Remote State Estimation in the Presence of an Active Eavesdropper

Kemi Ding♯, Xiaoqiang Ren†, Alex S. Leong‡, Daniel E. Quevedo∗, and Ling Shi‡

Abstract—We consider remote state estimation in the presence of an active eavesdropper. A sensor forwards local state estimates to a remote estimator over a network, which may be eavesdropped by an intelligent adversary. Aiming at improving the eavesdropping performance efficiently, the adversary may adaptively alternate between an eavesdropping and an active mode. In contrast to eavesdropping, the active attack enables the adversary to sabotage the data transfer to the estimator and improve the data reception to itself at the same time. However, launching active attacks may increase the risk of being detected. As a result, a tradeoff between eavesdropping performance and stealthiness arises. We present a generalized framework for active eavesdropping and propose a criterion based on the packet reception rate at the estimator to evaluate the stealthiness of the eavesdropper. Moreover, the tradeoff is formulated as a constrained Markov decision process. After deriving a sufficient condition under which at least one stationary policy satisfies the stealthiness constraint and also bounds the eavesdropping performance, we develop an optimal attack policy for the eavesdropper and focus on the structural analysis of the optimal policy. Furthermore, numerical examples are provided to illustrate the developed results.

Index Terms—State estimation, CPS privacy, active eavesdropper, Markov decision process.

I. INTRODUCTION

In recent years, there has been an increasing interest in cyber-physical systems (CPSs), an integration of the physical world, communication networks and control systems [1]. CPSs encompass large-scale smart grids and intelligent transportation systems, to middle-scale critical infrastructures (e.g., water supply systems) and mini-size wearable medical devices. While CPSs bring unprecedented efficiency and productivity, the embedded information system, especially if wireless communications are used, exhibits vast vulnerabilities for adversaries to explore, which are widely studied.

Three main types of attacks on CPSs were identified in [2]: integrity attack, denial-of-service (DoS) attack, and eavesdropping. The first can stealthily manipulate the transmitted data packet; the availability of information is compromised by DoS attacks via jamming the communication channels. Both of these malicious attacks may further lead to severe system deterioration [3]. Moreover, systems may be exposed to eavesdropping attacks, wherein private information is disclosed. Notice that the broadcast nature (or use of a shared medium) of wireless communications makes it difficult to perfectly shield transmitted signals from intended recipients. Hence, by eavesdropping the communication channels between the sensors and the estimator in remote state estimation, adversaries may infer the state of the physical system and further violate the sensors’ (or users’) privacy. For example, in medical systems, a patient’s sensitive data (e.g., gender or health condition) may be exploited by third parties for the purpose of targeted advertisement or surveillance. In smart grids, the eavesdropper may infer the residents’ private information, including load taxonomy, from the fine-grained electricity consumption data provided by smart meters. From this information, the attacker can infer the daily schedules of the inhabitants and easily break into the house when nobody is at home [4]. The above situations show that eavesdropping attacks may cause severe economic losses or even threaten human survival.

In this paper, we focus on the confidentiality/privacy issue for remote state estimation. A sensor monitors a physical process and forwards its data to a remote estimator. At the same time, an eavesdropper attempts to infer the state information as accurately as possible. Besides the conventional (passive) eavesdropping attacks, we also consider a new type of eavesdropping attack, called active eavesdropping [5]. In this mode, the eavesdropper can not only overhear the transmitted signal sent by the sensor (i.e., passive mode), but also invade the communication channels between the sensor and the estimator by blocking the data transfer and thereby degrade the estimation accuracy of the remote estimator. As this performance deterioration may trigger the sensor to transmit data packets using a higher power or more frequently, the attackers’ packet arrival rate is increased and hence more accurate state information is captured. Note that the active eavesdropper launching jamming-style attacks is different from the DoS attacker. With active eavesdropping, the jamming is utilized to facilitate the eavesdropping performance as well. In contrast, a DoS attacker is indifferent to its own data reception, and merely seeks to degrade the estimator’s performance. Specifically, in addition to jamming, the active attacker is able to improve its data reception directly through pilot contamination (of which the details are demonstrated in [5]). Compared with passive eavesdropping, the active attack mode improves the eavesdropping capability of the eavesdropper significantly and...
undoubtedly poses a more serious privacy threat to remote state estimation in CPSs.

Related work: Most of the existing literature which consider active eavesdropping attacks has been confined to communication systems. In particular, [5] provided a unified framework of active eavesdropping in practical communication systems, and proposed an energy-efficient design of transmitting pilot-contamination signals to achieve a satisfactory eavesdropping performance. Wu et al. [6] evaluated the effect of active eavesdropping on a massive multi-input multi-output (MIMO) communication system. Moreover, they used the asymptotic achievable secrecy rate in information theory to measure the system’s information leakage rate and showed how the latter can be improved using communication diversity technologies. In some situations, the transmitter can jam the active eavesdropper and further prevent information disclosure via broadcasting artificial interference along with the information signal [7]. Therefore, an interesting conflict between the transmitter and the eavesdropper occurs. The work [8] modeled the resulting interaction as a two-player zero-sum game with the ergodic MIMO secrecy rate as the payoff function.

For remote state estimation in CPSs, we consider underlying dynamic systems rather than the stationary sources studied in the communications literature. This requires tailored methods based on systems control theory. Several relevant literature proposed system-resilient ways to avoid information leakage for dynamic systems through building a tradeoff between the systems’ utility and privacy. Sandberg et al. [9] considered a power system and investigated the tradeoff between the differential privacy and the mean distortion of the state estimation. Tsiamis et al. [10] proposed a secrecy mechanism which involves randomly withholding sensor information. Under this mechanism the tradeoff between the user’s utility and the control-theoretic secrecy was explored via an optimization approach. Subject to minimizing the estimation error covariance at the remote estimator while guaranteeing a certain confidentiality constraint, [11] examined an optimization problem with the objective function being a linear combination of the sensor’s and the attacker’s expected error covariances. The counter-measure for CPS privacy discussed in the previous works is limited to passively eavesdropping, and overlooks the severity of active eavesdropping attacks. The latter needs more thorough analysis, and is the subject of the current paper.

Main results: In this work, we consider remote estimation in the presence of an eavesdropper with the capability of either passively or actively eavesdropping. Different from the previous works on CPS privacy, we investigate the tradeoff from the perspective of the active eavesdropper, who designs the attack policy based on knowledge of the optimal transmission policy (optimal in the absence of attacks) adopted by the sensor. For convenience, we assume that the transmission policy is pre-determined, and hence the eavesdropper can compute the optimal transmission policy following [12], [13]. This worst-case attack analysis provides a benchmark in designing advanced privacy-protection mechanisms from the perspective of the system designer. The eavesdropper, with the objective to infer the state information accurately, alternates between active and passive modes following an optimal attack policy.

Indeed, without the consideration of power consumption, the optimal policy for the eavesdropper would be to launch active attacks continuously (since the active attacks can improve the eavesdropper’s packet reception rate). However, such a manner of attacks will also render it easy to be detected, isolated, and removed. Consequently, the eavesdropper faces a dilemma of establishing an attack policy that corresponds to the maximum state information leakage, while also avoiding detection by the sensor/estimator. In this work, we design an active eavesdropping policy for such an intelligent eavesdropper, to simultaneously achieve a satisfactory eavesdropping performance, whilst degrading the detection performance of the sensor/estimator.

Compared with related works in the literature, the main contribution of our current work is threefold:
(1) We propose a unified attack model based on [5], and evaluate the eavesdropping performance using the attacker’s averaged estimation error. Moreover, a new control-theoretic definition is introduced to characterize the detection performance of the sensor/estimator.
(2) Based on the aforementioned criterion, the tradeoff between eavesdropping performance and detection risk is formulated systematically by a constrained Markov decision process (CMDP) framework (Problem 1). We prove that, under some feasibility conditions (Assumption 2), there exists at least one stationary policy for this CMDP problem.
(3) We use the Lagrangian dynamic programming method to convert the CMDP problem into an unconstrained Markov decision process (MDP) parameterized by a Lagrange multiplier. In Theorem 4, we establish structural results of the optimal attack policies, which admit an intuitive interpretation.

Structure: The remainder of the paper is organized as follows: Section II contains mathematical models of the system, and presents the generalized attack structure of the active eavesdropper. The CMDP framework is presented in Section III, while the feasibility analysis and structural results are provided in Section IV. Some examples and concluding remarks are presented in Sections V and VI, respectively.

Notations: $R^n$ is the $n$ dimensional Euclidean space and $\mathbb{N}_0 = \{0, 1, \ldots\}$. $S_n^+$ (or $S_n^{++}$) is the set of $n$ by $n$ positive semi-definite matrices (or positive definite matrices). Let $\mathbb{N}$ denote the set of natural numbers. When $X \in S_n^+$ (or $X \in S_n^{++}$), we write $X \succeq 0$ (or $X > 0$). For functions $h, g, h \circ g$ is defined as the function composition $h(g(\cdot))$. $\mathbb{E}[\cdot]$ is the expectation of a random variable, $\text{Pr}(\cdot)$ denotes the trace of a matrix and $\text{Pr}(\cdot)$ refers to the probability. Notation $y_{\theta}^T$ denotes the sequence $(y_0, \ldots, y_k)$. $\mathcal{N}(\mu, \Sigma)$ stands for the multivariate normal distribution with mean $\mu$ and covariance matrix $\Sigma$. Let $\rho(X)$ represent the spectral radius of matrix $X$, i.e., $\rho(X) = \max \{ |\lambda_i(X)| \}$, where $\lambda_i(X)$ is an eigenvalue of $X$. The superscripts $\top$ and $\star$ stand for the matrix transposition and the optimal solution, respectively. Moreover, the superscripts “S”, “E”, and “A” stands for sensor, estimator and attacker, respectively. The superscripts “pa”, “ac” and “ar” in the Appendix represent passive, active and arbitrary. The subscripts “s” and “d” represent stationary and deterministic, and we use the tilde “∼” to highlight the active eavesdropping attacks. The indicator function is defined as
Suppose that \( x \) is observable and \( (A, \sqrt{Q}) \) is controllable. Then it is well known that the error covariance matrix \( P_k \) converges exponentially to a unique fixed point \( P \) of \( h \circ g \) [15]. Without loss of generality, the initial state of the local Kalman filter is assumed to be in steady state, i.e., \( P_0 = P \). From [16], the steady-state error covariance \( P \) has the following property:

**Proposition 1.** For \( 0 \leq t_1 < t_2 \), the following inequality holds:

\[
\text{Tr}(P) \leq \text{Tr}[h^{t_1}(P)] < \text{Tr}[h^{t_2}(P)].
\]

### B. Attacker Model

After obtaining the local estimate \( \hat{x}_k^S \) of the process state, the sensor transmits this value as a data packet to the remote estimator through a vulnerable channel in the presence of an eavesdropper. Previous studies on CPS security and privacy often overlook the ability of the eavesdropper and assume the attacker may merely “listen” to the channel (i.e., conventional passive eavesdropping). In this paper, we consider a more sophisticated and novel adversary, one with the dual capability of either passively eavesdropping or actively disrupt the communication link between the sensor and the estimator with the purpose of improving its signal reception rate.

