52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Stochastic Game Approach for Replay Attack Detection

Fei Miao Miroslav Pajic George J. Pappas.

Abstract— **The existing tradeoff between control system performance and the detection rate for replay attacks highlights the need to provide an optimal control policy that balances the security overhead with control cost. We employ a finite horizon, zero-sum, nonstationary stochastic game approach to minimize the worst-case control and detection cost, and obtain an optimal control policy for switching between control-cost optimal (but nonsecure) and secure (but cost-suboptimal) controllers in presence of replay attacks. To formulate the game, we quantify game parameters using knowledge of the system dynamics, controller design and utilized statistical detector. We show that the optimal strategy for the system exists, and present a suboptimal algorithm used to calculate the system's strategy by combining robust game techniques and a finite horizon stationary stochastic game algorithm. Our approach can be generalized for any system with multiple finite cost, time-invariant linear controllers/estimators/intrusion detectors.**

I. Introduction

Cyber Physical Systems (CPS) feature tight integration of embedded computation, networks, and controlled physical processes [1]. This interaction between continuous physical dynamics, and discrete communication and computation substrates have made CPS vulnerable to malicious attacks beyond the standard cyber attacks [2]. Successful attacks on CPS could hamper the critical infrastructure with undesired consequences. For example, the Maroochy Water incident and the response discussed in [3] raised attention to security challenges and requirements for secure CPS [2], [1].

Several attack models, including physical attack, data deception attack, data denial-of-service attack (DoS), zero dynamics attack, and replay attack are analyzed in [4]. Most of these attacks assume some knowledge of the system's dynamics. On the other hand, even without any information about the system (including the controller's design), with replay attacks the attacker can record and resend delayed sensor measurements to the controller, causing deterioration in control performance (or even unstable behavior). Mo and Sinopoli designed a countermeasure method to increase the detection rate for replay attacks, which are undetectable by some statistical detector, like χ^2 detector [5]. By adding a Gaussian signal to the optimal control input, the control performance is sacrificed to increase the attack detection rate. Consequently, there is a need to provide a control strategy that balances the control and security requirements.

Game theory application for security has raised a lot of interest in recent years. The survey [6] provides a selected set of works that use game-theoretic approaches in computer networks security and privacy problems. Zhu *et al.* define a zero-sum stochastic game for the design of an Intrusion Detection System (IDS) and provide the stationary optimal strategy [7]. Robust algorithms against well-defined uncertainties are presented in [8]. A game theoretic formulation for minimax or robust estimation in the presence of faults, and the existence conditions for optimal solutions are discussed in [9]. Yet, there exist many challenges in applying these approaches for design of secure CPS. As stated in [6], one problem with game theoretic techniques in security modeling is the difficulty of quantifying security game parameters.

In this work we design a zero-sum, finite horizon, nonstationary stochastic game for minimizing the worst-case control and detection cost, and obtain an optimal control policy for switching between control-cost optimal (but nonsecure) and secure (but cost-suboptimal) controllers in presence of replay attacks. Therefore, in a system dynamic fashion we balance the control performance and security overhead for replay attack detection. To achieve this, we quantify game parameters using the knowledge of the system dynamics, controller design and utilized statistical detector. Our problem formulation satisfies the value existence conditions for nonstationary games [10], and we show that the optimal strategy for the system exists. A suboptimal algorithm is also developed, based on robust game techniques [11] and the finite horizon stationary stochastic game algorithm from [12].

The presented approach does not have to be constrained to this scenario (i.e., replay attacks, two controllers); it is possible to generalize our analysis, especially the quantification approach, for any system with multiple finite cost, time-invariant linear controllers/estimators/failure or intrusion detectors, facing different attacks. In this case, the system will choose from the library of control, estimation or detection methods under the uncertainty of attacker's particular choice.

This paper is organized as follows. In Section II, we present the system and replay attack model, and describe the control problem. In Section III, we formulate and quantify the nonstationary, zer-sum stochastic game between the system and the attacker. The existence of the system's optimal strategy is shown, followed by a suboptimal value iterative algorithm in Section IV. On several examples, in Section V we illustrate performance of the derived attack detection scheme. Finally, Section VI provides concluding remarks.

This material is based on research sponsored by DARPA under agreement number FA8750-12-2-0247. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

F. Miao, M. Pajic and G. J. Pappas are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA 19014. Email: {miaofei, pajic, pappasg}@seas.upenn.edu

II. Control System and Replay Attack Model

Before presenting the game formulation, we first introduce the system and attack models. We consider the setup from Figure 1, where a Linear Time-Invariant (LTI) plant is controlled by an LQR controller with a Kalman filter (acting as a state estimator), and a χ^2 detector that is used to detect any abnormal behavior. After we describe each component of the system, along with a model of replay attacks, we introduce the problem of balancing the control performance and detection rate.

LTI Plant: We consider LTI plants described as:

$$\begin{aligned} \mathbf{x}_{k+1} =& \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k, \\ \mathbf{y}_k =& \mathbf{C}\mathbf{x}_k + \mathbf{v}_k, \end{aligned} \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n, \mathbf{u}_k \in \mathbb{R}^p$ and $\mathbf{y}_k \in \mathbb{R}^m$ denote the plant's state, input and output vectors respectively, and $\mathbf{w}_k$ and $\mathbf{v}_k$ denote process and measurement noise at time k. We assume that $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q})$, $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R})$, and $\mathbf{x}_0 \sim \mathcal{N}(\bar{\mathbf{x}}_0, \Sigma)$ are independent and identically distributed (IID) Gaussian random variables, and that the plant is detectable.

