Stochastic Game in Remote Estimation Under DoS Attacks

Kemi Ding, Subhrakanti Dey, Daniel E. Quevedo, and Ling Shi

***Abstract*—This letter studies remote state estimation under denial-of-service (DoS) attacks. A sensor transmits its local estimate of an underlying physical process to a remote estimator via a wireless communication channel. A DoS attacker is capable to interfere the channel and degrades the remote estimation accuracy. Considering the tactical jamming strategies played by the attacker, the sensor adjusts its transmission power. This interactive process between the sensor and the attacker is studied in the framework of a zero-sum stochastic game. To derive their optimal power schemes, we first discuss the existence of stationary Nash equilibrium for this game. We then present the monotone structure of the optimal strategies, which helps reduce the computational complexity of the stochastic game algorithm. Numerical examples are provided to illustrate the obtained results.**

***Index Terms*—Stochastic game, DoS attack, cyber-physical systems security.**

I. Introduction

CYBER-PHYSICAL systems (CPSs) [1] have attracted much research interests in recent years. Taking advantage of information and communication technology, these systems equip the underlying physical processes with intelligent control which makes the system design and deployment much easier. With a promising future, CPSs, the next generation of engineering systems, have been applied in diverse areas including smart grids, intelligent transportation/architecture, self-driving cars and medical device network.

Despite the dramatic success of CPSs, there is an urgent need to address the safety problems as a result of the close integration of physical systems and the cyber world. Due to the vulnerability of open communication networks, CPSs are easily exposed to cyber attacks. For large-scale CPSs, such as transportation infrastructure and power grid, these attacks may lead to severe economic losses (e.g., traffic congestion and large-area power outage [2]) or even threaten human lives (e.g., Iran's nuclear centrifuges accident [3]). Cardenas *et al.* [4] categorized the cyber attacks into two typical classes: deception (integrity) attacks and denial-of-service (DoS) attacks. The former attempts to manipulate the transmitted data packets stealthily, while the latter affects the availability of data by blocking the communication channels. From the attacker's perspective, how to construct more subtle (or stealthy) attacks with more severe system consequences was studied in many preliminary works. Zhang *et al.* [5] investigated an energy-constrained attack policy taken by a DoS attacker to degrade the remote estimation accuracy. Teixeira *et al.* [6] characterized the disastrous attack policy (including replay, zero dynamics and bias injection attacks) in a general attack space. A set of effective countermeasures was developed against replay attacks [7], linear deception attacks [8], and DoS attacks [9].

In this letter, we focus on remote state estimation under DoS attacks. A sensor monitors a physical process and forwards its data to a remote estimator over a communication channel which can be interfered by an intelligent attacker. Our goal is to design an energy-efficient transmission scheme for the sensor to minimize the remote estimation error. We use game-theoretical tools to model and analyze the interaction between the sensor and the attacker. Several recent studies about CPS security under the game-theoretic framework have been carried out [7], [9], [10]. In [7] a finite-horizon zero-sum stochastic game model was employed to capture the interaction between a detector and a replay attacker. The authors used robust game techniques to calculate the non-stationary equilibrium solution, which, however, is more difficult to implement than the stationary one. Zhu and Basar [10] designed an optimal stationary cyber-policy to ensure system security through an infinite-horizon zero-sum stochastic game between a defender and an attacker. As for the Nash equilibria of the aforementioned stochastic games, their existence is easy to prove due to the finiteness of state space or the boundedness of immediate reward functions, and no structural results regarding the associated optimal policies are provided. A preliminary version of parts of the present manuscript [9] investigated a jamming game where the sensor and the attacker possess multiple discrete power levels. Different from [9], here we

Manuscript received March 6, 2017; revised May 24, 2017; accepted May 25, 2017. Date of publication June 1, 2017; date of current version June 14, 2017. The work of K. Ding and L. Shi was supported by the RGC General Research Fund under Grant 16222716. Recommended by Senior Editor J. Daafouz. *(Corresponding author: Kemi Ding.)*

K. Ding and L. Shi are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: kdingaa@ust.hk; eesling@ust.hk).

S. Dey is with the Department of Engineering Science, Uppsala University, 751 21 Uppsala, Sweden (e-mail: subhrakanti.dey@angstrom.uu.se).

D. E. Quevedo is with the Department of Electrical Engineering, Paderborn University, 33098 Paderborn, Germany (e-mail: dquevedo@ieee.org).

Digital Object Identifier 10.1109/LCSYS.2017.2711044

study continuous power sets and a general form of the packet-dropout rate, which embeds [9] as a special case. The proof of existence of the optimal solution also becomes more challenging due to the infinite state space and continuous power sets. Furthermore, [9] uses a computational method (Nash Q-learning) to solve the stochastic game numerically, while in this letter, we prove that the optimal strategies belong to a small class possessing certain structural properties. Similar idea has been seen in the sensor scheduling area [5], [11]. However, they focused only on one side, i.e., either the attacker or the defender. Compared with the previous works, our contributions consist of the following:

1) We model the remote estimation problem under strategic DoS attacks by a zero-sum stochastic game with countable states and continuous action sets. The existence of a stationary optimal policy pair for this game is proved via value iteration (Theorem 1).

2) Under suitable conditions, we develop the monotone structure of the optimal solutions for this game (Theorem 2). Furthermore, when the action set for each player consists of finite power levels, the optimal strategy for corresponding player has a threshold-type structure (Section III-D).

The remainder of this letter is organized as follows. Mathematical models of the system are described in Section II. In Section III, a stochastic game is introduced and the main results on the existence of the solution and structural analysis are presented. An example and some concluding remarks are shown in Sections IV and V. The proofs of lemmas are presented in the full version [12].