In Fig. 1 the communication links \( H, G \) suffering from channel fading (caused by multi-path propagation or environmental obstacles) are modeled by additive-white-Gaussian-noise (AWGN) channels. First, we consider the scenario where the attacker is in passive eavesdropping mode. To measure the transmission quality, we introduce the signal-to-noise ratio (SNR) for the estimator at time \( k \): \( SNR_k^E = \frac{\|H\|_q}{\sigma^E} \), in which \( \|H\|_q \geq 0 \) and \( \sigma^E \) correspond to the time-invariant channel gain\(^1\), the power of the transmitted packet at time \( k \), and the channel noise variance. As a consequence of channel fading, data packets from the sensor may arrive unsuccessfully to the remote estimator leading to packet errors. From [17], the reception rate, measured in symbol-error-rate (SER), depends critically on the signal-to-noise ratio (SNR), and we adopt a general function to show their relationship: \( SER_k = d^E(SNR_k^E) \). Generally, the function \( d^E(\cdot) \) has different forms for various channel models, but it is non-increasing regardless of the channel characteristics and the modulation schemes. Analogously, the dropout rate for the attacker can also be characterized by another decreasing function \( d^A(SNR_k^A) \) with \( SNR_k^A = \frac{\|G\|_q}{\sigma^A} \), in which \( \|G\| \) and \( \sigma^A \) are the corresponding time-invariant average channel gain and channel noise variance.

Next, we show the effect of the active eavesdropping without regard for its implementation details. Denote by \( SNR_k^A \) and \( SNR_k^E \) the SNR of the estimator and the attacker under the active attacks. When switching to the active mode, the malicious attacker can not only intercept the message sent from the sensor to the estimator, but also be capable to

1) degrade the packet reception at the estimator by launching random jamming noise to interfere with the estimator’s channel, i.e., \( SNR_k^E \leq SNR_k^E \);
2) enhance the packet reception at the eavesdropper via pilot contamination [5], i.e., \( SNR_k^A \geq SNR_k^A \). Specifically, the attacker will send the same pilot signals at the reverse training phase to fool the sensor about channel estimation, which may further affect the precoding design at the sensor side and then benefit the packet reception at the attacker side.

Note that pilot contamination may also affect the packet

\(^1\)According to [17], (average) SNR is defined based on the proportion between the averaged received signal energy and the noise energy per bit slot. For simplicity, we assume the average channel gain \( \|H\|_q \) remains constant, while the channel matrix \( H \) is random and the sensor repeatedly sends pilot signals to estimate the channel for precoder design.
reception at the remote estimator. Regarding the practical implementation of the active attack and the specific formulations of $\widehat{SNR}_E^E$ and $\widehat{SNR}_A^A$, interested readers are referred to [5] for more details.

In this work we focus on the mode selection of the attacker. Let $a_k = 0$ or 1 be the attacker’s binary decision variable at time $k$ qualifying whether it chooses the passive or active mode to attack. As the symbol error can be detected by the channel coding method [18], the packet will be dropped if it contains any error (akin to the erasure channel [17]). We utilize two Bernoulli processes $\eta_{k}^{E+1}$ and $\eta_{k}^{A+1}$ to describe the success of packet arrival (i.e., $\hat{x}_k^E$) for the estimator and the attacker, respectively. Let $\eta_{k}^{E+1} = 0 \ (\eta_{k}^{A+1} = 0)$ denote the loss of packet for the estimator (the attacker), and $\eta_{k}^{E+1} = 1 \ (\eta_{k}^{A+1} = 1)$ otherwise. Let $Pr(\eta_{k}^{E+1} = 0|a_k, q_k)$ and $Pr(\eta_{k}^{A+1} = 0|a_k, q_k)$ denote the probabilities of packet dropout at the estimator and the attacker when the transmission power is given by $q_k$ and the attacker’s decision variable is set to $a_k$. We have

\[
Pr(\eta_{k+1}^{E} = 0|a_k, q_k) = d^{E}(a_k, q_k)
\]

\[
= d^{E} \left( (1 - a_k)SNR_k^E + a_kSNR_k^E \right),
\]

\[
(2)
\]

\[
Pr(\eta_{k+1}^{A} = 0|a_k, q_k) = d^{A}(a_k, q_k)
\]

\[
= d^{A} \left( (1 - a_k)SNR_k^A + a_kSNR_k^A \right),
\]

in which we use the same terms $d^E(a_k, q_k)$ and $d^A(a_k, q_k)$, in an abuse of notation, to denote the packet dropout probabilities, as the terms $SNR_k^E, SNR_k^E, SNR_k^A$ and $SNR_k^A$ only rely on $q_k$ and $a_k$. Notice that these two random variables are dependent since both the estimator’s and the attacker’s packet-receptions depend on $q_k$ and $a_k$. However, due to the independent receiving antennas adopted by the estimator and the attacker, the arrival of packets are conditionally independent, that is,

\[
Pr(\eta_{k+1}^{E}, \eta_{k+1}^{A}|a_k, q_k) = Pr(\eta_{k+1}^{E}|a_k, q_k)Pr(\eta_{k+1}^{A}|a_k, q_k).
\]

(3)

For notational simplicity, we shall use a counter to record the number of consecutive drops. Let $\tau_k^E$ (or $\tau_k^A$) denote the holding time for the sensor (or the attacker) to indicate the time duration from the last successful transmission time to time $k$: $
\tau_k^E \triangleq k - \max_{0 \leq j \leq k} \{l: \eta_{l}^{E} = 1\}, \quad \tau_k^A \triangleq k - \max_{0 \leq j \leq k} \{j: \eta_{j}^{A} = 1\}$. The possible values for $\tau_k^E$ (or $\tau_k^A$) form a countable set $\mathbb{N}_0 = \{0, 1, \ldots\}$.

From the above discussion, we summarize the attacker model as follows:

**Assumption 1. [Attack Model]** We make the following assumptions about the ability of the attacker:

(i) The attacker has perfect knowledge of the global channel state information (CSI), that is, $G, H$, the distribution of $n^A, n^E$ and the functions $d^A(\cdot), d^E(\cdot)$.

(ii) The attacker in active mode may operate in a full-duplex mode [5], that is capable of performing the jamming, pilot contamination and eavesdropping simultaneously.

(iii) Due to the additive jamming noise and pilot contamination, we have $SNR_k^E \leq SNR_k^E$, while $SNR_k^A \geq SNR_k^A$.

(iv) The attacker can obtain the information $\{\eta_k^E, \ldots, \eta_k^E\}$ through intercepting the feedback channel from the estimator to the sensor.

In practice, the sensor can adopt pilot-aided channel estimation techniques to obtain the channel-state information of the estimators’ channel (e.g., the channel parameter $H$) [19]. However, these conventional techniques make no attempt to prevent CSI leakage to the attacker [20], and the perfect knowledge of the channel parameters are often assumed to be known by the attacker in the literature of physical layer security [7], [21]. Moreover, Assumption 1-(i) is reasonable since it follows the Kerckhoff’s principle, that is, the security of a system should not rely on its obscurity. As shown in Fig 1, the estimator explicitly informs the sensor whether the data packet is received successfully or not through acknowledgements (ACKs) contained in a feedback loop (e.g., TCP-like protocols). Consequently, the online information $\{\eta_k^E, \ldots, \eta_k^E\}$ is known to the sensor, which can be explored to design energy-efficient transmission schedules [12], [22]. As for the attacker, according to Assumption 1-(iv) it can observe the occurrence of packet dropouts at channel $H$ by intercepting these feedback ACKs [13], which brings benefits to design more efficient attack policies.

C. Transmission Policy

Recall that in this work we investigate the design of optimal attack schemes under a pre-determined transmission policy adopted by the sensor. As investigated in previous works [12], [22], the optimal transmission policy for the sensor in a communication environment (in the absence of attacks) was proved to be deterministic and stationary. For simplicity, we consider a transmission energy set with two discrete levels, i.e., $\{e_0, e_1\}$ with $e_0 < e_1$. Denote a deterministic and stationary transmission policy by $\theta$, which maps the holding time of the sensor to the energy set, i.e., $\theta: \mathbb{N}_0 \rightarrow \{e_0, e_1\}$. In order to strike a balance between the remote estimation performance and the transmission power, the sensor will adopt an optimal threshold-type policy [22]:

\[
\theta_0 = (e_0, e_0, e_1, e_1, \ldots),
\]

in which $m$ is a nonnegative integer related to the estimator’s channel characteristic. It means that if the current holding time of the sensor $\tau_k^E$ is less than $m$, then the sensor transmits the data packet $\hat{x}_k^S$ with the lower power level $p_k = e_0$, otherwise $p_k = e_1$. The intuition behind this threshold-type policy is that when $\tau_k^E$ is large, the sensor will spend great effort on transmission as $\hat{x}_k^S$ contains valuable information for remote estimation.
D. Remote Estimation

Let \( \hat{x}_E^k \) and \( \hat{x}_A^k \) denote the estimator’s and the attacker’s estimate of the state \( x_k \), respectively. After receiving the data packets, the estimation processes are the same for the estimator and the attacker: once receiving \( \hat{x}_E^k \) correctly (i.e., \( \eta_E^k = 1 \) or \( \eta_A^k = 1 \)), the estimator (or the attacker) synchronizes its estimate \( \hat{x}_E^k \) (or \( \hat{x}_A^k \)) with it; otherwise, it simply predicts the estimate with the previous estimate \( \hat{x}_E^{k-1} \) (or \( \hat{x}_A^{k-1} \)) using the system model (1). Hence, we have

\[
\hat{x}_E^k = \eta_E^k \cdot \hat{x}_E^{k-1} + (1 - \eta_E^k) \cdot A \hat{x}_E^{k-1},
\]

\[
\hat{x}_A^k = \eta_A^k \cdot \hat{x}_A^{k-1} + (1 - \eta_A^k) \cdot A \hat{x}_A^{k-1}.
\]

Define \( T_k^E \) as the set of all the information available at the remote estimator up to time \( k \), i.e.,

\[
T_k^E = \{ \eta_1^E \cdot x_1, \eta_2^E \cdot x_2, \ldots, \eta_k^E \cdot x_k \} \cup \{ \eta_1^A, \eta_2^A, \ldots, \eta_k^E \}. \tag{4}
\]

Similarly, we define the information set for the attacker as \( T_k^A \) by replacing the superscript “E” in (4) with “A”. To describe the estimation accuracy for the estimator (or privacy performance for the attacker), we introduce the error covariance matrix \( P_k^E \) (or \( P_k^A \)), as in [23], [24]:

\[
P_k^E \triangleq \mathbb{E}[(x_k - \hat{x}_E^k)(x_k - \hat{x}_E^k)^\top | T_k^E] = \eta_k^E \mathbf{P} + (1 - \eta_k^E) h(P_{k-1}^E),
\]

\[
P_k^A \triangleq \mathbb{E}[(x_k - \hat{x}_A^k)(x_k - \hat{x}_A^k)^\top | T_k^A] = \eta_k^A \mathbf{P} + (1 - \eta_k^A) h(P_{k-1}^A).
\]

Based on the definition of holding time, we have \( P_k^E = h^E_k(\mathbf{P}) \) and \( P_k^A = h^A_k(\mathbf{P}) \). The iterations of the holding times are as follows:

\[
\tau_k^E = (1 - \eta_k^E)(\tau_{k-1}^E + 1), \quad \tau_k^A = (1 - \eta_k^A)(\tau_{k-1}^A + 1),
\]

which suggests that \( \tau_k^E \) and \( \tau_k^A \) depend on the holding times at time \( k - 1 \), \( \eta_k^E \) and \( \eta_k^A \). Note that the transmission and attack actions \( a_k, q_k \) determine the probability of successful packet arrivals according to (2). Obviously, the pair of holding times \( \tau_k^E \) and \( \tau_k^A \) satisfies the Markov property, since given the current \( x_k, \tau_k^E, \tau_k^A, a_k, q_k \), its future conditional distribution probability is independent of the historical sequence \( \mathcal{H}_{k-1} \triangleq \{ \tau_1^E, \ldots, \tau_{k-1}^E, \tau_1^A, \ldots, \tau_{k-1}^A, a_1, \ldots, a_{k-1}, q_1, \ldots, q_{k-1} \} \):

\[
\mathbb{P}(\tau_{k+1}^E, \tau_{k+1}^A | \tau_k^E, \tau_k^A, a_k, q_k, \mathcal{H}_{k-1}) = \mathbb{P}(\tau_{k+1}^E, \tau_{k+1}^A | \tau_k^E, \tau_k^A, a_k, q_k).
\] \tag{5}

E. Attacker Detector Model

Compared with the passive case, the advent of active eavesdropping brings great performance improvement for the attacker, however, it also inevitably increases the risk of being detected. Next, we formulate the detectability of such active attacks in a mathematical way.