Kalman Filter: We assume the Kalman filter is already in steady state, with parameter $\mathbf{K}$ and initial condition $\mathbf{P} = \Sigma$. Thus, the filter acts as a fixed-gain estimator of the form:

$$\begin{aligned} \hat{\mathbf{x}}_{0|-1} &= \bar{\mathbf{x}}_0, \mathbf{P} = \Sigma, \mathbf{K} = \mathbf{P}\mathbf{C}^T(\mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{R})^{-1}, \\ \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_{k|k-1}), \\ \hat{\mathbf{x}}_{k+1|k} &= \mathbf{A}\hat{\mathbf{x}}_{k|k} + \mathbf{B}\mathbf{u}_k. \end{aligned} \tag{2}$$

Optimal LQG Controller: Using the state estimation $\hat{\mathbf{x}}_{k|k}$, the controller acts as a fixed gain compensator of the form $\mathbf{u} = \mathbf{u}_k^* = \mathbf{L}\hat{\mathbf{x}}_{k|k}$, where $L \triangleq -(\mathbf{B}^T\mathbf{S}\mathbf{B}+\mathbf{U})^{-1}\mathbf{B}^T\mathbf{S}\mathbf{A}$, and S is the solution of the Riccati equation:

$$\mathbf{S} = \mathbf{A}^T\mathbf{S}\mathbf{A} + \mathbf{W} - \mathbf{A}^T\mathbf{S}\mathbf{B}(\mathbf{B}^T\mathbf{S}\mathbf{B} + \mathbf{U})^{-1}\mathbf{B}^T\mathbf{S}\mathbf{A}, \tag{3}$$

for some $\mathbf{W}, \mathbf{U} \succ 0$. Thus, $\mathbf{u}_k^*$ satisfies:

$$\mathbf{u}_k^* = \arg\min_{\mathbf{u}_k} \lim_{T\to\infty} \mathbb{E}\frac{1}{T}[\sum_{k=0}^{T-1}(\mathbf{x}_k^T\mathbf{W}\mathbf{x}_k + \mathbf{u}_k^T\mathbf{U}\mathbf{u}_k)] \tag{4}$$

χ^2 Detector: For the described system, the Kalman filter residues $\mathbf{z}_i = \mathbf{y}_i - \mathbf{C}\hat{\mathbf{x}}_{i|i-1}$ are IID with Gaussian distribution $\mathcal{N}(0, \mathcal{P})$, where $\mathcal{P} = \mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{R}$. Therefore, at each time k, χ^2 detector takes the form:

$$g_k = \sum_{i=k-\tau+1}^{k} (\mathbf{y}_i - \mathbf{C}\hat{\mathbf{x}}_{i|i-1})^T\mathcal{P}^{-1}(\mathbf{y}_i - \mathbf{C}\hat{\mathbf{x}}_{i|i-1}) \lessgtr \alpha. \tag{5}$$

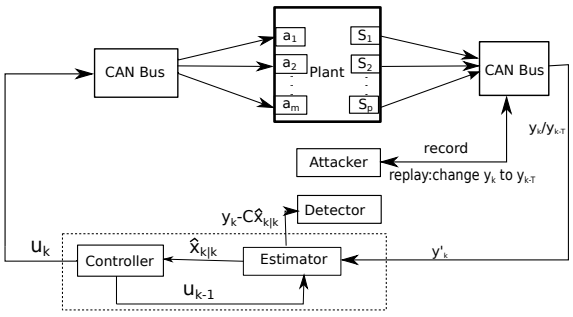


Fig. 1. System diagram with only one controller, where replay attacks can compromise sensor measurements delivered to the controller.

Here, τ is the detection window size and α is the alarm triggering threshold. Both parameters are predefined to provide a desired false alarm rate according to the distribution of g_k.

Attacker Model for Replay Attacks: We assume the attacker can record sensor measurements, choose the replay window size T, and at each time-step decide whether to send the correct or delayed plant outputs.[1] Thus, data $\mathbf{y}'_k$ $(k \geq 0)$ received by the estimator and detector can be described as:

$$\mathbf{y}'_k = \begin{cases} \mathbf{y}_k, & \text{sensor output is not changed at } k \\ \mathbf{y}_{k-T}, T > 0, & \text{replay attack occurs at } k \end{cases} \tag{6}$$

If $(A + BL)(I - KC) \triangleq \mathcal{A}$ is stable, the χ^2 detector is useless, since the expectation of detector statistics for the compromised residues $\mathbf{z}'_k = \mathbf{y}_{k-T} - \mathbf{C}\hat{\mathbf{x}}_{k|k-1}$ will converge to the same value as that for $\mathbf{z}_k = \mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_{k|k-1}$ [5]

$$\lim_{k\to\infty} \mathbb{E}[(\mathbf{z}'_k)^T\mathcal{P}^{-1}\mathbf{z}'_k] = \lim_{k\to\infty} \mathbb{E}[(\mathbf{z}_k)^T\mathcal{P}^{-1}\mathbf{z}_k] = m \tag{7}$$

To increase the detection rate, an IID Gaussian signal $\Delta\mathbf{u}_k \sim \mathcal{N}(0, \mathcal{L})$ can be added to the optimal controller

$$\mathbf{u}_k = \mathbf{u}_k^* + \Delta\mathbf{u}_k. \tag{8}$$

With this $\mathbf{u}_k$, (7) for $\mathbf{z}'_k$ will be $m + 2trace(\mathbf{C}^T\mathcal{P}^{-1}\mathbf{C}\mathcal{U})$, where $\mathcal{U}$ satisfies $\mathcal{U} - \mathbf{B}\mathcal{L}\mathbf{B}^T = \mathcal{A}\mathcal{U}\mathcal{A}^T$. On the other hand, applying the additional input Δu would increase the quadratic cost to $J' = J + trace[(\mathbf{U} + \mathbf{B}^T\mathbf{S}\mathbf{B})\mathcal{L}]$, where J denotes the optimal cost (for control input $\mathbf{u}_k^*$ from (4)) [5].