Notations: $\mathbb{R}^n$ is the n dimensional Euclidean space. $\mathbb{S}_+^n$ is the set of n by n positive semi-definite matrices. When $X \in \mathbb{S}_+^n$, it is written as $X \geq 0$. $X \geq Y$ if $X - Y \in \mathbb{S}_+^n$. $\mathbb{E}[\,\cdot\,]$ is the expectation of a random variable and $\mathtt{Tr}(\cdot)$ is the trace of a matrix. $|\mathcal{A}|$ represents the cardinality of set $\mathcal{A}$. For functions $f_1, f_2 : \mathbb{S}_+^n \to \mathbb{S}_+^n$, $f_1 \circ f_2$ is defined as $f_1 \circ f_2(X) \triangleq f_1\big(f_2(X)\big)$. The function $\delta_{kj}(\cdot) = \begin{cases} 1, \text{ if } k = j; \\ 0, \text{ otherwise.} \end{cases}$. $\pi^{1,2} \triangleq \{\pi^1, \pi^2\}$. $\rho(A)$ is the spectrum radius of A and $\mathtt{diag}(\cdot)$ represents a diagonal matrix.

II. Problem Setup

In this section, we introduce mathematical models of the system of interest, see Figure 1.

A. Local Kalman Filter

Consider the following linear time-invariant system:

$$x_{k+1} = Ax_k + w_k, \;\; y_k = Cx_k + v_k, \tag{1}$$

where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^m$ represent the system state vector and the sensor measurement at time k, respectively. The noises $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are assumed to be zero-mean i.i.d Gaussian sequences with $\mathbb{E}[w_k w_j^\top] = \delta_{kj} Q$ $(Q \geq 0)$, $\mathbb{E}[v_k v_j^\top] = \delta_{kj} R$ $(R > 0)$, and $\mathbb{E}[w_k v_j^\top] = 0$ $\forall j, k$. The initial state $x_0 \sim \mathcal{N}(0, \Pi_0)$ is uncorrelated with w_k and v_k. The pair (A, C) is detectable and $(A, \sqrt{Q})$ is stabilizable.

The sensor runs a Kalman filter locally to estimate the state x_k based on all collected measurements at time k.

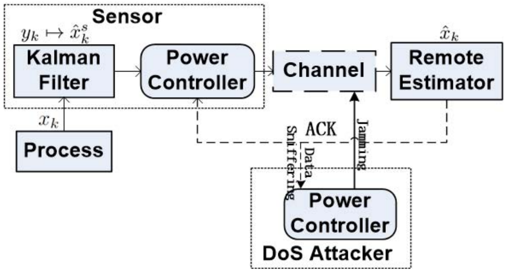

Fig. 1. Remote state estimation under DoS attacks.

Denote the local estimate as $\hat{x}_k^s$, i.e., $\hat{x}_k^s = \mathbb{E}[x_k | y_0, \ldots, y_k]$. The corresponding estimation error e_k^s and the error covariance matrix P_k^s are defined as $e_k^s \triangleq x_k - \hat{x}_k^s$ and $P_k^s \triangleq \mathbb{E}[(e_k^s)(e_k^s)^\top | y_0, \ldots, y_k]$. These terms are calculated via the standard Kalman filtering algorithm, where the iteration starts from $\hat{x}_0^s = 0$ and $P_0^s = \Pi_0$. For notational simplicity, we define the Lyapunov and Riccati operators h and $\tilde{g} : \mathbb{S}_+^n \to \mathbb{S}_+^n$ as $h(X) \triangleq AXA^\top + Q$, $\tilde{g}(X) \triangleq X - XC^\top[CXC^\top + R]^{-1}CX$.

Due to the detectability and stabilization assumptions, the error covariance P_k^s converges exponentially to a unique fixed point $\overline{P}$ of $\tilde{g} \circ h$ [13]. Since we consider infinite time horizon, we ignore the transient periods and assume that the Kalman filter at the sensor has entered steady state; i.e., $P_k^s = \overline{P}$, $k \geq 1$. To avoid trivial problems, we assume A is unstable. Define a function $h^s(\overline{P}) \triangleq \underbrace{h \circ \cdots h}_{s}(\overline{P})$, $s \in \{0, 1, \ldots\}$ with $h^0(\overline{P}) = \overline{P}$. As mentioned in [9], the function $h^s(\overline{P})$ has the following property.

Proposition 1: $\mathtt{Tr}[h^s(\overline{P})]$ is increasing in $s \in \{0, 1, \ldots\}$: for $0 \leq s_1 < s_2$, $\mathtt{Tr}[\overline{P}] \leq \mathtt{Tr}[h^{s_1}(\overline{P})] < \mathtt{Tr}[h^{s_2}(\overline{P})]$.

B. Communication Channel

We assume the channel between the sensor and the estimator is memoryless and it has independent additive white Gaussian noises (AWGN). To measure the non-ideal packet losses, we introduce the conventional packet-error-rate (PER), which is monotonically decreasing with the signal-to-noise-ratio (SNR) for any modulation scheme. An intelligent attacker launches DoS attacks by jamming the channel which reduces the packet arrival rate. Let $b_k \in \mathbb{E}^s$ and $a_k \in \mathbb{E}^a$ denote the transmission power of the sensor and the interference power of the attacker at time k, respectively, where $\mathbb{E}^s$ and $\mathbb{E}^a$ represent the available power sets for the sensor and the attacker. Define

$$d(\gamma_k) \triangleq d(a_k, b_k), \;\; \gamma_k \triangleq L\frac{h_2 b_k}{h_1 a_k + n_0}, \tag{2}$$

where $d(\cdot)$ is some general form PER depending on the specific modulation used, n_0 is the power of the additive white channel noise, the parameter h_2 (h_1 resp.) is the channel gain from the sensor (the attacker resp.) to the remote estimator, and L is the spreading gain of the communication system. These channel parameters are assumed to be time-invariant. According to [14], $d(\cdot)$ is strictly concave and decreasing in γ_k. For this erasure channel, the arrival of the packet can be characterized by a binary random process η_k, and $\eta_k = 0$ represents the packet loss. From the previous discussion, we have $\texttt{Pr}(\eta_k = 0) = d(a_k, b_k)$.