As mentioned previously, the active attacker may launch random jamming signals to corrupt the packet transmission from the sensor to the estimator, which may lead to anomalous transmission failure (i.e., packet dropout). In order to evaluate the anomaly and further detect the presence of active attacks, we use the notion of packet drop ratio (PDR), see [25]. To be specific, PDR is the ratio of the number of packets that are dropped to the number of received packets during a time window of length \( T \):

\[
PDR = \frac{\sum_{k=1}^{T} \mathbf{1}_{\eta_k^E \neq 1}}{T} = \frac{\sum_{k=1}^{T} \mathbf{1}_{\eta_k^E \geq 1}}{T}, \tag{6}
\]

which can be calculated either by the estimator, or by the attacker based on the online information set \( \{ \eta_1^E, \ldots, \eta_k^E \} \). Notice that a malicious attacker with an aggressive active attack policy (e.g., jamming all the time), which causes noticeable damage to the network quality, may be detected by the sudden increase of the PDR. However, the network congestion or channel noise may also lead to PDR degradation. According to [25], we can adopt a simple yet popular threshold mechanism based on the PDR value to distinguish an active attack from the aforementioned normal network failures. That is, if \( PDR > D \) (where \( D < 1 \) is the threshold parameter subject to the estimator’s channel condition), the intrusion detection system will be alerted, otherwise no anomaly is detected. With such a mechanism in place, the attacker should elaborate its attack tactics carefully to avoid triggering the detection condition (namely, make sure \( PDR \leq D \)).

III. FORMULATION AS A CONSTRAINED MDP

In this section, adopting an attacker’s perspective, we formulate the stealthy attack policy design problem as a constrained Markov decision process.

A. MDP Formulation

We mathematically describe the MDP by the quadruplet \( \mathcal{S}, \mathcal{A}, T, c(\cdot) > 0 \) with details as follows:

1) State Space: The state at time \( k \) is defined as the pair of stopping times: \( s_k \triangleq (\tau_k^E, \tau_k^A) \), which takes values in the countable state space \( \mathcal{S} \triangleq \mathbb{N}_0 \times \mathbb{N}_0 = \{ (0,0), (0,1), (1,0), \ldots \} \).  

2) Action Space: The action space is \( \mathcal{A} \triangleq \{ 0,1 \} \) with \( a_k \in \mathcal{A} \) representing the binary action to be selected by the attacker at time \( k \). Let \( a_k = 1 \) and \( a_k = 0 \) denote that the attacker chooses the active mode and eavesdropping only, respectively. Let \( \mathcal{P} \) denote the space of measures on \( \mathcal{B}(\mathcal{A}) \), the Borel subsets of \( \mathcal{A} \), endowed with the weak topology (hence it is a Polish space, a complete and separable metric space). The mixed strategy for the attacker, denoted by \( \pi_k \in \mathcal{P} \), is a probability distribution over the pure action space \( \mathcal{A} \).

The randomized decision rule for the attacker is a sequence of mixed strategies \( \pi \in \Pi \) with \( \Pi \) representing the randomized policy set, and \( \pi_k^\infty \triangleq (\pi_1, \pi_2, \ldots, \pi_k, \ldots) \) with \( \pi_k \in \mathcal{P} \). Among the admissible strategies for the attacker, we focus on randomized stationary ones, which can be found through offline computations and are easy to implement. According to [26], there is no loss of optimality in restricting attention to attack strategies of the stationary form. Let \( \Pi_s \) denote the family of all randomized stationary strategies with their stochastic kernels satisfying:
1. \( \pi_k \) is a mapping from the state space \( S \) to the measure space \( \mathcal{P} \), which means that the action is a (possibly) random function of the current state \( s \in S \). With slight abuse of notation, we denote by \( \pi_k(a|s) \) the probability with which a particular action \( a \in A \) at time \( k \) will be played in state \( s \); 2. for each \( s \in S \), there exists a probability measure \( \pi_k(s) \in \mathcal{P} \) such that \( \pi_k(s) = \pi_k(s) \cdot \forall k \geq 0 \), namely, the mapping \( \pi_k : S \to \mathcal{P} \) is time-independent.

One special case of the randomized stationary policy are deterministic one, where \( \pi \) is a deterministic function of current state, \( \pi : S \to A \). The corresponding deterministic policy set is denoted by \( \Pi_d \), and \( \Pi_d \subseteq \Pi_e \).

3) Transition Probability: As the sensor can only observe the feedback ACKs \( \{\theta^E_1, \ldots, \theta^E_k\} \) sent from the estimator, its transmission policy depends on \( \theta^E_1 \) instead of the pair of holding times \( s_k \). Given the control policies \( (\theta, \pi) \) played by the sensor and the attacker, the random process \( s_k \) becomes a controlled Markov chain as depicted in Figure 2. Let \( s_k = (i_1, i_2) \), then the stationary transition probabilities are computed from (3) and (5):

\[
\begin{align*}
\Pr(s_{k+1}|s_k, a_k, q_k) = \\
\begin{cases}
& d^E(a_k, q_k)d^A(a_k, q_k), & \text{if } s_{k+1} = (i_1 + 1, i_2 + 1); \\
& d^E(a_k, q_k)[1 - d^A(a_k, q_k)], & \text{if } s_{k+1} = (i_1 + 1, 0); \\
& [1 - d^E(a_k, q_k)]d^A(a_k, q_k), & \text{if } s_{k+1} = (0, i_2 + 1); \\
& [1 - d^E(a_k, q_k)][1 - d^A(a_k, q_k)], & \text{if } s_{k+1} = (0, 0).
\end{cases}
\end{align*}
\]

As a shorthand notation, we denote by \( d^E(0) \triangleq d^E(q_k = e_0, a_k = 0) \) and \( d^A(0) \triangleq d^A(q_k = e_0, a_k = 0) \) resp. the packet-dropping rate for the estimator (the attacker resp.) when a passive attack is launched on the low-power packet. If a high transmission power \( e_1 \) is used, then “0” is replaced by “1” in the dropout-rate notation. Moreover, a tilde “~” stands for the active attack, e.g., \( d^E(1) \triangleq d^E(q_k = e_1, a_k = 1) \) and \( d^A(1) \triangleq d^A(q_k = e_1, a_k = 1) \). Notice that \( q_k \) relies on \( \theta^E_1 \), and then the transition of \( \tau^E_k \) depends heavily on \( \theta^E_k \) (as demonstrated in Figure 2). Hence, the Markov chain \( s_k \) is no longer a simple combination of \( \tau^E_k \) and \( \tau^A_k \). This may cause difficulty when investigating the feasibility of attack policies and the solution structure.

4) Immediate Cost: Let \( c(\cdot) \) denote the immediate cost function for the attacker with \( c : S \times A \to \mathbb{R} \). As the attacker attempts to obtain an accurate state estimate \( \hat{x}_k \) without much energy investment, its immediate cost function is chosen as:

\[
c(s_k, a_k, q_k) \triangleq \text{Tr}[P_k^A] + \beta q_k,
\]

in which coefficient \( \beta \) represents the proportion of the energy term in the cost. For the attacker, \( \beta \) characterizes the tradeoff between its estimation performance improvement and the energy consumption by adopting the active attacks. The stealthiness requirement will be taken into account via a constraint of the form \( \text{PDR} \leq D \), see Problem 1 below.

**B. Problems of Interest**

As discussed previously, the sensor transmits the local estimate \( \hat{x}_k^E \) to the remote estimator under the risk of information leakage. In this work we investigate an active eavesdropper, which poses a more serious threat to CPSs than that caused by traditional passive eavesdropping [7]. The active attack is specified in Section II-B. Based on the observed data packet, the active eavesdropper will compute an estimate \( \hat{x}_k^A \) of the process state. Considering an infinite time-horizon, the eavesdropping performance, under a pair of transmission and attack policies \((\pi, \theta)\), is described by the average expected cost:

\[
J_{s_0}(\pi, \theta) = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}_{\pi, \theta}[c(s_k, a_k, q_k)|s_0],
\]

in which \( s_0 \) is the initial state. Recall that the sensor adopts an optimal deterministic stationary and threshold-type transmission policy \( \theta_0 \) as discussed in Section II-C. Moreover, the attacker can calculate the optimal transmission policy from Assumption 1 (i) and based on results in [12] and [22]. In order to infer the state \( x_k \) as accurately as possible, the attacker, using knowledge of the transmission policy \( \theta \), seeks an optimal schedule \( \pi \) to minimize \( J_{s_0}(\pi, \theta) \).

Beyond minimizing \( J_{s_0}(\pi, \theta) \), the attacker should design \( \pi \) carefully to avoid detection by the sensor/estimator. Note that PDR in (6) describes the packet drop frequency during a given finite time window. To characterize the detection criterion in the infinite-time horizon, we adopt the asymptotic PDR as follows:

\[
D_{s_0}(\pi, \theta) = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}_{\pi, \theta}[	au_k^E \geq 1|s_0, \pi, \theta] = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}_{\pi, \theta}[	heta_k^E \geq 1|s_0].
\]

In conclusion, it is of interest for the attacker to devise a stealthy attack scheme capable of achieving a high eavesdropping performance. Specifically, with the knowledge of the optimal transmission policy, the attacker wishes to find an attack policy such that the average eavesdropping cost is minimized under a detection constraint:

**Problem 1.**

\[
\min_{\pi \in \Pi} J_{s_0}(\pi, \theta_0), \quad \text{s.t.} \quad D_{s_0}(\pi, \theta_0) \leq D,
\]

in which \( D \in \mathbb{R}^+ \) is the detection threshold. The optimal stationary strategy will be denoted as \( \pi^* \).