The above analysis of control performance and detection rate is over infinite time horizon. To balance the security overhead and the control cost, we consider applying different controllers according to the system dynamics in finite time. Consequently, our goal is to design an optimal policy for switching between the (cost) optimal controller (i.e., $\mathbf{u}_k^*$) and the controller from (8) which allows for the detection of replay attacks, as shown in Figure 2. To achieve this, we frame the problem as a non-cooperative stochastic game between the system and the attacker, since the system has limited knowledge about the attacker's decisions, while the detector only provides a probabilistic detection rate of malicious behavior.

III. Stochastic Game Formulation of Replay Detection

To obtain a switching control policy that minimizes the expected cost for the considered closed-loop system, we formulate a zero-sum finite horizon stochastic game between the

[1] It is worth noting here that when a replay attack occurs, values from all sensors can be compromised.

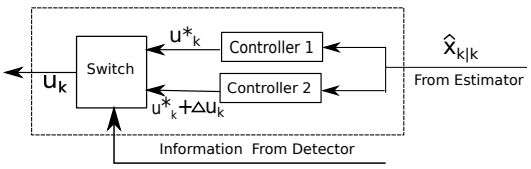


Fig. 2. Block diagram of the switching controller – Controller 1 is cost optimal, while Controller 2 provides a higher attack detection rate.

system and the attacker. We consider K-stage games, where each stage corresponds to n time-steps of the dynamical model from (1). For simplicity, in this section we present the case for $n = 1$, i.e., every game stage is also one time step of the physical model. We denote the attacker as the maximizer (the row player) and the system as the minimizer (the column player) in the game. While each player is able to observe the state of the game, neither player has exact information about the other player's previous behavior.

We assume that once the detector triggers the alarm, the system stops its execution to check for malicious behavior. In this case, if the attacker has been active it will be detected and the system will remain safe, using only the optimal controller in the future (i.e., the system wins). Otherwise, the system pays a significantly large penalty for triggering the false alarm.

Figure 3 illustrates the stochastic game model. The game state space S is stationary. At each stage k, the nonstationary parameters include action spaces for the attacker (A_{tk}) and system (A_{sk}), the state transition probability matrix $\mathbb{P}_k$, and the immediate payoff matrix r_k. We consider mixed strategies $\mathbf{F}_k, \mathbf{G}_k$ for both players. Formally, we define the game as a sequence of tuples $(S, A_{tk}, A_{sk}, \mathbf{F}_k, \mathbf{G}_k, \mathbb{P}_k, r_k)$, $k \in \{1, \cdots, K\}$. Every parameter is quantified using the system and attack model described in Section II.

*1) **Game State Space**:* $S = \{s_1, s_2, s_3\}$ denotes the set of the stochastic game states. Absorbing state $s_1 = safe$ describes that the system has already successfully detected a replay attack; $s_2 = no\ detection$ specifies that the alarm has not been triggered; finally, the system enters the state $s_3 = false\ alarm\ trigger$ when the alarm is triggered before attacks occur.

*2) **Attacker's Action Space**:* At each stage k, the attacker has M candidate actions, i.e., $A_{tk} = \{a_{1k}, \cdots, a_{Mk}\} = \{\mathbf{y}_k, \mathbf{y}_{k-t_2}, \cdots, \mathbf{y}_{k-t_M}\}$ is the attacker's action space, where y_k is the real sensor value and $\mathbf{y}_{k-t_i}$ denotes the replay attack with a window size t_i.

*3) **System's Action Space**:* $A_{sk} = \{u_{1k}, u_{2k}\} = \{\mathbf{u}_k^*, \mathbf{u}_k^* + \Delta\mathbf{u}_k\}$ is the system's action space facing a replay attack at stage k. Here, $u_{1k} = \mathbf{u}_k^*$ is the cost optimal input, while $u_{2k} = \mathbf{u}_k^* + \Delta\mathbf{u}_k$ provides a higher detection rate.[2]

*4) **Mixed Strategy**:* Let $f_k^i(s)$ and $g_k^j(s)$ be the probabilities that at stage k and state s the attacker and the system chose actions $a_{ik} \in A_{tk}$ and $u_{jk} \in A_{sk}$, respectively. We define $\mathbf{F}_k$ and $\mathbf{G}_k$ as the sets of strategies of the attacker and the system at stage k:

$$\mathbf{F}_k := \{\mathbf{f}_k = [\mathbf{f}_k(s_1), \mathbf{f}_k(s_2), \mathbf{f}_k(s_3)] |\ \mathbf{f}_k(s) \in \mathbb{R}^M, \forall s \in S, f_k^i(s) \geq 0, \forall a_{ik} \in A_{tk}, \sum_i f_k^i(s) = 1\},$$

$$\mathbf{G}_k := \{\mathbf{g}_k = [\mathbf{g}_k(s_1), \mathbf{g}_k(s_2), \mathbf{g}_k(s_3)] |\ \mathbf{g}^k(s) \in \mathbb{R}^2, \forall s \in S, g_k^j(s) \geq 0, \forall u_{jk} \in A_{sk}, \sum_j g_k^j(s) = 1\},$$

We denote with $H_k = H_1\mathbf{F}_1\mathbf{G}_1 \cdots \mathbf{F}_{k-1}\mathbf{G}_{k-1}$ the concatenation of the strategies until stage k, where $H_1 \subseteq S$ is the initial state set, and each strategy history $h_k \in H_k$ can be described as $h_k = h_1\mathbf{f}_1\mathbf{g}_1 \cdots \mathbf{f}_{k-1}\mathbf{g}_{k-1}$ for $h_1 = s_l$.