C. Remote State Estimation

Based on the received data packets up to time k, the remote estimator computes its own estimate of x_k, which is denoted by $\hat{x}_k$ with corresponding error covariance P_k. The estimate $\hat{x}_k$ is obtained as follows [9]: $\hat{x}_k = \eta_k \hat{x}_k^s + (1 - \eta_k)A\hat{x}_{k-1}$. Consequently, the recursion of the error covariance P_k is

$$P_k \triangleq \eta_k \overline{P} + (1 - \eta_k)h(P_{k-1}). \tag{3}$$

Recall that $\overline{P}$ is the error covariance in the steady state. According to (3), the possible values that P_k can take form an infinite set $\{\overline{P}, h(\overline{P}), h^2(\overline{P}), \ldots\}$.

To simplify the notations, define $\tau_k \in \mathbb{Z}$ as the holding time: $\tau_k \triangleq k - \max_{0 \le l \le k}\{l : \eta_l = 1\}$, which represents the intervals between the present moment k and the most recent time that the data packet has successfully arrived at the remote estimator. By definition, the iteration of the holding time is

$$\tau_k = \begin{cases} 0, & \text{if } \eta_k = 1, \\ \tau_{k-1} + 1, & \text{otherwise.} \end{cases} \tag{4}$$

Without loss of generality, we assume $\eta_0 = 1$. Then it is straightforward to show that $P_k = h^{\tau_k}(\overline{P})$. We assume that η_k is obtained by the sensor based on some short acknowledgement frames (ACKs) up to time k, which are sent back from the estimator through a reliable channel. The ACKs are also known to the attacker via eavesdropping the feedback channel.

D. Problem of Interest

The sensor and the attacker aim to design some energy-efficient transmission/jamming power schemes with opposite goals, i.e., the sensor wishes to minimizes the remote estimation error while the attacker tries to maximize it. This interactive decision-making process is analyzed under a stochastic game framework in the next section.

III. Main Results

In this section, a dynamic game framework is introduced to model the strategic interaction between the sensor and the attacker. Notice that we consider infinite state space and continuous action sets for this game. Analyzing the solution of such game is well-known to be difficult and challenging [15]. Besides an existence result, some structural properties of the solution are also provided.

A. Game Model

We introduce a stochastic game which is composed of five tuples [16]: $\mathcal{G} \triangleq < \mathcal{I}, \mathcal{S}, \mathcal{A}, q, \mathcal{R} >$.

1) Player: $\mathcal{I} = \{1, 2\}$ is the set of the players, where $i = 1$ represents the attacker, and $i = 2$ the sensor. For the subsequent equilibrium analysis, we assume both players are rational.

2) State: $\mathcal{S} = \{0, 1, 2, \ldots\}$ is the state space. We define the state of the game at time step k as the holding time, i.e., $s_k \triangleq \tau_k$ and $s_k \in \mathcal{S}$. Note that the state is equivalent to the estimation error covariance P_k.

3) Action: $\mathcal{A} = \prod_{i \in \mathcal{I}} \mathcal{A}_i$ is the joint action set, where $\mathcal{A}_i$ is a compact set denoting the actions available to player i. At time k, the attacker selects the jamming power $a_k \in \mathcal{A}_1 \triangleq [a_{\min}, a_{\max}]$, and the sensor decides the transmission power $b_k \in \mathcal{A}_2 \triangleq [b_{\min}, b_{\max}]$. Let $\mathscr{P}_i$ denote the space of measures on $\mathbf{B}(\mathcal{A}_i)$, the Borel subsets of $\mathcal{A}_i$, endowed with the weak topology (hence it is Polish space, a complete and separable metric space [17]). Denote by $\pi_k^1 \in \mathscr{P}_1$ and $\pi_k^2 \in \mathscr{P}_2$ mixed strategies for the attacker and the sensor, respectively. The joint action (or pure strategy) at time k is denoted by (a_k, b_k).

4) Transition Probability: The transition probability $q(s_{k+1}|s_k, a_k, b_k) \in [0, 1] : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$. From (4), s_k has the Markov property. Let $m, m' \in \mathcal{S}$ and $(a, b) \in \mathcal{A}$, then $\forall k \ge 0$, $q(s_{k+1} = m'|s_k = m, a_k = a, b_k = b)$

$$= \begin{cases} d(a, b), & \text{if } m' = m + 1; \\ \underline{d}(a, b) = 1 - d(a, b), & \text{if } m' = 0; \\ 0, & \text{otherwise.} \end{cases}$$

5) Immediate Reward: $\mathcal{R} = \{R_i, i \in \mathcal{I}\}$ is the reward set and R_i represents the immediate reward function for player i with $R_i : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$. The reward function for the attacker is given by

$$R_1(m, a, b) = r(m) + \theta(a) + \vartheta(b), \;\; m \in \mathcal{S}, (a, b) \in \mathcal{A},$$

where $r(m) \triangleq \texttt{Tr}[h^m(\overline{P})]$ with $r(0) = \texttt{Tr}[\overline{P}]$, $\theta(\cdot) \le 0$ is a decreasing function in a and $\vartheta(\cdot) \ge 0$ is an increasing function in b. Notice that the attacker attempts to maximize the state estimation error in an energy-efficient way. More transmission energy consumed by the sensor is also preferred by the attacker. The immediate reward function for the sensor is $R_2(m, a, b) \triangleq -R_1(m, a, b)$. For simplicity, we ignore the subscription of $R_1(\cdot)$ and instead represent the immediate reward for the attacker by $R(m, a, b)$ without ambiguity.