In the following sections, we study the design problem with the aid of MDP theory, and analytically develop a simple and optimal dynamic schedule. We shall answer the following questions: Under what circumstances does there exists an optimal policy? What properties does the policy satisfy?
IV. MAIN RESULTS

In this section, we first study the stability condition under which the averaged expected cost will converge, and the feasibility condition of the optimal attack schemes. The latter guarantees the existence of at least one stationary policy satisfying the stealthiness constraint. Next, analyse the structure of the optimal policy by reformulating the original CMDP as an unconstrained MDP with the aid of the Lagrange multiplier method.

A. Feasibility of Constrained MDP

The Markov chain process $s_k$ is shown in Fig. 2. As the attacker might have infinite cost (namely, when $\tau_k^A \rightarrow \infty$), the stability issue is a precondition for the existence of an optimal attack policy for Problem 1. Moreover, the role of the stealthiness constraint in Problem 1 is to ensure reasonable and undetectable attack policies (The policies satisfying the detection constraint are called “admissible policies”). According to [27], an optimal solution for Problem 1 among these admissible policies can be achieved by a stationary one, which is also independent of the initial state. This implies that we may restrict our search for an optimal policy to the dominating class $\Pi_s$ (introduced in Section III-A2). We will present in Theorem 1 a condition on the dropout rates (i.e., $d^E(0), d^E(1), d^A(1)$ and $d^A(0)$ and so on, which are defined after (7)) and the average detection constraint, under which there exist stationary admissible policies inducing a positive recurrent Markov chain with bounded average cost.

Some assumptions are necessary to obtain the existence of an optimal stationary policy in the countable state space case.

Assumption 2. There exists a convex combination of the two points $(\bar{d}^E(0), \bar{d}^E(1), \bar{d}^A(1))$ and $(d^E(0), d^E(1), d^A(1))$, denoted by

\[(\bar{d}^E_0, \bar{d}^E_1, \bar{d}^A_1) = \alpha(d^E(0), \bar{d}^E(1), \bar{d}^A(1)) + (1 - \alpha)(d^E(0), \bar{d}^E(1), d^A(1)) \quad \alpha \in [0, 1],\]

such that the following conditions are satisfied:

(i) $\frac{1}{D} \frac{1 - \alpha}{1 - \alpha} \geq 1 - D$;

(ii) $\bar{d}^A_1 < \frac{1}{\rho(A)}$.

Remark 1. Assumption (i) restricts the packet-dropout rate of the estimator’s channel ($\bar{d}^E_0$ and $\bar{d}^E_1$) under passive/active attacks to guarantee the attack policy is stealthy. The left-hand-side of inequality in Assumption (i) is non-increasing in $\bar{d}^E_1$. If Assumption (i) is not satisfied (which means $\bar{d}^E_0$ and $\bar{d}^E_1$ are large enough), then the number of packet drops for the remote estimator during a specific time window may be increased and the intrusion detection system will be easily alerted. Assumption 2 (ii) restricts the packet-dropout rate of the eavesdropper employing active attacks. This type of assumption is common in state estimation problems with intermittent observations [28]. Intuitively, for unstable systems, if (ii) is not satisfied, then the eavesdropper using active attacks may still not receive enough packets leading to an unbounded estimation error (or the averaged expected cost $J_{s_0}(\pi, \theta) = \infty$). An example illustrating Assumption 2 is shown in Figure 3, in which $\frac{1}{\rho(A)} = 0.8$, $1 - D = 0.2$ and the threshold $m = 3$, and the possible values of $(\bar{d}^E_0, \bar{d}^E_1, \bar{d}^A_1)$ forms a cylinder with red edges. Moreover, we consider the extreme case $(\bar{d}^E(0), \bar{d}^E(1), \bar{d}^A(1)) = (0, 0, 1)$ and $(d^E(0), d^E(1), d^A(1)) = (1, 1, 0)$. As shown in Figure 3, the convex combination of the two points is the blue line. One may say that there exists a constant $\alpha \in [0, 1]$ such that the inequalities in Assumption 2 hold, since a subset of the blue line is contained in the cylinder (i.e., the dashed blue line).

Theorem 1. If Assumption 2 holds, then there exists a feasible stationary schedule that satisfies the stability condition, namely, $\exists \pi \in \Pi_s$ such that for any initial state $s_0 \in S, J_{s_0}(\pi, \theta_0) < \infty$ and $D_{s_0}(\pi, \theta_0) \leq \bar{D}$.

Proof: See Appendix A.

Notice that from the assumption and the proof (Appendix A), we can construct a feasible policy such that the two random processes (i.e., $\tau_k^E$ and $\tau_k^A$) are independent. However, the optimal policy may be dependent on the state and hence the analysis of performance becomes difficult. Next, we shall investigate the structure of the optimal policy, and its implementation.

B. Lagrange Dynamic Programming

As mentioned previously, the Lagrangian approach allows us to transform a constrained MDP problem into an equivalent minmax unconstrained problem. Thereby, we use the Lagrange multiplier method to reformulate the constrained MDP as a parameterized unconstrained MDP, and adopt the Lagrangian multiplier to design policies that achieve the stealthiness condition. For any Lagrange multiplier $\lambda \geq 0$, we define the immediate Lagrangian cost at time $k$ by

$$r^\lambda(s_k, a_k, q_k) = c(s_k, a_k, q_k) + \lambda 1_{t^E \geq 1}.$$  

Accordingly, define the Lagrangian average cost for a policy $\pi$ as

$$J_{s_0}^\lambda(\pi, \theta) = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} E_{\pi, \theta}[r^\lambda(s_k, a_k, q_k)|s_0].$$
Since the strong duality holds based on Theorem 12.7 in [27], the optimal cost function of the primal problem (i.e., Problem 1) satisfies

\[ J_{s_0}(\pi^*, \theta_0) = \sup_{\lambda \geq 0} \min_{\pi \in \Pi_{s_0}} J^\lambda_{s_0}(\pi, \theta_0) - \lambda D. \]

Obviously, the solution of the aforementioned optimization problem relies on the solvability of the corresponding average cost unconstrained MDP problem for given \( \lambda \geq 0 \), defined as follows:

**Problem 2.**

\[ \min_{\pi \in \Pi_{s_0}} J^\lambda_{s_0}(\pi, \theta_0), \]

of which \( \pi^{2*}_{\lambda} \) denotes the average cost optimal policy.

The optimal stationary policy always exists for the unconstrained MDP problem with average cost if the state space \( S \) is finite. However, this is no longer guaranteed the case when \( S \) is denumerably infinite (e.g., \( S = \{0, 1, \ldots\} \)). Indeed, an optimal policy of any sort (stationary or non-stationary) may not even exist. In the case of average cost MDPs, one can still derive the existence and the structure of the optimal policy by viewing the average cost MDP model as the limit of a sequence of discounted cost MDPs with discount factors \( \xi \rightarrow 1 \). Define the discounted Lagrangian cost associated with policy \( \pi \) and discount factor \( \xi \in [0, 1) \) by

\[ J^\lambda_{s_0}(\pi, \theta_0) = \limsup_{T \to \infty} \mathbb{E}_{\pi, \theta_0} \sum_{k=1}^{T} \epsilon_k^\lambda(s_k, a_k, q_k)|s_0|, \]

and the related discounted cost MDP problem as

**Problem 3.**

\[ \min_{\pi \in \Pi_{s_0}} J^\lambda_{s_0}(\pi, \theta_0), \]

of which \( \pi^{3*}_{\lambda, \xi} \) is the optimal policy for the \( \xi \)-discount cost.

In particular, we present the results of this section following the steps below:

1. Analyze the existence and structure of the (discounted cost) optimal policy \( \pi^{3*}_{\lambda, \xi} \) of Problem 3 for any \( \lambda \geq 0 \) and \( \xi \in [0, 1) \);
2. Discuss the relationship between \( \pi^{3*}_{\lambda, \xi} \) and the optimal policy of Problem 2 (i.e., \( \pi^{2*}_{\lambda} \));
3. Compute the optimal policy \( \pi^* \) of Problem 1 by the aid of Lagrangian multiplier and \( \pi^{2*}_{\lambda} \).

First, we discuss the solution existence and structure of the discounted cost MDP (Problem 3) for a given Lagrangian multiplier \( \lambda \). We denote by \( V_{\lambda, \xi}(s) : S \to \mathbb{R} \) the expected total discounted reward function of the optimal policy, which satisfies Bellman’s equation of optimality:

\[ V_{\lambda, \xi}(s) = \min_{a \in A} \left\{ r(s, q, a) + \xi \sum_{s' \in S} \mathbb{P}(s'|s, q, a) V_{\lambda, \xi}(s') | \theta_0 \right\}. \]

Let \( K \) be the set of all real-valued functions on \( S \). With the knowledge of the transmission policy \( \theta_0 \), we define an operator \( C : S \times A \times K \to \mathbb{R} \) as

\[ C(s, a, f) \triangleq \sum_{\tilde{s} \in S} \mathbb{P}(\tilde{s}|s, q, a) f(\tilde{s}), \]

in which \( f(\cdot) \in K \). Based on (9), we define a dynamic programming operator \( W_{\xi} : K \to K \) as

\[ W_{\xi} f(s) \triangleq \min_{a \in A} \left[ r(s, q, a) + \xi C(s, a, f) \right], \]

Moreover, we define a partial order on the state space \( S \) before discussing the monotone structure of the optimal policy \( \pi_{\lambda, \xi}^{3*} \). Let \( s, s' \in S \), \( s = (i, j) \) and \( s' = (i', j') \). We say that \( s \leq s' \) if \( i \leq i' \) and \( j \leq j' \). Let \( s \lor (\land) s' \) denote the join (meet) on the lattice \( (\leq, S) \). Notice that the threshold \( m \) separates the state space \( \mathbb{N}^2 \) into two disjoint state regions \( S_{i \leq m} \) and \( S_{i > m} \), where we denote \( s = (i, j) \) with \( 0 \leq i \leq m \) as \( S_{i \leq m} \), and \( S_{i > m} \), otherwise.

**Theorem 2.** Suppose \( 0 \leq \xi < 1 \) and \( \lambda \geq 0 \). There exists an optimal stationary and deterministic policy \( \pi_{\lambda, \xi}^{3*} \) such that (9) holds. Additionally, there exists a switching curve \( z(s) = 0 \) which further divides each of the state regions \( S_{i \leq m-1} \) and \( S_{i \geq m+1} \) into two disjoint parts such that

\[ \pi_{\lambda, \xi}^{3*}(s) = \begin{cases} 1, & \text{if } z(s) \geq 0; \\ 0, & \text{if } z(s) < 0. \end{cases} \]

**Proof:** See Appendix B.

**Remark 2.** While focusing on the monotonicity of the optimal attack policies, Theorem 2 does not provide the specific representation of the switching curve \( z(s) \), as it is difficult to obtain and its property is closely related to the system parameters. Notice that the threshold of the transmission policy adopted by the sensor divides the state space into two parts, and the structure of the attack policy along the line \( i = m \) is intricate. Its structural analysis needs more rigorous conditions on the spectral radius of the system dynamic matrix and packet drop ratios. Readers are suggested to refer to Figure 4(b) in Section V-B for an explicit understanding of the structure of the optimal attack policy. The monotone structure of the optimal attack policies follows the intuition: when the eavesdropper has an inaccurate estimate of the state (or a large \( \tau_{\Omega}^k \)), it is prone to adopt active attacks to improve its eavesdropping performance; while if the remote estimator does not receive data packets for a long time (corresponding to an increase of \( \tau_{\Omega}^k \)), rather than risk being discovered, the eavesdropper may adaptively switch from the active mode to the passive one. Consequently, by exploiting its local information and the feedback information from the remote estimator, the eavesdropper can achieve the tradeoff between eavesdropping performance and a favourable stealthiness.