*5) **State Transition Probability**:* $\mathbb{P}_k$ is the state transition probability set at k, where $\tilde{P}_k \in \mathbb{P}_k$ satisfies:

$$\tilde{P}_k(s'|h_k, s) = [\tilde{P}_k^{ij}(s'|h_k, s) \geq 0] \in \mathbb{R}^{M\times 2}, s', s \in S,$$
$$\sum_{s'\in S} \tilde{P}_k^{ij}(s'|h_k, s) = 1, \forall (a_{ik}, u_{jk}) \in A_{tk} \times A_{sk}, s \in S.$$

Here, $\tilde{P}_k^{ij}(s'|h_k, s)$ is the probability provided by the detector (like χ^2 detector), that the system transits from state s at stage k to state s' at stage $k+1$, given a history $h_k \in H_k$ and both players' actions (a_{ik}, u_{jk}).[3]

*6) **Immediate Payoff Function**:* We define the immediate payoff matrix set at stage k as $r_k \subseteq \mathbb{R}^{M\times 2}$, and $\tilde{r}_k \in r_k$, where $\tilde{r}_k(h_k, s) = [\tilde{r}_k^{ij}(h_k, s) \geq 0]$, denotes the payoff of every action pair (a_{ik}, u_{jk}) for strategy history h_k and state s. Given system state $\mathbf{x}_k(h_k)$ and control input $\gamma_k(h_k, a_{ik}, u_{jk})$, similar to the LQG cost we have $\tilde{r}_k^{ij}(h_k, s_l) = \tilde{r}_{tk}^{ij}(h_k, s_l)$(attacker) $= -\tilde{r}_{sk}^{ij}(h_k, s_l)$(system):

$$\tilde{r}_k^{ij}(h_k, s_1) = \mathbf{x}_k^T(h_k)\mathbf{W}\mathbf{x}_k(h_k) + \gamma_k^T(h_k, a_{1k}, u_{jk})\mathbf{U}\gamma_k(h_k, a_{1k}, u_{jk}),$$
$$\tilde{r}_k^{ij}(h_k, s_2) = \mathbf{x}_k^T(h_k)\mathbf{W}\mathbf{x}_k(h_k) + \gamma_k^T(h_k, a_{ik}, u_{jk})\mathbf{U}\gamma_k(h_k, a_{ik}, u_{jk}),$$
$$\tilde{r}_k^{ij}(\hat{h}_k, s_3) = false\ alarm\ trigger\ penalty.$$

Note that at state s_1 the system wins, so the payoff is determined by real sensor data $a_{1k} = \mathbf{y}_k$.

*7) **Control System Dynamics**:* With all the above definition, for any strategy history h_k, given the system and attack models from Section II, we can obtain the expected system behavior under the stochastic game formulation (for simplicity we omit the $\mathbb{E}$), which is used to compute the expected payoff for different strategies (defined in the next section). Given initial $\hat{\mathbf{x}}_{1|0}(h_1) = \bar{\mathbf{x}}_0$, $\mathbf{x}_1(h_1) = \mathbf{x}_0$, at each stage k the *LTI plant* and *Kalman Filter* evolve as:

$$\mathbf{x}_k(h_k) = \mathbf{A}\mathbf{x}_{k-1}(h_{k-1}) + \mathbf{B}\mathbf{u}_{k-1}(h_k) + \mathbf{w}_{k-1},$$
$$a_{1k}(h_k) = \mathbf{y}_k(h_k), \qquad a_{ik}(h_k) = \mathbf{y}_{k-t_i}(h_k), i \neq 1,$$
$$\hat{\mathbf{x}}_{k|k}(h_k, a_{ik}) = \hat{\mathbf{x}}_{k|k-1}(h_k) + \mathbf{K}(a_{ik}(h_k) - \mathbf{C}\hat{\mathbf{x}}_{k|k-1}(h_k)),$$
$$\hat{\mathbf{x}}_{k+1|k}(h_k, a_{ik}, u_{jk}) = \mathbf{A}\hat{\mathbf{x}}_{k|k}(h_k, a_{ik}) + \mathbf{B}\gamma_k(h_k, a_{ik}, u_{jk}).$$

To specify behavior of the χ^2 *Detector* we define residues

$$\mathbf{z}_{k+1}(h_k, a_{ik}, u_{jk}) = a_{ik}(h_k) - \mathbf{C}\hat{\mathbf{x}}_{k+1|k}(h_k, a_{ik}, u_{jk}).$$

[2]Note that if $N > 2$ controllers/estimators/detectors are used, the system's action space should be of size N and the method presented in this paper could still be used.

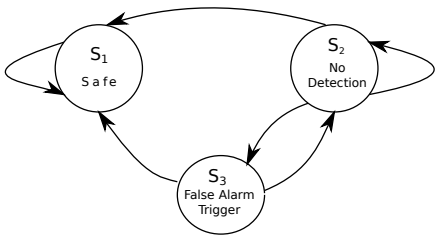


Fig. 3. Stochastic Game Model, s_1 is an absorbing state.