The game is played as follows.Whenever the system is in state s_k at time $k \ge 0$, they independently and simultaneously take actions (a_k, b_k) according to randomized stationary policies to be described in details in the next paragraph. As a consequence, the following happens: the attacker receives an immediate reward $R(s_k, a_k, b_k)$ (or a cost $-R(s_k, a_k, b_k)$ is incurred for the sensor simultaneously); the system moves to a new state s_{k+1} with a transition probability determined by $q(s_{k+1}|s_k, a_k, b_k)$. The goal of the attacker is to maximize its reward (whereas the sensor aims to minimize its cost) with respect to the discounted performance criterion $J(\cdot)$, which is defined in (6).

The decision rule for player i is a sequence of mixed strategies $\pi^i = \{\pi_0^i, \pi_1^i, \ldots, \pi_k^i, \ldots\}$ with $\pi_k^i \in \mathscr{P}_i, i \in \{1, 2\}$. Among the admissible strategies for each player, we consider the randomized *stationary* ones. Denote by Π_i the family of all randomized *stationary* strategies for player $i \in \{1, 2\}$ with their stochastic kernels satisfying: for each $k \ge 0$ and $s \in \mathcal{S}$, $\pi_k^i(\cdot|s)$ is a probability measure on $\mathcal{A}_i$, and for every $D \in \mathbf{B}(\mathcal{A}_i)$, $\pi_k^i(D|s)$ is a Lebesgue measurable function; for each $s \in \mathcal{S}$, there is a probability measure $\pi^i(\cdot|s) \in \mathscr{P}_i$ such that $\pi^i(\cdot|s) = \pi^i_k(\cdot|s), \forall k \geq 0$.

Hence, the pair of *stationary* strategies for the players depends only on the current state $s \in \mathcal{S}$ and with abuse of notation, the stationary strategy for player i is denoted by $\pi^i(\cdot|s)$ with $\pi^i_k(D|s)$ representing the probability with which a particular set of action $D \in \mathbf{B}(\mathcal{A}_i)$ will be played. As a shorthand notation, for each pair of strategies $(\pi^1, \pi^2) \in \Pi \triangleq \Pi^1 \times \Pi^2$, the associated transition probabilities are defined as follows:

$$q(s'|s, \pi^1, \pi^2) \triangleq \int_{\mathcal{A}} q(s'|s, a, b)\pi^1(da|s)\pi^2(db|s), \;\; s' \in \mathcal{S}, \tag{5}$$

and $R(s, \pi^1, \pi^2)$ is defined similarly, with $R(s, a, b)$ in lieu of $q(s'|s, a, b)$. Note that $R(s, a, b)$ and $q(s'|s, a, b)$ are continuous on $\mathcal{A}$, which implies that $q(s'|s, \pi^1, \pi^2)$ and $R(s, \pi^1, \pi^2)$ are continuous on Π.

The accumulated expected reward for the attacker under the discounted criterion is defined as follows:

$$J_\alpha(m_0, \pi^{1,2}) \triangleq \mathbb{E}^{\pi^{1,2}}\left[\sum_{k=0}^{\infty} \alpha^k R(s_k, a_k, b_k)|s_0 = m_0\right], \tag{6}$$

in which $m_0 \in \mathcal{S}$ and $\alpha \in [0, 1)$ is the discount coefficient.

Considering the stationary property, we introduce the solution concept for the stochastic game, a *stationary Nash equilibrium (SNE)*:

Definition 1: For the zero-sum stochastic game $\mathcal{G}$, a stationary policy $\pi^{i,\star}, \forall i \in \mathcal{I}$ is a stationary Nash equilibrium if for all players $i \in \mathcal{I}$, for all $\pi^i \in \mathscr{P}_i$ and for any initial state $m_0 \in \mathcal{S}$,

$$J_\alpha(m_0, \pi^{1,\star}, \pi^2) \geq J_\alpha(m_0, \pi^{1,\star}, \pi^{2,\star}) \;\; \geq J_\alpha(m_0, \pi^1, \pi^{2,\star}).$$

Thus the SNE is a sequence of distribution over actions from which no player is motivated to deviate unilaterally. The stochastic game problem is to find $(\pi^1, \pi^2) \in \Pi$ to obtain

$$J_\alpha(s) \triangleq \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} J_\alpha(s, \pi^1, \pi^2), \;\; \forall s \in \mathcal{S}. \tag{7}$$

The optimal strategies and the value of the game is denoted by $\pi^{1,\star}, \pi^{2,\star}$ and $J_\alpha(s)$, respectively.

B. Relation to a One-Stage Game

For a given state s, the sup-inf problem in (7) is analogous to a one-stage game $\mathcal{G}'$ (specifically, a two-player zero-sum game on the unit square), in which the expected reward function $J_\alpha(s, \pi^1, \pi^2)$ is replaced by the utility function $U(a, b)$. Hence, before discussing the existence and the structural property of the SNE for the stochastic game $\mathcal{G}$, we introduce some preliminary results for the one-stage game.

