{k-1}(\mathcal{A}^j)^{\mathsf{T}}C^{\mathsf{T}}\Sigma^{-1}C\mathcal{A}^j$ exists. Thus

$$\overline{L}_\infty \triangleq \lim_{k\to\infty} \overline{L}_k = \mathcal{B}^{\mathsf{T}}\mathcal{O}\mathcal{B} + \Psi^{\mathsf{T}}\Sigma^{-1}\Psi. \qquad (24)$$

By letting $E_k \triangleq \sum_{j=0}^{k-1} C\mathcal{A}^j\mathcal{B}$ we have

$$\lim_{k\to\infty} E_k = \lim_{k\to\infty} \sum_{j=0}^{k-1} C\mathcal{A}^j\mathcal{B} = C(I - \mathcal{A})^{-1}\mathcal{B},$$

where the last equality follows because $\mathcal{A}$ is a stable matrix. Moreover, a straightforward computation results in $\widehat{L}(k) = E_k^{\mathsf{T}}E_k - \overline{L}_k$. By taking limits on both sides we have

$$\begin{aligned}\overline{L}_\infty \triangleq \lim_{k\to\infty} \widehat{L}_k &= \lim_{k\to\infty} \left(E_k^{\mathsf{T}}E_k - \overline{L}_k\right) \\ &= \mathcal{B}^{\mathsf{T}}\underbrace{(I - \mathcal{A})^{-\mathsf{T}}C^{\mathsf{T}}C(I - \mathcal{A})^{-1}}_{\triangleq \mathcal{M}}\mathcal{B} - \overline{L}_\infty \\ &= \mathcal{B}^{\mathsf{T}}\left[\mathcal{O} - \mathcal{M}\right]\mathcal{B} - \Psi\Sigma^{-1}\Psi. \qquad (25)\end{aligned}$$

By substituting (24), and (25) in (23) it follows that

$$\overline{P}_\infty^D \triangleq \lim_{k\to\infty} \overline{P}^D(k) = \frac{m + \mathrm{diag}(\overline{L}_\infty)^{\mathsf{T}}p + p^{\mathsf{T}}\widehat{L}_\infty p}{\tau}. ■$$

Proof of Proposition 4.1: Recall that $x^e(k) = x(k) + \gamma(k)$, $\mathbb{E}[x(0)] = 0$, and $x(k)$ is independent of $\gamma(k)$. Hence,

$$\begin{aligned}\mathbb{E}\left[\sum_{k=0}^{T-1} x^e(k+1)x^e(k+1)\right] = \sum_{k=0}^{T-1} &\mathbb{E}\left[x(k+1)x(k+1)\right] \\ &+ \mathbb{E}\left[\gamma(k+1)\gamma(k+1)\right]\end{aligned}$$

As the first term does not depend on the optimization variable p, for purpose of optimization, we can treat it as a constant. Instead, from (6) and (15) it follows that

$$\sum_{k=0}^{T-1} \gamma(k+1)^{\mathsf{T}}\gamma(k+1) = \delta(a_{0:T-1})^{\mathsf{T}}\mathcal{H}_{T-1}^{\mathsf{T}}\mathcal{H}_{T-1}\delta(a_{0:T-1}).$$

Taking the expectation on both sides of the above equation and following the same procedure as we did in proof of Lemma 3.3 (for analyzing $\mathbb{E}[\epsilon(k)^{\mathsf{T}}\Sigma^{-1}\epsilon(k)]$), the statement of the proposition follows. ■