[3]Full description of the game formulation can be found at `https://sites.google.com/site/miaofeiatpenn/publications`

Then g_{k+1} used to extract the state transition probability is

$$g_{k+1}(h_k, a_{ik}, u_{jk}) = \sum_{t=k-\tau+2}^{k} [\mathbf{z}_t(h_t)]^T \mathcal{P}^{-1} \mathbf{z}_t(h_t) + [\mathbf{z}_{k+1}(h_k, a_{ik}, u_{jk})]^T \mathcal{P}^{-1} \mathbf{z}_{k+1}(h_k, a_{ik}, u_{jk}) \quad (9)$$

Updating the Model With Strategy At Stage k: If the strategies at stage k are $\mathbf{f}_k$ and $\mathbf{g}_k$, then $h_{k+1} = h_k \mathbf{f}_k \mathbf{g}_k$. Furthermore, let $p(s_k^{l'})$ be the probability that the system is at state $s_{l'}$ at stage k (note that $p(s_1^{l'})$ is given). Then we update the following parameters for stage $k+1$:

$$\mathbf{u}_k(h_{k+1}) = \sum_{j=1}^{2} \sum_{i=1}^{M} \sum_{l=1}^{3} p(s_k^l) f_k^i(s_l) g_k^j(s_l) \gamma_k(h_k, a_{ik}, u_{jk}),$$

$$\mathbf{y}_k(h_{k+1}) = \sum_{i=1}^{M} \sum_{l=1}^{3} p(s_k^l) f_k^i(s_l) a_{ik}(h_k),$$

and the probability of the system being at state $s_{l'}$ at stage $k+1$ is:

$$p(s_{k+1}^{l'}) = \sum_{l=1}^{3} p(s_k^l) [\mathbf{f}_k(s_l)]^T \tilde{P}_k(s_{l'}|h_k, s_l) \mathbf{g}_k(s_l).$$

Therefore, when we involve the system dynamics to define the game, the resulting formulation utilizes a nonstationary immediate payoff matrix and a transition probability matrix.

IV. Existence of An Optimal Strategy and Suboptimal Algorithm

Based on the game formulation, in this section we discuss the existence of an optimal solution for the system, and present an algorithm to compute a suboptimal system strategy.

A. Existence of the System's Optimal Strategy

We define the concatenation of strategies for K-stage game of each player ($\mathbf{f}$ for attacker and $\mathbf{g}$ for system) as

$$\mathbf{f} = \mathbf{f}_1 \cdots \mathbf{f}_K, \mathbf{f}_k \in \mathbf{F}_k, \quad \mathbf{g} = \mathbf{g}_1 \cdots \mathbf{g}_K, \mathbf{g}_k \in \mathbf{G}_k.$$

Let the random variable ζ_k describe the state of the game at stage k, and let us define the conditional expected total payoff till $\tilde{K}$ for any $\mathbf{f}, \mathbf{g}$, given initial state $\zeta_1 = s$ as

$$R_{\tilde{K}}(s, \mathbf{f}, \mathbf{g}) = \sum_{k=1}^{\tilde{K}} \sum_{l=1}^{3} p(\zeta_k = s_l | \zeta_1 = s) [\mathbf{f}_k(s_l)]^T \tilde{r}_k(h_k, s_l) \mathbf{g}_k(s_l).$$

Since the immediate payoff satisfies $0 \leq \tilde{r}_k^{ij}(h_k, s_l) < \infty$, we have that $R_{\tilde{K}}(s, \mathbf{f}, \mathbf{g})$ is a nonnegative real-valued, nondecreasing function with $\tilde{K}$. Furthermore, for finite K

$$R_K(s, \mathbf{f}, \mathbf{g}) < \infty, \forall s, \mathbf{f}, \mathbf{g}. \quad (10)$$

Definition 1 ([10]): A two-person zero-sum K-stage stochastic game is said to have a value vector v_K^* if

$$v_{K,s}^* = \underline{v}_{K,s} = \bar{v}_{K,s}, \text{ for any } s \in S, \quad \text{where}$$

$$\underline{v}_{K,s} = \sup_{\mathbf{f}} \inf_{\mathbf{g}} R_K(s, \mathbf{f}, \mathbf{g}), \quad \bar{v}_{K,s} = \inf_{\mathbf{g}} \sup_{\mathbf{f}} R_K(s, \mathbf{f}, \mathbf{g}).$$

For the finite value K-stage stochastic game, strategies $\mathbf{g}^*$ and $\mathbf{f}^*$ are called optimal for player two (the system) and player one (the attacker), respectively, if for all $s \in S$

$$v_{K,s}^* = \sup_{\mathbf{f}} R_K(s, \mathbf{f}, \mathbf{g}^*), \quad v_{K,s}^* = \inf_{\mathbf{g}} R_K(s, \mathbf{f}^*, \mathbf{g}). \quad \square$$

The existence conditions of the value and optimal strategies for a general finite horizon zero-sum nonstationary stochastic game are shown in [10]. The game defined in this paper has a finite state space, finite action spaces, and satisfies (10). Therefore, using the same approach as in the proofs from [10] we can prove the following theorem:

Theorem 1: There exists the value of the considered game and an optimal strategy for the system. $\square$

B. Suboptimal Algorithm For the Nonstationary Game

Existing value iterative algorithms for stationary stochastic games can not be used to solve our game, since the nonstationary game parameters depend on the previous history, which is only available in the future algorithm iterations. Hence, we design a suboptimal algorithm based on the value iteration method for finite horizon stationary stochastic game from [12] and robust game techniques from [11]. Algorithm 1 provides an upper bound for the game value and the corresponding nonstationary suboptimal strategy for the system. The idea is to solve a robust game at each iteration step – i.e., minimize the worst-case caused by extreme points of the nonstationary payoff and state transition probability polyhedra, or $(r_k, \mathbb{P}_k)$, defined for all possible histories.

The value iteration algorithm for finite horizon stationary stochastic game (with fixed payoff r and state transition probability P at every stage) works in the way that if a player knew how to play in the game optimally from the next stage on, then, at the current stage, he would play with such strategies [12]. The value of K-stage game is finally provided by the last step of iteration. Similarly, the Algorithm 1 of the nonstationary stochastic game starts from the last stage, gets the matrix game value and the optimal strategy related to nonstationary $(r_k, \mathbb{P}_k)$ at each stage, and returns an upper bound for the value of the total payoff in K-stages. To estimate the values at each step, we consider the payoff and state transition probability set $(r_k, \mathbb{P}_k)$ as an uncertain parameter set for the one shot robust game [11].