First, we define *a point of the spectrum* [16] of mixed strategy for $\mathcal{G}'$ as follows.

Definition 2: A pure strategy z_0 for a player is called *a point of the spectrum* of the mixed strategy Z if $\forall \varepsilon > 0, \omega \triangleq \{z{:}|z - z_0| < \varepsilon\}$, the integral $\int_\omega dZ(z)$ is positive.

Note that this definition degenerates to the definition of *support* when considering finite action set. Based on (7), we define an operator for $\mathcal{G}'$ to find its game value: $val(U) = \sup_{\pi_1 \in \mathscr{P}_1} \inf_{\pi_2 \in \mathscr{P}_2} \int_{a,b} U(a, b)d\pi_1(a)d\pi_2(b)$ (π_1 and π_2 are the mixed strategies or probability distributions over action sets $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively). Here, we have the following properties for $\mathcal{G}'$.

Lemma 1: For the two-player zero-sum game $\mathcal{G}'$ on the unit square, we have

(1) If one player possesses a pure optimal strategy $a^\star$, the other player also plays a pure optimal strategy: $val(U) = \min_b U(a^\star, b)$;

(2) If the utility function $U(a, b)$ is continuous in both variables, for a pair of arbitrary optimal strategies $(\pi_1^\star, \pi_2^\star)$, and for any a_0 (b_0 resp.), a point of the spectrum of $\pi_1^\star$ ($\pi_2^\star$ resp.), $val(U) = U(a_0, \pi_2^\star) = U(\pi_1^\star, b_0)$.

Next, we illustrate the connection between the stochastic game $\mathcal{G}$ and $\mathcal{G}'$ via value iteration method. Let $\mathcal{K}$ be the set of all real-valued functions on $\mathcal{S}$. Define the operator $C : \mathcal{S} \times \mathcal{A} \times \mathcal{K} \to \mathbb{R}$ as

$$\begin{aligned} C(s, a, b, f) &\triangleq \mathbb{E}[f(s_{k+1})|s_k = s, a_k = a, b_k = b] \\ &= d(a, b)f(s_{k+1} = s + 1) + \underline{d}(a, b)f(s_{k+1} = 0), \end{aligned}$$

in which $f(\cdot) \in \mathcal{K}$ and $d(a, b)$ is the packet-dropout rate when the pair of action (a, b) is taken. Moreover, define the operator $W : \mathcal{S} \times \mathcal{A} \times \mathcal{K} \to \mathbb{R}$ as

$$W(s, a, b, f) \triangleq R(s, a, b) + \alpha C(s, a, b, f).$$

For given (s, f), we regard $C(s, f)$ and $W(s, f)$ as a one-stage game with utility functions $C(s, a, b, f)$ and $W(s, a, b, f)$, and similar to (5), $C(s, \pi_1, \pi_2, f)$ and $W(s, \pi_1, \pi_2, f)$ are well-defined. Denote by $T_\alpha f(s) : \mathcal{K} \to \mathcal{K}$ the operator associated with the stochastic game $\mathcal{G}$, that is, $T_\alpha f(s) = valW(s, f), \forall s \in \mathcal{S}$. The property of this operator T_α is shown in Lemma 2.

Assumption 1: The maximum packet-dropout rate and the system matrix A satisfy that $d(a_{max}, b_{min})\rho^2(A) < 1$.

Notice that this type of assumption is common in state estimation problem with intermittent observations [18]. For unstable systems, if the condition is not satisfied, the attacker can choose to attack with a certain jamming power constantly to gain an unbounded expected error covariance (i.e., the attacker dominates this game), thus there exists no equilibrium. A statement and proof of the equilibrium existence result are demonstrated in Theorem 1 and Lemma 2 below.

Lemma 2: Suppose **Assumption 1** holds, then

(1) J_α is the unique solution to *the optimality equation*, i.e., $J_\alpha(s) = T_\alpha J_\alpha(s)$;

(2) Any stationary policy pair (π^1, π^2) achieving $valW(s, J_\alpha(s)), \forall s \in \mathcal{S}$ is optimal for the game $\mathcal{G}$;

(3) For every initial value function $f(s) \in \mathcal{K}$, then $\{T_\alpha^n f(s)\}$ is a convergent sequence in $\mathcal{K}$ and $\lim_{n\to\infty} T_\alpha^n f(s) = J_\alpha(s)$.

Theorem 1: There exists a SNE for the attacker-sensor game $\mathcal{G}$ if **Assumption 1** is satisfied, and for each state s, it is a NE solution to a one-stage game with utility function $W(s, a, b, J_\alpha(s))$.

The proof for this theorem follows directly from Lemma 2. Furthermore, Lemma 2 implies that the α-discounted *value iteration algorithm* converges to the unique point (i.e., the game value). Intuitively speaking, **Assumption 1** restricts the

$$\begin{aligned}\Delta_1 &\triangleq W(m, a_2, \nu(\cdot|m), J_\alpha) - W(m, a_1, \nu(\cdot|m), J_\alpha) - [W(m-1, a_2, \nu(\cdot|m-1), J_\alpha) - W(m-1, a_1, \nu(\cdot|m-1), J_\alpha)]\\ &< 0. \qquad (8)\end{aligned}$$

$$\begin{aligned}\Delta_2 &\triangleq W(m, \mu(\cdot|m), b_2, J_\alpha) - W(m, \mu(\cdot|m), b_1, J_\alpha) - [W(m-1, \mu(\cdot|m-1), b_2, J_\alpha) - W(m-1, \mu(\cdot|m-1), b_1, J_\alpha)]\\ &> 0. \qquad (9)\end{aligned}$$