Many works have developed the theory for relating the optimal solution of the \( \xi \)-discount cost Problem 3 to the average cost Problem 2, see, e.g., [26] and [30]. Specifically, under suitable conditions given in [31], the average cost optimal policies inherits the structure of the discounted cost optimal policies with discount factors approaching to 1. Let \( \{\xi_k\} \triangleq \{\xi_1, \xi_2, \ldots, \xi_l, \ldots\} \) be some increasing sequence of discount factors converging to 1. The associated sequence of discount optimal stationary policies is denoted by \( \{\pi_{\lambda, \xi_k}^{3*}\} \), which converges to a limit \( \pi_{\lambda}^{3*} \). This limit is a candidate for the optimal policy for Problem 2. In the following theorem,
we demonstrate that it indeed corresponds to the average cost optimal policy \( \pi_1^{2,*} \).

**Theorem 3.** Any stationary deterministic policy \( \pi_0 \) obtained as mentioned above, is an average cost optimal policy for Problem 2 for given \( \lambda \). That is, for any initial state \( s_0 \), \( J_{s_0}^{\lambda}(s, \pi_0) \leq J_{s_0}^{\lambda}(s, \pi) \), \( \forall s \in S, \pi \in \Pi \). Moreover, there exists at most one \( D \) and \( \pi \) satisfying the constraint. Moreover, it is easy to verify (C2) in the state \( s \). That is, for any initial state \( s \) we demonstrate that it indeed corresponds to the average cost optimal policy \( \pi_1^{2,*} \).

**Theorem 4.** There exists a stationary attack policy \( \pi^* \) that is the optimal solution of the primal CMDP problem, such that \( \pi^* \) is a randomized mixture of two deterministic policies \( \pi_{\lambda_1}^{2,*} \) and \( \pi_{\lambda_2}^{2,*} \) chosen with probability \( 1 - \varphi \) and \( \varphi \). Both deterministic policies \( \pi_{\lambda_1}^{2,*} \) and \( \pi_{\lambda_2}^{2,*} \) satisfy the structural condition of Theorem 2. Furthermore, there exists at most one state \( s \) such that \( \pi_{\lambda_1}^{2,*} \neq \pi_{\lambda_2}^{2,*} \).

**Proof:** The result follows directly from [32, Theorem 4.4] and [33, Theorem 1].

Denote \( s_1 \) to be the state such that \( \pi_{\lambda_1}^{2,*} \neq \pi_{\lambda_2}^{2,*} \). According to \( D_{s_0}(\pi^*) = D \) and the theory of regenerative processes [34, Proposition 5.9], we can compute the randomization parameter for the optimal mixed policy. This parameter is given by

\[
\varphi = \frac{I_1}{I_1 + I_2} \frac{D - D_{s_1}(\pi_{\lambda_1}^{2,*})}{D_{s_0}(\pi_{\lambda_1}^{2,*}) - D},
\]

in which \( I_1 \) and \( I_2 \) are the probability of state \( s_1 \) in the stationary distribution of Markov chains induced by the two deterministic policies \( \pi_{\lambda_1}^{2,*} \) and \( \pi_{\lambda_2}^{2,*} \), respectively. The limiting distribution can be easily computed based on the transition probability (7).

Last, we discuss the structural result for the primal CMDP with different constraints. Note that the asymptotic PDR introduced in (8) is closely related to the estimation performance at the estimator. Under given stationary transmission and attack policies \((\pi, \theta)\), a steady state distribution of \( P_k^E = h_k^E(\mathcal{P}) \) exists. The asymptotic PDR is a probabilistic statement of the estimator’s estimation accuracy [35] since \( D_{s_0}(\pi, \theta) = 1 - \Pr\left( \lim_{k \to \infty} P_k^E \leq \mathcal{P} \right) \). Interestingly, the results in Theorem 4 can be extended to a scenario where the detection constraint is replaced by a function of the error covariance, denoted by \( f(\tau_k^E) \). In particular, this scenario includes the probabilistic estimation-performance criterion (or the asymptotic PDR) with corresponding \( f(\tau_k^E) = 1 + \epsilon \geq 1 \) and the expected value of error covariance criteria with \( f(\tau_k^E) = \text{Tr}[P_k^E] = \text{Tr}[h_k^E(\mathcal{P})] \). The corresponding structural results is summarized as follows:

**Corollary 1.** Consider the MDP \( \mathcal{P} \) with different constraints, which has the following general form:

\[
D_{s_0}(\pi, \theta) \triangleq \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} E[f(\tau_k^E)|\pi, \theta, s_0],
\]

in which \( f(\cdot) \) is a non-decreasing general function in \( \tau_k^E \). Then the optimal stationary attack policy for this CMDP problem still has the threshold-like structure as mentioned in Theorem 4.

**Proof:** We can still use the Lagrangian dynamic programming method to transform the primal CMDP to an unconstrained one, but with a different immediate Lagrangian cost:

\[
r^E(s_k, a_k, q_k) = c(s_k, a_k, q_k) + \lambda f(\tau_k^E).
\]

Then, the structural analysis of the optimal policy for this new CMDP is similar to the aforementioned analysis. Details are left to the interesting reader.

**C. Practical Implementation**

As the state space \( S \) is infinite, from the perspective of practical implementation, we need to truncate the state space without sacrificing much performance. Indeed, we approximate the infinite set \( \mathcal{N} \) by a finite size set \( \mathcal{N}_0^t \triangleq \{0, 1, \ldots, N\} \), and hence the truncated state space is denoted by \( S^t \triangleq \mathcal{N}_0^t \times \mathcal{N}_0^t \). The truncation operator assumes that holding times \( \tau_k^E \) or \( \tau_k^B \) larger than \( N \) will be regarded as \( N \). As the optimal policy induces a positive recurrent Markov chain, based on Lemmas 1 and 2, we can prove that the probability of the state \( s_k \) belonging to the set of states \( \{N + 1, N + 2, \ldots\} \) will go to zero as the approximation parameter \( N \) goes to infinity [36]. It means that the effect of the truncation operator on the system performance can be ignored. The constrained MDP with finite state space can be solved by some conventional MDP numerical algorithms, for example, linear programming [27], primal-dual methods [37] and reinforcement learning [38]. Moreover, some modification taking the structure of the optimal policies into account can be used to accelerate the algorithm speed [39].
### Table I
SUMMARY FOR PARAMETERS

| Estimator | \(d^E(0)\) | \(d^E(1)\) | \(\tilde{d}^E(0)\) | \(\tilde{d}^E(1)\) | \(m\) |
|-----------|-------------|-------------|-------------------|-------------------|-----|
| 0.4       | 0.2         | 0.6         | 0.4               | 4                 |
| Attacker  | \(d^A(0)\) | \(d^A(1)\) | \(d^A(0)\) | \(d^A(1)\) | \(\beta\) |
| 0.7       | 0.3         | 0.35        | 0.3               | 2                 |

### V. Simulation and Example

In this section, we illustrate the structure of the optimal policy (Theorem 4) via numerical examples. In particular, we consider the following vector system with parameters \(A = \begin{pmatrix} 1 & 0.5 \\ 0 & 1.25 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 \end{pmatrix}, Q = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}\) and \(R = 0.5\), where the steady-state error covariance is \(P = \begin{pmatrix} 0.38 & 0.28 \\ 0.28 & 1.69 \end{pmatrix}\).

The values of parameters such as packet-dropout rate under passive/active eavesdropping, the threshold for transmission policy \(\theta_0\) and the weight \(\beta\) are summarized in Table I. Moreover, the constraint value is \(\bar{D} = 0.50\) and the truncation operator uses \(N = 12\). First, we verify the satisfaction of Assumption 2: if \(0 < \alpha < 0.56\), then the inequality in (1) is satisfied; Since \(\tilde{d}^A(1) = 0.3 < 0.64\), (2) always holds. Assume that at the beginning both the estimator and the attacker receive the data packet \(\tilde{x}_0\) (namely, the initial state is \((0, 0)\)).

#### A. Structure of the Optimal Policy for Problem 2

In this subsection, we present the structure of the unconstrained optimal attack policy \(\pi_2^{2,\star}\) under different Lagrangian multipliers, while other parameters (i.e., threshold \(m\) and truncation \(N\)) remain unchanged. The unconstrained optimal attack policies under different Lagrangian multipliers are plotted in Figures 4(a) and 4(b), 5(a) and 5(b):

1. Figure 4(a) depicts the unconstrained optimal policy \(\pi_2^{2,\star}\) obtained by relative value iteration for \(\lambda = 13.39\). One may see that \(x_k^E = 3\) (which corresponds to the threshold value \(m = 3\) of the transmission policy \(\theta_0\)) splits the state space into two parts and the unconstrained optimal attack policy in each part has the threshold structure stated in Theorem 3. Under different Lagrangian multipliers, the optimal policy has the threshold structure over the whole state space as shown in Figures 5(a) and 5(b) with \(\lambda = 1.00\) and \(\lambda = 41.00\), respectively.

2. By comparing Figures 5(a), 4(b) and 5(b), as the Lagrange multiplier increases, the attacker is more inclined to adopt passive attacks. This phenomenon supports the intuition that as the active eavesdropping is at the risk of being discovered, more severe punishment of detection leads to more conservative attack policies for the attacker.

#### B. Structure of the Optimal Policy for Problem 1

In this subsection, we demonstrate the structure of the optimal policy \(\pi^\star\) for Problem 1. Note that \(\pi^\star\) is a randomized mixture of two deterministic policies \(\pi_{\lambda_1}^{2,\star}\) and \(\pi_{\lambda_2}^{2,\star}\), and the constraint is active under such randomization, i.e., \(D_{s_0}(\pi^\star) = \bar{D}\). By employing linear programming [27], we obtain the two deterministic policies \(\pi_{\lambda_1}^{2,\star}\) and \(\pi_{\lambda_2}^{2,\star}\), as shown in Figures 4(a) and 4(b). Two possible lagrangian multipliers for \(\pi_{\lambda_1}^{2,\star}\) and \(\pi_{\lambda_2}^{2,\star}\) are \(\lambda_1 = 13.45\) and \(\lambda_2 = 12.98\). It is observed that \(\pi_{\lambda_2}^{2,\star}\) is different from \(\pi_{\lambda_1}^{2,\star}\) in one state \(s = (2, 0)\). Moreover, we compute that \(I_1 = 0.0639, D_{s_0}(\pi_{\lambda_1}^{2,\star}) = 0.4938, I_2 = 0.0627, D_{s_0}(\pi_{\lambda_2}^{2,\star}) = 0.5029,\) and the randomization parameter \(\varphi = 0.41\).