To quantify the bounded polyhedra $(r_k, \mathbb{P}_k)$, we need the expected system dynamics $\mathbf{x}_k, \mathbf{u}_k, \mathbf{y}_k, k = 1, \cdots, K$ defined in Section III-.7, which is determined by the strategy history. The extreme points for the uncertain set $(r_k, \mathbb{P}_k)$ depend on pure strategy histories. Let $H_k^p \subset H_k$ be the concatenation of pure strategies until stage k, where a pure history $h_k^p \in H_k^p$, $h_k^p = s_1 \mathbf{f}_1^p \mathbf{g}_1^p \cdots \mathbf{f}_{k-1}^p \mathbf{g}_{k-1}^p$ satisfies that all $f_t^p(s), g_t^p(s)$ have only one non-zero element (i.e., they choose the corresponding action or the *pure* strategy). By using any h_k^p in the game model of Section III-.7, we get the corresponding $r_k^p(h_k^p, s_l)$ and $P_k^p(h_k^p, s_l)$ for stages 1 to K. Thus, the extreme points set $(r_{kp}, \mathbb{P}_{kp})$ for $(r_k, \mathbb{P}_k)$ is:

$$\{(r_k^p(h_k^p, s_l), P_k^p(h_k^p, s_l)) |\ h_k^p \in H_k^p, l \in \{1, 2, 3\}\}.$$

We define the pure strategy backup matrix set $\mathbb{Q}_{kp}$, $k = 1, \cdots, K-1$ as:

$$\mathbb{Q}_{kp} = \{Q_k^p(h_k^p, s_l) = r_k^p(h_k^p, s_l) + \sum_{s' \in S} P_k^p(s'|h_k^p, s_l) v_{k+1}^{s'}(h_{k+1}^p)\}, \quad (11)$$

where $v_{k+1}^{s'}(h_{k+1}^p) \geq 0$ is the robust game value resulting from the iteration at stage $k+1$ of Algorithm 1, and relates

to matrix games defined by $\mathbb{Q}_{(k+1)p}$. In addition, we define the backup matrix set $\mathbb{Q}_k$ for all possible h_k, as

$$\mathbb{Q}_k = \{\tilde{Q}_k(h_k, s_l) = \tilde{r}_k(h_k, s_l) + \sum_{s' \in S} \tilde{P}_k(s'|h_k, s_l) v_{k+1}^{s'}(h_{k+1}^p)\}.$$

Finally, for the stage K we define $Q_K^p(h_K^p, s_l) = r_K^p(h_K^p, s_l)$ and $\tilde{Q}_K(h_K, s_l) = \tilde{r}_K(h_K, s_l)$.

Now, consider the iteration for calculating $v_k^{s_l}(h_k^p)$ from all matrix games $Q_k^p(h_k^p, s_l) \in \mathbb{Q}_{kp}$ in Algorithm 1. We denote the non-pure history value of k as $v_k^{s_l}(h_{k-1}^p)$, which is calculated from the $\tilde{Q}_k(h_k, s_l)$ that have the same pure strategies as h_k^p in all stages $1, \cdots, k-2$, and any strategy at stage $k-1$. We use the following result to show that at every k, $v_k^{s_l}(h_k^p)$ is greater than or equal to $v_k^{s_l}(h_{k-1}^p)$.

Theorem 2: Consider the value iteration for stage k as a one shot robust game. Based on $v_{k+1}^{s'}(h_{k+1}^p) \geq 0$ of previous iteration, We define the robust game value obtained at k as

$$v_k^{s_l}(h_k^p) = \max_{Q_k^p(h_k^p, s_l) \in \mathbb{Q}_{kp}} v^*[Q_k^p(h_k^p, s_l)], \tag{12}$$

where v^* is the function that yields the value of a zero-sum matrix game. Then for $k = 2, \cdots, K$, $v_k^{s_l}(h_{k-1}^p)$ is upper bounded by $v_k^{s_l}(h_k^p)$ (i.e., $v_k^{s_l}(h_{k-1}^p) \leq v_k^{s_l}(h_k^p)$). □

Proof: Since $v_{k+1}^{s'}(h_{k+1}^p)$ is a nonnegative scalar, the extreme points of $\mathbb{Q}_k$ can only come from the extreme points of the tuple $(r_k, \mathbb{P}_k)$, i.e., by considering the matrix game value of every $Q_k^p(h_k^p, s_l)$ defined in (11), we will get the maximum value from extreme points of $\mathbb{Q}_k$. Now consider the following optimization problem for the system (14) and for any attacker's strategy vector $\mathbf{f}$

$$\min_{\mathbf{g}} \quad z \tag{13}$$

$$\text{subject to} \quad z \geq \max_{\tilde{Q}_k(h_k, s_l) \in \mathbb{Q}_k} \mathbf{f}^T[\tilde{Q}_k(h_k, s_l)]\mathbf{g}. \tag{14}$$

As proven by Lemma 5 in [11], (14) is equivalent to the following constraint that considers only the extreme points

$$z \geq \max_{Q_k^p(h_k^p, s_l) \in \mathbb{Q}_{kp}} \mathbf{f}^T[Q_k^p(h_k^p, s_l)]\mathbf{g}, \tag{15}$$

For the worst-case f, the above is also true. Hence, let

$$v_k^{s_l}(h_k^p) = \max_{Q_k^p(h_k^p, s_l) \in \mathbb{Q}_{kp}} \min_{\mathbf{g}} \max_{\mathbf{f}} \mathbf{f}^T[Q_k^p(h_k^p, s_l)]\mathbf{g}. \tag{16}$$