$$\begin{aligned}\Delta_1 &= \mathbb{E}^{\nu(\cdot|m)}\big[R(m, a_2, b) + \alpha C(m, a_2, b, J_\alpha) - R(m, a_1, b) - \alpha C(m, a_1, b, J_\alpha)\big]\\ &\quad - \mathbb{E}^{\nu(\cdot|m-1)}\big[R(m-1, a_2, b) + \alpha C(m-1, a_2, b, J_\alpha) - R(m-1, a_1, b) - \alpha C(m-1, a_1, b, J_\alpha)\big]\\ &= \alpha[J_\alpha(m+1) - J_\alpha(0)]\mathbb{E}^{\nu(\cdot|m)}[d(a_2, b) - d(a_1, b)] - \alpha[J_\alpha(m) - J_\alpha(0)]\mathbb{E}^{\nu(\cdot|m-1)}[d(a_2, b) - d(a_1, b)]. \qquad (10)\end{aligned}$$

$$\begin{aligned}&W(m, a_1, b_2, J_\alpha) - W(m, a_1, b_1, J_\alpha) - [W(m-1, a_2, b_2, J_\alpha) - W(m-1, a_2, b_1, J_\alpha)]\\ &\quad = \alpha[d(a_1, b_2) - d(a_1, b_1)][J_\alpha(m+1) - J_\alpha(0)] - \alpha[d(a_2, b_2) - d(a_2, b_1)][J_\alpha(m) - J_\alpha(0)]. \qquad (11)\end{aligned}$$

value of the maximum data-packet dropout rate to guarantee the existence of SNE. For subsequent structural analysis, we present the property of the game value function in Lemma 3.

Assumption 2: There exists $s_0 \geq 0$ such that $\forall s \geq s_0$, $r(s+1) - c_0 r(s) + (c_0 - 1) r(0) \geq c_1$, where constants $c_0 > 1$ and $c_1 = (c_0 - 1)[\theta(a_{\min}) + \vartheta(b_{\max}) - \theta(a_{\max}) - \vartheta(b_{\min})]$.

Similar assumptions on the reward function (for example, linearity [19]) are widely adopted in the analysis of structural strategies for stochastic games. For scalar systems, **Assumption 2** holds if and only if $A^2 \geq c_0$. An example to demonstrate its feasibility in a higher dimensional system is given in Section IV. Notice that **Assumption 2** is critical in the proof of the following lemma.

Lemma 3: For the stochastic game $\mathcal{G}$,

(1) The function $J_\alpha(s)$ is monotonically increasing in s;

(2) For any $c_0 > 1$, if **Assumption 2** holds, then the value function $J_\alpha(s)$ satisfies $\frac{J_\alpha(s+1) - J_\alpha(0)}{J_\alpha(s) - J_\alpha(0)} \geq c_0$ for $s \geq s_0$.

C. Structure of Optimal Stationary Strategies

The existence of SNE is proved via value iteration convergence in Lemma 2, which can be computed as a one-stage game with the optimal value function $J_\alpha(s)$. In this section, we present some special structure of the SNE of $\mathcal{G}$.

Definition 3: For a stationary strategy taken by the attacker $\{\mu(\cdot|s), s \in \mathcal{S}\}$ with $\mu(\cdot|s) \in \mathscr{P}_1$, denote by $a^{\sup}(\pi^1 = \mu|s)$ ($a^{\inf}(\pi^1 = \mu|s)$ resp.) the greatest (smallest resp.) point $a \in \mathcal{A}_1$ in the spectrum of $\mu(\cdot|s)$. This strategy is strongly monotone nondecreasing if $a^{\sup}(\pi^1 = \mu|s = s_2) \leq a^{\inf}(\pi^1 = \mu|s = s_1)$ holds for any $s_1, s_2 \in \mathcal{S}$ with $s_2 < s_1$.

Note that similar definition is applied to the stationary strategy played by the sensor. Before establishing the structure of the SNE for the stochastic game $\mathcal{G}$, we introduce the following result.

Lemma 4: For any SNE $(\mu, \nu) \in \{(\pi^{1,\star}, \pi^{2,\star})\}$ of the stochastic game $\mathcal{G}$, if (8) (or (9)), as shown at the top of this page, holds for any $a_1 > a_2$ (or $b_1 > b_2$) and for any $m \in \mathcal{S}$, then μ (or ν) is monotone nondecreasing.

The inequalities in (8) and (9) are referred to the submodularity and super-modularity for a two-player zero-sum stochastic game. Note that in order to prove Δ_2 is positive, it suffices to show that for any $b_1 > b_2$ and any a_1 and a_2, $W(m, a_1, b_2, J_\alpha) - W(m, a_1, b_1, J_\alpha) - [W(m-1, a_2, b_2, J_\alpha) - W(m-1, a_2, b_1, J_\alpha)] > 0$. We now present the structure of the SNE.

Theorem 2: Suppose **Assumption 2** holds, then any of the SNE $\{\mu, \nu\}$ of the attacker-sensor game is (partially) nondecreasing monotone for $s \geq s_0$.