We also compute the optimal policies \(\pi^\star\) under different values of parameter \(\beta\) (other parameters are the same as that in Table I). We observed that the optimal policy under \(\beta = 6\) is deterministic, in which the attacker is apt to choose passive mode more frequently compared with that of \(\beta = 2\). By calculation, the stealthiness condition is naturally satisfied since \(D_{s_0}(\pi^\star) = 0.4523 < \bar{D} = 0.5\). It follows the intuition that the larger \(\beta\) is, the more energy cost it incurs for each transmission. Hence, the attacker may select passive attacks more frequently, and the number of packet dropouts at the remote estimator will not grow too much to trigger the detection system. Note that if the stealthiness constraint is satisfied, Problem 1 becomes a special case of Problem 2 with \(\lambda = 0\).
C. Comparison with other Policies

Here, we compare the eavesdropping performance and stealthiness of the optimal attack policy $\pi^*$ with that of the conservative/radical policies (i.e., $\pi^{pa}$, $\pi^{ac}$)\(^2\). With the employment of the pair of policies $(\pi, \theta)$, the pair of holding times $(\tau_k^E, \tau_k^A)$ forms a two-dimensional Markov chain. We implement the constrained optimal attack policy $\pi^*$ as well as the conservative/radical policy with the system parameters as described above. The simulation results (about 10000 iterations) are plotted in Figures 5(c) and 5(d). Performance is the expected cost and PDR. Notice that here we use the accumulated average cost $\frac{1}{k} \sum_{i=0}^{k} c(s_k, a_k, q_k)$ and packet-delivery frequency $\frac{1}{k} \sum_{i=0}^{k} 1_{i \geq 1}$ to approximate the average expected cost and PDR. In comparison, we conclude that: The radical policy $\pi^{ac}$ with constant active eavesdropping will provide the best eavesdropping performance while sacrificing stealthiness. Rather than having a high risk of being detected, the attacker will adaptively switch from the active mode to the passive mode by adopting $\pi^*$. The conservative policy $\pi^{pa}$ provides the upper bound of the accumulated average cost and the lowest bound of the detection rate. These figures reveal the tradeoff between eavesdropping performance and the stealthiness for the attacker.

VI. CONCLUSION

This work has investigated remote estimation in CPSs under privacy threats caused by an advanced eavesdropper. Rather than merely passively intercepting the data packet sent from the sensor to the estimator, the eavesdropper also has the ability to jam the transmission channel and launch pilot-contamination attacks to improve its eavesdropping performance. Aiming at providing a test bed for system-resilient methods against the advanced eavesdropper, from the perspective of the attacker, we propose a stealthy structural attack policy, which instructs the attacker to choose the attack mode adaptively. In this work, we restrict ourselves to the design of the attack policy under a pre-determined transmission policy adopted by the sensor. The situation becomes more complicated, and requires further investigation, when the sensor is aware of the attack policy and may adjust its own transmission policy accordingly.

APPENDIX A

PROOF OF THEOREM 1

Proof: As depicted in Figure 2, the transition of $\tau_k^E$ is mixed with that of $\tau_k^A$, which causes difficulty to study the existence conditions for a stationary schedule in the feasible set bounding $J_{s_0}(\pi, 0\theta)$. However, under two special policies (namely, $\pi^{pa}$ and $\pi^{ac}$, which will be introduced later), the state $s_k$ is decoupled into two independent Markov chains with states $\tau_k^E$ and $\tau_k^A$, respectively. Then, from Lemmas 1 and 2, we find that for any policies $\pi \in \Pi$, the performance measures $J_{s_0}(\pi, 0\theta)$ and $D_{s_0}(\pi, 0\theta)$ are bounded by those of policies $\pi^{pa}$ and $\pi^{ac}$. Motivated by this property, we can construct a stationary policy in the feasible set to satisfy the stability condition. The proof details are as follows:

First, we study the expected average cost $J_{s_0}(\pi, 0\theta)$ and detection probability $D_{s_0}(\pi, 0\theta)$ under two special attack policies. Let $\pi^{ac} = (1, 1, \ldots)$ ($\pi^{pa} = (0, 0, \ldots)$, resp.) be a special stationary deterministic attack policy in which the attacker always launches the passive attacks (always passive eavesdropping, resp.) indifferent to the state $s_k$. Under the policy $\pi^{pa}$ and $\theta_0 = (0_{m_0}, 1, 1, \ldots)$, the holding time $\tau_k^E$ induces a time-homogeneous Markov process independent of $\tau_k^A$, and its transition probability matrix is as follows:

$$
T^E(\pi = \pi^{pa}, \theta = \theta_0) = \begin{pmatrix}
1 - d^E(0) & d^E(0) \\
1 - d^E(0) & d^E(0) \\
& & \ddots \\
& & & 1 - d^E(1) & d^E(1)
\end{pmatrix}
$$

(11)

where the $(i + 1, j + 1)$-th entry in the matrix represents the transition probability from $\tau_k^E = i$ to $\tau_k^E + 1 = j$, and the other default entries are 0. Then, we denote the transition probability matrix of $\tau_k^E$ under constant active attacks as $T^E(\pi^{ac}, \theta_0)$, which has similar structure as (11) except that $d^E(\cdot)$ is replaced by $d^E(\cdot)$.

Under the stationary transmission and attack policies $(\pi, \theta)$, and $\bar{\pi}$, a steady state distribution of the Markov chain with state $s_k = (\tau_k^E, \tau_k^A)$ exists independent of the initial state $s_0$ [26]. The invariant state distribution is defined as $\Omega(\pi, \theta) \equiv \{(\omega_{i,j}(\pi, \theta), \omega_{i,j}(\pi, \theta), \omega_{i,j}(\pi, \theta), \ldots)\}$. Furthermore, the marginal distribution of holding time $\tau_k^E$ is denoted by $\overline{\pi}_k(\pi, \theta) \equiv (\omega_{0,0}(\pi, \theta), \omega_{1,0}(\pi, \theta), \omega_{0,1}(\pi, \theta), \ldots)$, in which $\omega_{i,j}(\pi, \theta) = \sum_{j=0}^{\infty} \omega_{(i,j)}(\pi, \theta)$. Similarly, we define the marginal distribution for $\tau_k^A$ as $\overline{\Omega}(\pi, \theta)$ and $\omega_{i,j}(\pi, \theta) = \sum_{j=0}^{\infty} \omega_{(i,j)}(\pi, \theta)$.

Consider the Markov chain with state $3\tau_k^E$ and transition kernel (11), for which the invariant distribution satisfies:

$$
\overline{\pi}(\pi^{pa}, \theta_0) = \overline{\pi}(\pi^{pa}, \theta_0)T^E(\pi^{pa}, \theta_0), \sum_{i=0}^{\infty} \overline{\pi}(\pi^{pa}, \theta_0) = 1.
$$

Hence, we can obtain the closed-form stationary distribution:

$$
\omega_{i}(\pi^{pa}, \theta_0) = \begin{cases}
\omega_{0}(\pi^{pa}, \theta_0)[d^E(0)]^i, & \text{if } i \leq m; \\
\omega_{0}(\pi^{pa}, \theta_0)[d^E(0)]^m[d^E(q_k = 1)]^{i-m}, & \text{otherwise},
\end{cases}
$$

(12)

in which

$$
\omega_{0}(\pi^{pa}, \theta_0) = \frac{[1 - d^E(0)][1 - d^E(1)]}{1 - d^E(1) - [d^E(0)]^m[d^E(0) - d^E(1)]}.
$$

(13)

\(^2\)The conservative policy $\pi^{pa}$ and the radical policy $\pi^{ac}$ are two special stationary deterministic attack policies defined in Appendix A. In short, $\pi^{pa} = (0, 0, \ldots)$ and $\pi^{ac} = (1, 1, \ldots)$.

\(^3\)As shown in Fig. 2, different from $\tau_k^E$, the holding time $\tau_k^A$ cannot build a Markov chain.
For notational simplicity, we define a function \( y(d_0, d_1) \) with \( d_0, d_1 \in [0, 1] \) as follows:

\[
y(d_0, d_1) \triangleq \frac{(1 - d_0)(1 - d_1)}{1 - d_1 - d_0^m (d_0 - d_1)} = \frac{1}{\sum_{k=0}^{m-1} d_0^k + d_0^m},
\]

in which the second equation implies that the function \( y(d_0, d_1) \) is decreasing with \( d_0 \) and \( d_1 \). We have \( \bar{\nu}_0(\pi^{pa}, \theta_0) = y(d^E(0), d^E(1)) \) and the limit of the corresponding asymptotic PDR in (8) exists:

\[
D_{\pi_0}(\pi^{pa}, \theta_0) = \sum_{i=1}^{+\infty} \bar{\nu}_i(\pi^{pa}, \theta_0) = 1 - y(d^E(0), d^E(1)).
\]

(15)

Similarly, we can obtain the values of \( \bar{\nu}_i(\pi^{ac}, \theta_0) \) and \( D(\pi^{ac}, \theta_0) \) for constant active attacks \( \pi^{ac} \) via replacing \( d^E(\cdot) \) with \( d^E(\cdot) \).

Next, we will show that for arbitrary stationary attack schedule, the following property holds:

**Lemma 1.** Suppose that the sensor adopts a stationary transmission policy \( \theta^{ar} \in \Theta_s \). For any stationary attack policy \( \pi^{ar} \in \Pi_s \), the limiting distribution \( \Omega_{i,j}(\pi^{ar}, \theta^{ar}) \) of holding time \( \tau^E \) has the following properties: \( \forall \theta^{ar} \in \Theta_s \),

\[
\bar{\nu}_{n+1}(\pi^{ar}, \theta^{ar}) \leq \bar{\nu}_n(\pi^{ar}, \theta^{ar}),
\]

\[
\bar{\omega}_{n+1}(\pi^{ar}, \theta^{ar}) \leq \bar{\omega}_n(\pi^{ar}, \theta^{ar}),
\]

\[
\bar{\nu}_0(\pi^{pa}, \theta^{ar}) \geq \bar{\omega}_0(\pi^{ar}, \theta^{ar}) \geq \bar{\nu}_0(\pi^{ac}, \theta^{ar}),
\]

\[
\bar{\omega}_0(\pi^{pa}, \theta^{ar}) \leq \bar{\omega}_0(\pi^{ar}, \theta^{ar}) \leq \bar{\nu}_0(\pi^{ac}, \theta^{ar}).
\]

**Proof:** From the transition kernel in (7), we have for \( i \geq 1 \) that

\[
\bar{\nu}_i(\pi^{ar}, \theta^{ar}) = \sum_{j=1}^{+\infty} \omega_{(i,j)}(\pi^{ar}, \theta^{ar}) + \omega_{(i,0)}(\pi^{ar}, \theta^{ar})
\]

\[
= \sum_{j=1}^{+\infty} d^E(\pi^{ar}(j-1), \theta^{ar}(i-1)) d^A(\cdot) \omega_{(i-1,j-1)}(\pi^{ar}, \theta^{ar})
\]

\[
+ \sum_{l=0}^{+\infty} d^E(\pi^{ar}(l), \theta^{ar}(i-1)) (1 - d^A(\cdot)) \omega_{(i-1,l)}(\pi^{ar}, \theta^{ar})
\]

\[
= \sum_{j=0}^{+\infty} d^E(\pi^{ar}(j), \theta^{ar}(i-1)) \omega_{(i-1,j)}(\pi^{ar}, \theta^{ar})
\]

\[
\leq \sum_{j=0}^{+\infty} \omega_{(i-1,j)}(\pi^{ar}, \theta^{ar}) = \bar{\nu}_{i-1}(\pi^{ar}, \theta^{ar}).
\]