For optimal policies $\mathbf{f}^*(h_k^p, s_l)$ and $\mathbf{g}^*(h_k^p, s_l)$, the above optimization (16) results in cost $\max_{Q_k^p(h_k^p, s_l) \in \mathbb{Q}_{kp}} v^*[Q_k^p(h_k^p, s_l)]$. However, $(\mathbf{f}^*(h_k^p, s_l), \mathbf{g}^*(h_k^p, s_l))$ can be non-pure strategies, meaning that when used in III-.7, they will not result in extreme points of $\mathbb{Q}_{k+1}$. The non-pure history value $v_{k+1}^{s'}(h_k^p)$ of iteration for stage $k+1$ satisfies $v_{k+1}^{s'}(h_k^p) \leq v_{k+1}^{s'}(h_{k+1}^p)$. Replacing $v_{k+1}^{s'}(h_{k+1}^p)$ by $v_{k+1}^{s'}(h_k^p)$ in (11) will decrease every element in matrix Q_k^p, since $r_k^{ij} \geq 0$ and $P_k^{ij} \geq 0$. With a similar argument in the next iteration for stage $k-1$, we have $v_k^{s_l}(h_{k-1}^p) \leq v_k^{s_l}(h_k^p)$. ■

According to Theorem 2, we use Algorithm 1 to compute an upper bound of the value and the corresponding suboptimal strategy for every step. The function π computes the strategy and robust value as defined in (12). Note that for

Algorithm 1 : Suboptimal Algorithm for A Finite Nonstationary Stochastic Game

Input: System model parameters and game parameters.
Initialization: Compute the set of $(r_{kp}, \mathbb{P}_{kp})$ for $k = 1, ...K$. Compute the game at stage K: $Q_K^p(h_K^p, s_l) = r_K^p(h_K^p, s_l)$, $f^*(h_K^p, s_l), g^*(h_K^p, s_l), v_K^{s_l}(h_K^p) \leftarrow \pi(Q_K^p(h_K^p, s_l))$.
Iteration: For $k = (K-1), \cdots, 1$,
get the backup matrix set $\mathbb{Q}_{kp}$ for all $h_k^p \in H_k^p$, where each matrix is defined in (11), then calculate:
$f^*(h_k^p, s_l), g^*(h_k^p, s_l), v_k^{s_l}(h_k^p) \leftarrow \pi(Q_k^p(h_k^p, s_l))$.
$\mathbf{f}_k^* = [\mathbf{f}^*(h_k^p, s_l), l = 1, 2, 3]$, $\mathbf{g}_k^* = [\mathbf{g}^*(h_k^p, s_l), l = 1, 2, 3]$.
Return: the strategy concatenation pair $\mathbf{f}_a = \mathbf{f}_1^* \cdots \mathbf{f}_K^*, \mathbf{g}_a = \mathbf{g}_1^* \cdots \mathbf{g}_K^*$ and the value upper bound $v_1^{s_l}, l = 1, 2, 3$.

the replay attacks considered in this paper, at state s_1 the system wins the game and replay is harmless. Thus, the optimal strategies of s_1 is $\mathbf{g}_k(s_1) = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$, $\mathbf{f}_k(s_1) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$. The nonstationary game values $v_k^{s_l}(h_{k-1}^p)$ and $v_k^{s_l}(h_k^p)$ that result from value iteration of two strategies only differ at stage $k-1$ (i.e., same and pure from stages 1 to $k-2$). By value iteration backward to stage 1, we compare the game value (for all possible strategies) and the robust game value $v_1^{s_l}$ of Algorithm 1 in the following theorem.

Theorem 3: Algorithm 1 results in an upper bound $v_1^{s_l}$ for the value of the K-stage game, together with suboptimal strategies $\mathbf{f}_a$ and $\mathbf{g}_a$. □

Proof: The strategies $\mathbf{f}_a, \mathbf{g}_a$ of Algorithm 1 are possibly not pure. According to Theorem 2, $v_k^{s_l}(h_{k-1}^p) \leq v_k^{s_l}(h_k^p)$, and the proof holds for every $k = 2, \cdots, K$. Consider the value iteration for $k = 1$, with $v_2^{s_l}(h_1^p) \leq v_2^{s_l}(h_2^p)$,

$$Q_1^{ij}(h_1, s_l) = \tilde{r}_1^{ij}(h_1, s_l) + \sum_{s' \in S} \tilde{P}_1^{ij}(s'|h_1, s_l) v_2^{s'}(h_1^p) \leq Q_1^{ij}(h_2^p, s_l),$$

thus $v^*[Q_1(h_1, s_l)] \leq v_1^{s_l}$. Iterative value based on pure strategy back up matrix sets $\mathbb{Q}_{kp}, k = 1, \cdots, K$ obtained from Algorithm 1 is an upper bound for the game value. ■

V. Example

To illustrate our stochastic game approach, we consider the four input four output system examined in Section 3.1 of [13]. We assume that the attacker's action space (i.e., replay window size) contains $t_2 = 10, t_3 = 20, t_4 = 30, t_5 = 40$, and that the initial system state is s_2 (i.e., $p(s_2^1) =$