Proof: First, suppose the optimal strategy ν is monotone nondecreasing. In order to prove the monotonicity of μ, it is sufficient to verify $\Delta_1 < 0$ in (10), as shown at the top of this page, for any $a_1 > a_2$ and $m \in \mathcal{S}$. Define $\delta_1(a_1, a_2, b) = d(a_2, b) - d(a_1, b) < 0$. One easily obtains that $\frac{\partial \delta_1(a_1, a_2, b)}{\partial b} < 0$ since $d(a, b)$ is concave in $\gamma(a, b)$. As $b^{\inf}(\nu_m) \geq b^{\sup}(\nu_{m-1})$, $\delta_1^{\sup}(a_1, a_2, b|\nu, m) \triangleq \delta_1(a_1, a_2, b^{\inf}(\nu|m)) \leq \delta_1^{\inf}(a_1, a_2, b|\nu, m-1) \triangleq \delta_1(a_1, a_2, b^{\sup}(\nu|m-1))$. Hence, $\mathbb{E}^{\nu(\cdot|m)}\delta_1(a_1, a_2, b) \leq \mathbb{E}^{\nu(\cdot|m)}\delta_1(a_1, a_2, b) < 0$. Based on Lemma 3, $J_\alpha(m+1) - J_\alpha(0) > J_\alpha(m) - J_\alpha(0) > 0$ and it follows that $\Delta_1 < 0$.

Next, we establish the monotonicity of ν via checking a sufficient condition for $\Delta_2 > 0$ as mentioned previously. For any a_1, a_2 and $b_1 > b_2$, we have (11), as shown at the top of this page. Define $\delta_2(b_1, b_2, a) = d(a, b_2) - d(a, b_1) > 0$ and we have $\frac{\partial \delta_2(b_1, b_2, a)}{\partial a} = \frac{\partial d(a, b_2)}{\partial a} - \frac{\partial d(a, b_1)}{\partial a} < 0$. If (11) is positive, $\frac{J_\alpha(m+1) - J_\alpha(0)}{J_\alpha(m) - J_\alpha(0)} > \frac{\delta_2(b_1, b_2, a_2)}{\delta_2(b_1, b_2, a_1)}$ should be satisfied. Define $\delta_3(a_1, a_2, b_1, b_2) \triangleq \frac{\delta_2(b_1, b_2, a_2)}{\delta_2(b_1, b_2, a_1)}$, in which $a_1, a_2 \in \mathcal{A}_1$, $b_1, b_2 \in \mathcal{A}_2$ and $b_1 \geq b_2$. Since $\delta_3(\cdot)$ is continuous on a compact set with removable discontinuous points, according to the extreme value theorem the function $\delta_3(\cdot)$ has an upper bound, which is denoted by $\delta_3^{\max} > 1$. Hence, we obtain a sufficient condition for the monotonicity structure of optimal strategies $\{\mu, \nu\}$, that is, $\frac{J_\alpha(m+1) - J_\alpha(0)}{J_\alpha(m) - J_\alpha(0)} > \delta_3^{\max}$. According to Lemma 3, this inequality holds for $m \geq m_0$ with $c_0 = \delta_3^{\max}$. ■

The special structure of the optimal policy pair (μ, ν) helps reduce the computational complexity of the stochastic game algorithm significantly. From previous discussions, the optimal SNE $\{\mu, \nu\}$ for $\mathcal{G}$ follows the monotonicity when $s \geq s_0$; however, the structure of SNE is irregular for $s < s_0$.

D. Discussion of Discrete Power Levels

Next, we demonstrate the structure of $\{\mu, \nu\}$ for some special cases. Without loss of generality, s_0 is assumed to be zero. Consider the scenario where the action space $\mathcal{A}_i, i \in \mathcal{I}$ is discretized with finite power levels. According to Lemma 1 (1), $\mu(\cdot|s_1)$ and $\nu(\cdot|s_2)$ are mixed strategies iff $s_1 = s_2$ since $\{\mu(\cdot|s), \nu(\cdot|s)\}$ are NE for the one-stage game with utility function $W(s, J_\alpha)$. Moreover, based on Def. 3, it follows that the randomization of the monotone nondecreasing stationary

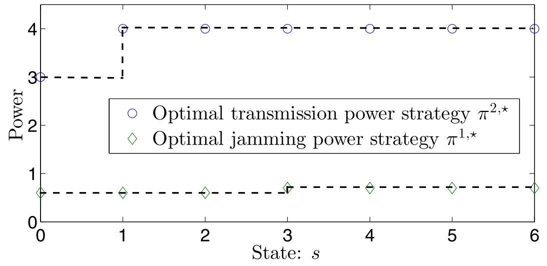

Fig. 2. SNE for the attacker/sensor under different states.

strategy μ (ν resp.) is over no more than $\min(|\mathcal{A}_1, \mathcal{A}_2|) - 1$ states. Assume there are only two actions for each player (i.e., $\mathcal{A}_1 = \{a_{\min}, a_{\max}\}$ and $\mathcal{A}_2 = \{b_{\min}, b_{\max}\}$) and **Assumption 2** is satisfied. Then a SNE for $\mathcal{G}$ is constructed with the threshold structure: $\mu(\cdot|s) = \begin{cases} (1, 0), & \text{if } s < s_1; \\ (p_1, \underline{p_1}), & \text{if } s = s_1; \\ (0, 1), & \text{otherwise,} \end{cases}$ and $\nu(\cdot|s) = \begin{cases} (1, 0), & \text{if } s < s_2; \\ (p_2, \underline{p_2}), & \text{if } s = s_2; \\ (0, 1), & \text{otherwise,} \end{cases}$ where the parameters p_1, p_2, s_1 and s_2 satisfy either $s_1 = s_2$ or $p_1 = p_2 = 1$.