(16)
Analogously, we can obtain \( \omega_{n+1}(\pi^{ar}, \theta^{ar}) \leq \omega_n(\pi^{ar}, \theta^{ar}) \).
Next, we will prove that \( \omega_0(\pi^{ac}, \theta^{ar}) \geq \omega_0(\pi^{ac}, \theta^{ar}) \) by contradiction. From (16), we have
\[
\omega_1(\pi^{ar}, \theta^{ar}) = \sum_{i=0}^{+\infty} d^E(\pi^{ar}(j), \theta^{ar}(i-1))\omega(i-1,j)(\pi^{ar}, \theta^{ar}) \leq \omega_0(\pi^{ac}, \theta^{ar}),
\]
which the inequality (a) is derived as \( d^E(a_k, \theta^{ar}(j)) \leq d^E(\theta^{ar}(j)) \) for any \( a_k \) and \( j \) according to Assumption 1 (iii), and the inequality (b) is obtained by using the inequality (a) recursively. Suppose \( \omega_0(\pi^{ar}, \theta^{ar}) < \omega_0(\pi^{ac}, \theta^{ar}) \), then
\[
\sum_{i=0}^{+\infty} \omega_i(\pi^{ar}, \theta^{ar}) \leq \sum_{i=0}^{+\infty} \prod_{l=0}^{i-1} d^E(\theta^{ar}(l))\omega_0(\pi^{ar}, \theta^{ar}) < \sum_{i=0}^{+\infty} \prod_{l=0}^{i-1} d^E(\theta^{ar}(l))\omega_0(\pi^{ac}, \theta^{ar}) = \omega_1(\pi^{ac}, \theta^{ar}) = 1,
\]
which contradicts with the fact that \( \sum_{i=0}^{+\infty} \omega_i(\pi^{ar}, \theta^{ar}) = 1 \). The equality (a) holds because if \( \pi^{ar} = \pi^{ac} \), then \( d^E(\pi^{ar}(j), \theta^{ar}(i-1)) = d^E(\theta^{ar}(i-1)) \) for all \( i \geq 1 \) and the equations in (17) is satisfied. Consequently, \( \omega_0(\pi^{ar}, \theta^{ar}) \geq \omega_0(\pi^{ac}, \theta^{ar}) \). Analogously, we can obtain \( \omega_0(\pi^{ar}, \theta^{ar}) \leq \omega_0(\pi^{pa}, \theta^{ar}) \) for the conservative policy \( \pi^{pa} \). Notice that \( d^A(a_k, \theta^{ar}(j)) \geq d^A(\theta^{ar}(j)) \) for any \( a_k \) and \( j \) according to Assumption 1 (iii). The proof for the last inequality is similar to the above and we omit the details here to avoid redundancy.

With the help of Lemma 1, we next attempt to investigate the “weak majorization” order among arbitrary policy \( \pi^{ar} \), conservative policy \( \pi^{pa} \) and radical policy \( \pi^{ac} \), as demonstrated in the following lemma:

**Lemma 2.** Suppose that the sensor adopts a stationary transmission policy \( \theta^{ar} \in \Theta_s \). Then the preorder on the invariant distribution vectors under the three policies \( \pi^{ar}, \pi^{pa} \) and \( \pi^{ac} \) is:\(^4\)
\[
\Omega(\pi^{pa}) \succ \Omega(\pi^{ar}) \succ \Omega(\pi^{ac}),
\]
\[
\Omega(\pi^{ac}) \succ \Omega(\pi^{ar}) \succ \Omega(\pi^{pa}),
\]
where \( \succ \) denotes the weak majorization.

**Proof:** Define \( m_0 = \min\{m : \omega_m(\pi^{ac}) > \omega_m(\pi^{ar})\} \). Based on Lemma 1, we can obtain that for all \( m \leq m_0 \),
\[
\sum_{i=0}^{m} \omega_i(\pi^{ar}) \geq \sum_{i=0}^{m} \omega_i(\pi^{ac}).
\]
If \( m \geq m_0 \), we have
\[
\omega_m(\pi^{ar}) \leq \omega_{m_0}(\pi^{ac}) \prod_{l=m_0}^{m} d^E(\theta^{ar}(l)) < \omega_{m_0}(\pi^{ac}) \prod_{l=m_0}^{m} d^E(\theta^{ar}(l)) = \omega_m(\pi^{ac}).
\]
Hence, we can obtain that for all \( m \geq m_0 \),
\[
\sum_{i=0}^{m} \omega_i(\pi^{ar}) \geq \sum_{i=m_0+1}^{m} \omega_i(\pi^{ac}) = \sum_{i=0}^{m} \omega_i(\pi^{ac}).
\]
The proof of \( \Omega(\pi^{pa}) \succ \Omega(\pi^{ar}) \succ \Omega(\pi^{ac}) \) is similar and we omit details here due to the limited space.

Based on Lemma 2 and Proposition 1, we can obtain \( J_{s_0}(\pi^{ar}, \theta_0) \leq J_{s_0}(\pi^{ac}, \theta_0) \leq J_{s_0}(\pi^{pa}, \theta_0) \) and \( D_{s_0}(\pi^{co}, \theta_0) \geq D_{s_0}(\pi^{ac}, \theta_0) \geq D_{s_0}(\pi^{pa}, \theta_0) \).

**Lemma 3.** With the employment of transmission and attack policies \( (\pi^{ar}, \theta_0) \), \( J_{s_0}(\pi^{ac}, \theta_0) \) converges absolutely if and only if \( d^A(1) < [r(A)]^{-2} \).

**Proof:** Under the attack policy \( \pi^{ac} \), we have
\[
J_{s_0}(\pi^{ac}, \theta_0) = \sum_{i=0}^{+\infty} \omega_i(\pi^{ac})h^i(\mathcal{T}) = \sum_{i=0}^{m} \omega_0(\pi^{ac})[d^A(0)]^i h^i(\mathcal{T}) + \sum_{i=m+1}^{+\infty} \omega_0(\pi^{ac})[d^A(0)]^m[d^A(1)]^{i-m} h^i(\mathcal{T}),
\]
the convergence of which is only related to the series \( \sum_{i=m+1}^{+\infty}[d^A(1)]^{i-m} h^i(\mathcal{T}) \). From [12], we know that the series \( \sum_{i=0}^{+\infty} A_i Q A_i^T \) converges if and only if \( r(A) < 1 \). Hence, the sufficient and necessary condition for the convergence of (18) is \( r(\sqrt{d^A(1)A}) < 1 \). Based on the conservative policy \( \pi^{pa} \) and radical policy \( \pi^{ac} \), we may construct a stationary randomized policy \( \pi^{co} = \alpha \pi^{ac} + (1 - \alpha) \pi^{pa} \), of which the attacker randomly chooses \( \pi^{pa} \) or \( \pi^{ac} \) following a Bernoulli distribution with coefficient \( \alpha \in [0, 1] \). Similar to the calculation of \( D_{s_0}(\pi^{pa}, \theta_0) \) in (15), we compute that \( D_{s_0}(\pi^{co}, \theta_0) = 1 - \psi(d^A_{\pi^{pa}}, d^E_{\pi^{pa}}) \). Moreover, we have \( J_{s_0}(\pi^{co}, \theta_0) = \alpha J_{s_0}(\pi^{pa}, \theta_0) + (1 - \alpha) J_{s_0}(\pi^{ac}, \theta_0) \). If Assumption 2 is satisfied, according to Lemma 3 we conclude that the admissible stationary policy \( \pi^{co} \in \Pi_s \) induces bounded expected error covariance \( J_{s_0} \) for the attacker, and simultaneously satisfies the stealthy condition \( D_{s_0}(\pi^{co}, \theta_0) \leq D \).

**APPENDIX B**

**PROOF OF THEOREM 2**

**Existence:** The existence of a stationary, deterministic discounted cost optimal policy follows straightforwardly from [26].
Structure: The optimal policy $\pi^3_{\lambda,\xi}$ can be obtained by value iteration, which will converge for all initial conditions. In order to prove the monotonicity of the optimal policy, it is sufficient to show the monotonicity and submodularity of $V_{\lambda,\xi}(s)$:
1. monotonicity: $V_{\lambda,\xi}(s) \leq V_{\lambda,\xi}(s')$ if $s \leq s'$;
2. submodularity: $V_{\lambda,\xi}(s) + V_{\lambda,\xi}(s') \geq V_{\lambda,\xi}(s \wedge s') + V_{\lambda,\xi}(s \vee s')$ for any $s, s' \in S$.

Considering the structure of the transmission policy, the monotonicity of the optimal attack policy has been discussed for two disjoint state regions $S_{\leq m}$ and $S_{> m}$. Since in each region the transmission power is a constant, we can omit the transmission power $q$ in the following proof for notational convenience.

First, we consider the case when $s, s' \in S_{> m}$.

Monotone: Suppose $s \leq s'$ and $f(s) \in K$ is a measurable increasing function. Notice that $r^\lambda(s, a)$ is increasing in $s$. Hence, for each state $s, s' \in S_{> m}$, and for any $a \in A$, we have
$$r^\lambda(s, a) + \xi C(s, a, f) \leq r^\lambda(s', a) + \xi \sum_{s' \in S} \Pr(s|s', a)V_{\lambda,\xi}(s'),$$
which implies that the monotonicity of function $f(s)$ is preserved by the operator $W_{\xi}$, namely, $W_{\xi}f(s') \geq W_{\xi}f(s)$. By adopting the operator $W_{\xi}$ repeatedly, we obtain
$$\lim_{n \to \infty} W_{\xi}^n f(s) = V_{\lambda,\xi}(s)$$

Notice that, according to [40], convergence of $V_{\lambda,\xi}(s)$ is independent of the monotonicity assumption of $f(s)$.

Submodular: First, we assume that $f(s)$ is a measurable submodular function, i.e.,
$$f(s) + f(s') \geq f(s \wedge s') + f(s \vee s'), \forall s, s' \in S_{\leq m}.$$ Notice that the convergence of $V_{\lambda,\xi}(s)$ is also independent of the submodularity assumption of $f(s)$. Motivated by the previous monotonicity discussion, in order to prove the submodularity of $V_{\lambda,\xi}(s)$, it suffices to prove that the submodularity of $f(s)$ propagates through the operator, i.e., to prove the submodularity of $W_{\xi}f(s)$. Due to the linear combination structure of the immediate reward function $r^\lambda(s, a)$, it is sufficient to prove that
$$\min_{a \in A} \beta f + \frac{\lambda}{\lambda} \sum_{s \in S} \Pr(s|s', a) f(\tilde{s}) + \min_{a \in A} \beta f + \frac{\lambda}{\lambda} \sum_{s \in S} \Pr(s|s', a) f(\tilde{s}).$$

Consequently, $\min[\beta f + \min[\beta f + \min[\beta f + \min[\beta f + \min[\beta f]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]}
The other scenario when \( \partial E \partial A > \partial \tilde{E} \partial A \) can be proved by showing
\[
C(s \lor s',1,f) + C(s \land s',0,f) - C(s,0,f) - C(s',1,f) \leq 0 \text{ in a similar way.}
\]
(3) Case \( \alpha^*(s) = \alpha^*(s') = 1 \) or \( \alpha^*(s) = \alpha^*(s') = 0 \): It can be easily proved by the fact that \( f(s) \) is submodular when \( s \in S_{i<m} \).