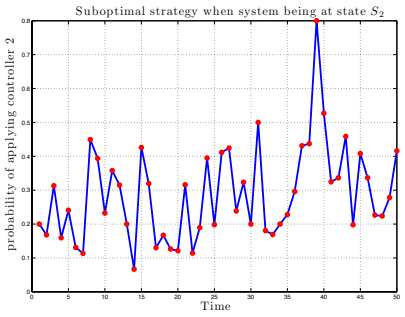
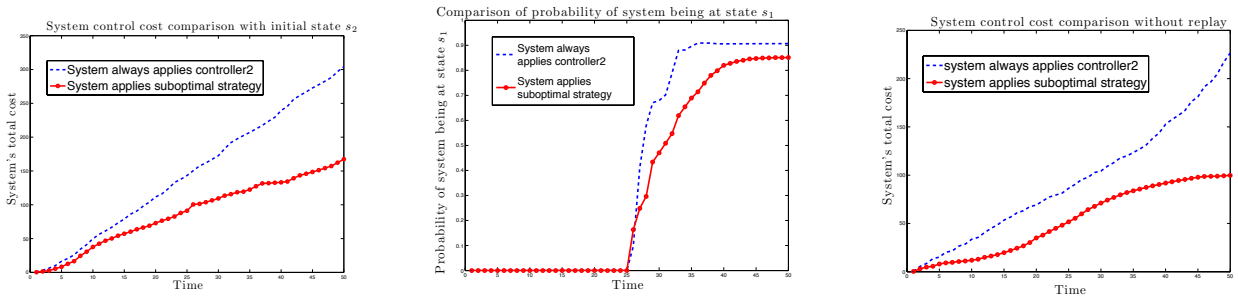


Fig. 4. System's suboptimal strategy at state s_2 – the probability of switching to Controller 2 at every stage.



(a) Control cost comparison when the system applies different strategies; initial state is s_2

(b) Probability of system being at state s_1 (safe, already successfully detected a replay attack).

(c) Control cost comparison when no replay occurs; initial state is s_2.

Fig. 5. Comparison on the system control cost and detection rate under different scenarios

1). Using Algorithm 1, we obtained a suboptimal system's strategy $\mathbf{g}_a$, for the $K = 50$-stage game. Figure 4 shows the probability of switching to Controller 2 at every step.

We compare the system's control cost in two cases, when the system applies the suboptimal game strategy $\mathbf{g}_a$ and when it only uses Controller 2 (i.e., always adds noise by applying input $\mathbf{u}_k^* + \Delta\mathbf{u}_k$ at every step k); all plots presented in this section are obtained by averaging results of 10000 simulations. In Figures 5(b) and 5(a), we present the system's total cost when the attacker does not follow the strategy $\mathbf{f}_a$ but rather replays previous sensor measurements with delay of $T = 25s$, starting from time $t = 26s$. We compare two strategies for the system, using Controller 2 in each step and following the strategy $\mathbf{g}_a$. Figure 5(b) shows the probability that the system is at state s_1 (successfully detected a replay attack) over time. These results illustrate that the suboptimal system strategy $\mathbf{g}_a$ results in a lower control cost compared to the non-optimal Controller 2, although Controller 2 insures a higher detection rate. Note that in regular operation modes, when no attacks occur, it is inefficient to sacrifice control performance at every time step. By switching controllers we reduce the performance loss, as shown in Figure 5(c), while being able to detect potential attacks.

Besides finding the optimal or suboptimal strategy, we can also use the stochastic game framework to find the 'best reply' strategy for a specific attacker behavior. For example, the attacker's strategy can be to always apply a certain type of attack before being detected. When the computational complexity of Algorithm 1 is high and we want a faster iteration, we can calculate the best reply strategy with respect to one attack type from the attack action space, run the algorithm for a different action type each time, and find the worst-case cost. This approximation is reasonable since when we do not have expectation that a certain attack type will appear, then our goal is to keep the system performance acceptable under the worst-case attack type.

VI. CONCLUSION

In this paper, we have proposed the use of noncooperative stochastic games to design a suboptimal switching control policy that balances control performance with the intrusion detection rate for replay attacks. We have presented the detailed quantification process for the game parameters, which utilizes knowledge of the system's dynamics. To solve the nonstationary stochastic game, we have developed a suboptimal value iteration algorithm by considering each iteration as a robust game. Note that the quantification process and the proposed algorithm can be generalized and applied for optimal design of LTI plants with finite number, finite cost components facing attacks. In future, we will explore the use of game theoretic methods to provide system resiliency against other types of attacks, and revise the presented algorithm to reduce its computational complexity.

REFERENCES

[1] K.-D. Kim and P. R. Kumar, "Cyber-physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. special Centennial-Issue, 2012.

[2] A. Cardenas, S. Amin, and S. S. Sastry, "Research challenges for the security of control systems." in *Proceedings of the 3rd USENIX Workshop on Hot topics in security*, 2008, Article 6.

[3] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in *Critical Infrast. Protection*, 2007, pp. 73–82.

[4] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *1st international conference on HiCoNS*, 2012, pp. 55–64.

[5] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *47th Annual Allerton Conference on Communication, Contro, and Computing*, 2009, pp. 911–918.

[6] M. Manshaei, Q. Zhu, T. Alpcan, T. Basar, and J.-P. Hubaux, "Game theory meets network security and privacy," *ACM Comput. Surv.*, vol. 45, no. 3, pp. 25:1–25:39, 2013.

[7] Q. Zhu and T. Basar, "Dynamic policy-based ids configuration," in *48th IEEE Conference on Decision and Control (CDC)*, 2009, pp. 8600–8605.

[8] A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *Workshops at the Int. Conf. on Distributed Computing Systems*, 2008, pp. 495–500.

[9] S. Verdu and H. Poor, "On minimax robustness: A general approach and applications," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 328–340, 1984.

[10] A. S. Nowak, "Approximation theorems for zero-sum nonstationary stochastic games," in *American Mathematical Society*, vol. 92, no. 3, 1984.

[11] M. Aghassi and D. Bertsimas, "Robust game theory," *Math. Program.*, no. 1, pp. 231–273, 2006.

[12] M. J. Kearns, Y. Mansour, and S. P. Singh, "Fast planning in stochastic games," in *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, 2000, pp. 309–316.

[13] Y. M. Chabukswar, R. and B. Sinopoli, "Detecting integrity attacks on scada systems," in *Proceedings of the 18th IFAC World Congress*, vol. 18, 2011.