IV. Simulation

In this section, we illustrate the structural equilibrium result with an example arising in a collision avoidance system. Consider the remote state estimation of relative distance and relative speed between two unmanned-ground-vehicles (UGVs) [20], which is based on noisy measurements of the distance between them. This is a simplified version of navigation, as we here only focus on one directional motion. Define the state vector consisting of distance L_k and velocity V_k as $x_k = [L_k \ \ V_k]^\top$. The state equation described in (1) is, for discrete time with sampling period T, $x_{k+1} = \begin{bmatrix} 1 & (1+\frac{\varsigma}{2})T \\ 0 & 1+\varsigma \end{bmatrix} x_k + w_k$, in which ς is the ratio of the acceleration to the velocity (i.e., $V_{k+1} - V_k = \varsigma V_k$). The measurement is imperfect, i.e., $y_k = [1 \ 0]x_k + v_k$. Consider the following parameters $\varsigma = 0.25, T = 0.5, Q = \mathtt{diag}(0.5, 0.5), R = 0.5$. Notice that the two UGVs transmit local estimates to the remote estimator through a fading channel, which is vulnerable to DoS attacks. The energy levels are $\mathcal{A}_1 = [0.6 \ 0.7]$ and $\mathcal{A}_2 = [3 \ 4]$. The channel noise is 0.5 and the packet-dropout rate (notice that PER follows $d(\gamma) = \gamma^{-\frac{1}{2}}$ [21]) is $D = \begin{bmatrix} 0.61 & 0.52 \\ 0.63 & 0.55 \end{bmatrix}$ where $D(i, j)$ represents PER under the i-th jamming power level and the j-th transmission power level. As for the discounted criterion, $\alpha = 0.96$, $\theta(a) = -a$ and $\vartheta(b) = 2b$. By calculation, we have $c_0 = 1.04, c_1 = 0.09$ and $r(0) = 2.79$. It is easy to verify that **Assumption 1** is satisfied and **Assumption 2** holds for $s_0 = 0$. We apply the Nash-Q learning algorithm[1] to obtain SNE for this game. We summarize the optimal stationary strategies for each player in Figure 2: the optimal transmission strategy for the sensor is to send packets with minimum power 3 under states $s = 0$ and use power 4 for state $s \geq 1$; while the attacker adopts jamming power 0.6 when $s \leq 3$ and 0.7 for other states. This demonstrates the threshold-type strategies for the attacker-sensor game $\mathcal{G}$ with discrete action sets.

V. Conclusion

This letter studied remote estimation under strategic DoS attacks, which is formulated under a stochastic game framework. It was shown that the SNE possesses some special structures. In particular, when the actions sets are discrete, a threshold-type strategy can be obtained, which helps design algorithm to calculate optimal strategies with improved computation efficiency.

References

[1] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proc. IEEE*, vol. 100, pp. 1287–1308, May 2012.

[2] M. Assante. *Confirmation of a Coordinated Attack on the Ukrainian Power Grid*. Accessed on May 24, 2017. [Online]. Available: https://ics.sans.org/blog/2016/01/09/confirmation-of-a-coordinated-attack-on-the-ukrainian-power-grid

[3] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.

[4] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *Proc. IEEE 28th Int. Conf. Distrib. Comput. Syst. Workshops*, Beijing, China, 2008, pp. 495–500.

[5] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal Denial-of-service attack scheduling with energy constraint," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 3023–3028, Nov. 2015.

[6] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proc. 1st Int. Conf. High Confidence Netw. Syst.*, Beijing, China, 2012, pp. 55–64.

[7] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *Proc. 52nd IEEE Conf. Decis. Control*, Florence, Italy, Dec. 2013, pp. 1854–1859.

[8] Y. Li, L. Shi, and T. Chen, "Detection against linear deception attacks on multi-sensor remote state estimation," *IEEE Trans. Control Netw. Syst.*, to be published, doi: 10.1109/TCNS.2017.2648508.

[9] Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "SINR-based DoS attack on remote state estimation: A game-theoretic approach," *IEEE Trans. Control Netw. Syst.*, to be published, doi: 10.1109/TCNS.2016.2549640.

[10] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Syst.*, vol. 35, no. 1, pp. 46–65, Feb. 2015.

[11] A. S. Leong, S. Dey, and D. E. Quevedo, "Sensor scheduling in variance based event triggered estimation with packet drops," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1880–1895, Apr. 2017.

[12] K. Ding, S. Dey, D. E. Quevedo, and L. Shi. *Stochastic Game in Remote Estimation under DoS Attacks*. Accessed on May 24, 2017. [Online]. Available: http://www.ece.ust.hk/~eesling/2017_submitted.pdf

[13] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[14] S. Loyka, V. Kostina, and F. Gagnon, "Error rates of the maximum-likelihood detector for arbitrary constellations: Convex/concave behavior and applications," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1948–1960, Apr. 2010.

[15] A. Haurie, J. B. Krawczyk, and G. Zaccour, *Games and Dynamic Games*. Singapore: World Sci. Books, 2012.

[16] N. N. Vorob'ev, *Game Theory: Lectures for Economists and Systems Scientists*, vol. 7. New York, NY, USA: Springer, 2012.

[17] P. Billingsley, *Convergence of Probability Measures*. New York, NY, USA: Wiley, 1999.

[18] B. Sinopoli *et al.*, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.

[19] T. Başar and A. Haurie, *Advances in Dynamic Games and Applications*, vol. 1. New York, NY, USA: Springer, 2013.

[20] M. Pajic *et al.*, "Robustness of attack-resilient state estimators," in *Proc. ACM/IEEE 5th Int. Conf. Cyber Phys. Syst.*, Berlin, Germany, 2014, pp. 163–174.

[21] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[1] The algorithm is omitted due to limited space. Interested readers are suggested to see [9] for reference.