The monotonicity and the submodularity of function \( W_x f(s) \) also holds when \( s \in S_{i<m} \), of which the proof is similar to the above.

REFERENCES

[1] K.-D. Kim and P. Kumar, “Cyber-physical systems: A perspective at the centennial,” Proceedings of the IEEE, May 2012.
[2] A. A. Cardenas, S. Amin, and S. Sastry, “Secure control: Towards survivable cyber-physical systems,” in Proc. IEEE 28th Int. Conf. Distributed Computing System Workshops, 2008, pp. 495–500.
[3] Y. Mo, S. Weerakkody, and B. Sinopoli, “Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs,” IEEE Trans. on Control Systems, vol. 35, no. 1, pp. 93–109, Feb 2015.
[4] I. Rouf, H. Mustafa, M. Xu, W. Xu, R. Miller, and M. Gruteser, “Neighborhood watch: security and privacy analysis of automatic meter reading systems,” in Proceedings of the 2012 ACM conference on Computer and communications security, pp. 462–473.
[5] X. Zhou, B. Maham, and A. Hjorungnes, “Pilot contamination for active eavesdropping,” IEEE Trans. on Wireless Communications, vol. 11, no. 3, pp. 903–907, March 2012.
[6] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, “Secure massive MIMO transmission with an active eavesdropper,” IEEE Trans. on Information Theory, vol. 62, no. 7, pp. 3880–3900, July 2016.
[7] M. Bloch and J. Barros, Physical-layer security: from information theory to security engineering. Cambridge University Press, 2011.
[8] A. Makkehje and A. L. Swindlehurst, “Jamming games in the MIMO wiretap channel with an active eavesdropper,” IEEE Trans. on Signal Processing, vol. 61, no. 1, pp. 82–91, Jan 2013.
[9] H. Sandberg, G. Dán, and R. Thobaben, “Differentially private state estimation in distribution networks with smart meters,” in 2015 IEEE 54th Annual Conf. on Decision and Control (CDC), 2015, pp. 4492–4498.
[10] A. C. Tsinias, K. Gatsis, and G. J. Pappas, “State estimation with secrecy against eavesdroppers,” IFAC-PapersOnLine, vol. 50, no. 1, pp. 8385 – 8392, 2017, 20th IFAC World Congress. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405896317321559
[11] A. S. Leong, D. E. Quevedo, D. Doliz, and S. Dey, “Remote state estimation over packet dropping links in the presence of an eavesdropper,” IFAC-PapersOnLine, vol. 50, no. 1, pp. 7339 – 7344, 2017, 20th IFAC World Congress. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405896317320591
[12] Z. Ren, P. Cheng, J. Chen, L. Shi, and H. Zhang, “Dynamic sensor transmission power scheduling for remote state estimation,” Automatica, vol. 50, no. 4, pp. 1235–1242, 2014.
[13] T.-Y. Liu, P.-H. Lin, S.-C. Lin, Y.-W. P. Hong, and E. A. Jorswieck, “To avoid or not to avoid csi leakage in physical layer secret communication systems,” IEEE Communications Magazine, vol. 53, no. 12, pp. 19–25, 2015.
[14] Y.-S. Shiu, S. Y. Chang, H.-C. Wu, S. C.-H. Huang, and H.-H. Chen, “Physical layer security in wireless networks: A tutorial,” IEEE wireless Communications, vol. 18, no. 2, 2011.
[15] X. Ren, J. Wu, K. H. Johansson, G. Shi, and L. Shi, “Infinite horizon optimal transmission power control for remote state estimation over fading channels,” IEEE Trans. on Automatic Control, 2017.
[16] L. Shi, P. Cheng, and J. Chen, “Sensor data scheduling for optimal state estimation with communication energy constraint,” Automatica, vol. 47, no. 8, pp. 1693–1698, 2011.
[17] A. S. Leong, S. Dey, and D. E. Quevedo, “Sensor scheduling in variance based event triggered estimation with packet drops,” IEEE Trans. on Automatic Control, vol. 62, no. 4, pp. 1880–1895, 2017.
[18] W. Xu, W. Trappe, Y. Zhang, and T. Wood, “The feasibility of launching and detecting jamming attacks in wireless networks,” in Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing, 2005, pp. 46–57.
[19] M. L. Puterman, Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons, 2014.
[20] E. Altman, Constrained Markov decision processes. CRC Press, 1999, vol. 7.
[21] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, “Kalman filtering with intermittent observations,” IEEE Trans. on Automatic Control, vol. 49, no. 9, pp. 1453–1464, 2004.
[22] D. M. Topkis, Supermodularity and complementarity. Princeton university press, 2011.
[23] O. Hernández-Lerma and J. B. Lasserre, Discrete-time Markov control processes: basic optimality criteria. Springer Science & Business Media, 2012, vol. 30.
[24] R. Cazavos-Cadena and L. I. Sennott, “Comparing recent assumptions for the existence of average optimal stationary policies,” Operations research letters, vol. 11, no. 1, pp. 33–37, 1992.
[25] F. J. Beutler and K. W. Ross, “Optimal policies for controlled markov chains with a constraint,” Journal of mathematical analysis and applications, vol. 112, no. 1, pp. 236–252, 1985.
[26] K. W. Ross, “Randomized and past-dependent policies for markov decision processes with multiple constraints,” Operations Research, vol. 37, no. 3, pp. 474–477, 1989.
[27] S. M. Ross, Applied probability models with optimization applications. Courier Corporation, 2013.
[28] M. Epstein, L. Shi, A. Tiwari, and R. M. Murray, “Probabilistic performance of state estimation across a lossy network,” Automatica, vol. 44, no. 12, pp. 3046–3053, 2008.
[29] K. Ding, Y. Li, D. E. Quevedo, S. Dey, and L. Shi, “A multi-channel transmission schedule for remote state estimation under DoS attacks,” Automatica, vol. 78, pp. 194–201, 2017.
[30] V. S. Borkar, “An actor-critic algorithm for constrained markov decision processes,” Systems & control letters, vol. 54, no. 3, pp. 207–213, 2005.
[31] P. Geibel, “Reinforcement learning for mdps with constraints,” in European Conference on Machine Learning. Springer, 2006, pp. 646–653.
[32] D. V. Djinin and V. Krishnamurthy, “Q-learning algorithms for constrained markov decision processes with randomized monotone policies: Application to mimo transmission control,” IEEE Trans. on Signal Processing, vol. 55, no. 5, pp. 2170–2181, 2007.
[33] O. Hernández-Lerma and J. B. Lasserre, Further topics on discrete-time Markov control processes. Springer Science & Business Media, 2012, vol. 42.
Kemi Ding received the B.S. degree in Electronic and Information Engineering from Huazhong University of Science and Technology, China, in 2014 and the Ph.D. degree in the Department of Electronic and Computer Engineering from Hong Kong University of Science and Technology in 2018. She is currently a postdoctoral researcher at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Prior to this, she was a postdoctoral researcher in the School of Electrical, Computer and Energy Engineering, Arizona State University from September 2018 to August 2019. Her current research interests include cyber-physical system security/privacy, networked state estimation, game theory and social networks.

Alex S. Leong was born in Macau in 1980. He received the B.S. degree in mathematics and B.E. degree in electrical engineering in 2003, and the Ph.D. degree in electrical engineering in 2008, all from the University of Melbourne, Australia. He is currently a Research Associate at Paderborn University, Germany. He was with the Department of Electrical and Electronic Engineering at the University of Melbourne from 2008 to 2015. His research interests include networked control systems, signal processing for sensor networks, and statistical signal processing. Dr. Leong was the recipient of the L. R. East Medal from Engineers Australia in 2003, an Australian Postdoctoral Fellowship from the Australian Research Council in 2009, and a Discovery Early Career Researcher Award from the Australian Research Council in 2012.

Xiaoqiang Ren is a professor at the School of Mechatronic Engineering and Automation, Shanghai University, China. He received the B.E. degree in Automation from Zhejiang University, Hangzhou, China, in 2012 and the Ph.D. degree in control and dynamic systems from Hong Kong University of Science and Technology in 2016. Prior to his current position, he was a postdoctoral researcher in the Hong Kong University of Science and Technology in 2016, Nanyang Technological University from 2016 to 2018, and KTH Royal Institute of Technology from 2018 to 2019. His research interests include security of cyber-physical systems, sequential decision, and networked estimation and control.

Ling Shi received the B.S. degree in electrical and electronic engineering from Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2002 and the Ph.D. degree in Control and Dynamical Systems from California Institute of Technology, Pasadena, CA, USA, in 2008. He is currently an associate professor at the Department of Electronic and Computer Engineering, and the associate director of the Robotics Institute, both at the Hong Kong University of Science and Technology. His research interests include cyber-physical systems security, networked control systems, sensor scheduling, event-based state estimation, and exoskeleton robots. He is a senior member of IEEE. He served as an editorial board member for The European Control Conference 2013-2016. He was a subject editor for International Journal of Robust and Nonlinear Control (2015-2017). He has been serving as an associate editor for IEEE Transactions on Control of Network Systems from July 2016, and an associate editor for IEEE Control Systems Letters from Feb 2017. He also served as an associate editor for a special issue on Secure Control of Cyber Physical Systems in the IEEE Transactions on Control of Network Systems in 2015-2017. He served as the General Chair of the 23rd International Symposium on Mathematical Theory of Networks and Systems (MTNS 2018).

Daniel E. Quevedo (S’97–M’05–SM’14) received Ingeniero Civil Electrónico and M.Sc. degrees from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2000, and in 2005 the Ph.D. degree from the University of Newcastle in Australia. He is a Professor of Cyberphysical Systems with the School of Electrical Engineering and Robotics, Queensland University of Technology (QUT), Brisbane, Australia. Before joining QUT, he was head of the Chair of Automatic Control at Paderborn University, Germany.

Dr. Quevedo was supported by a full scholarship from the alumni association during his time at the Universidad Técnica Federico Santa Maria and received several university-wide prizes upon graduating. He received the IEEE Conference on Decision and Control Best Student Paper Award in 2003 and was also a finalist in 2002. In 2009 he was awarded a five-year Research Fellowship from the Australian Research Council. He is co-recipient of the 2018 IEEE Transactions on Automatic Control George S. Axelby Outstanding Paper Award.

Prof. Quevedo is Associate Editor of the IEEE Control Systems Magazine and past Chair of the IEEE Control Systems Society Technical Committee on Networks & Communication Systems. His research interests are in control of networked systems and of power converters.