An Event-Based Stealthy Attack on Remote State Estimation

Peng Cheng, *Member, IEEE*, Zeyu Yang, *Student Member, IEEE*, Jiming Chen, *Fellow, IEEE*, Yifei Qi, and Ling Shi, *Senior Member, IEEE*

***Abstract*—Security issues in cyber-physical systems (CPSs) have gained increasing attention in recent years due to the importance and unavoidable vulnerability of CPSs. This article focuses on designing an intelligent online attack, which can compromise a sensor, eavesdrop measurements, and inject false feedback information, against remote state estimation. From the viewpoint of the attacker, we design an event-based attack strategy to degrade the estimation quality with an arbitrary communication rate stealthy constraint. The approximate minimum mean-squared error estimation algorithm from the viewpoint of the attacker is derived under a Gaussian assumption. Furthermore, the relation between the attack threshold and the scheduling threshold is obtained in a closed form. We show that the mean-squared stability condition of the estimation is weakened under the attack. Two examples are provided to demonstrate the main results.**

***Index Terms*—Event-based scheduling, state estimation, stealthy attack.**

I. Introduction

Cyber-physical systems (CPSs), such as integrity information and physical elements, have attracted great research interest in the past decade. Driven by the integration of control, communication, and computation, CPSs are applied in many industrial fields, such as energy, transportation, and manufacturing, to name a few [2]–[4]. Since the Iran's nuclear facilities and Ukraine's power grid are attacked by Stuxnet and BlackEnergy, causing significant damage to the system, security issues in CPSs have been investigated from different perspectives [5]–[7].

Remote state estimation (RSE) is a critical problem in CPSs, since its accuracy is the precondition of the system performance. However, integrity [8] and availability [9] of data in RSE are vulnerable to the increasing cyber attack. Guo *et al.* [10] proposed an optimal linear data integrity attack on RSE evading the Chi-squared detector. Compared with integrity attacks, the vulnerability of availability is much easier to be utilized, due to the unmanned guarded sensors and the unprotected communication channel between sensors and the estimator. Zhang *et al.* [11] studied the optimal denial-of-service (DoS) attack strategy against state estimation. The authors proved that the optimal jamming strategy for energy-constrained attacker is consecutive attack and provided the performance reduction with a closed form. In [12], Ding *et al.* investigated the decision-making procedures for defending sensors and malicious attackers in a multichannel communication network. They proposed a Nash Q-learning algorithm to solve the optimal strategies for both players. In [13], Peng *et al.* considered the optimal attack power schedule to degrade RSE performance in multisystems under the average energy constraints. In [14], Qin *et al.* considered the optimal attack scheduling with energy constraints to maximize the average RSE error in packet-dropping networks.

It should be pointed out that most of the existing works mainly assume that the sensor has enough computational capability to obtain the local estimation and has the ability to send the local estimation to the remote estimator. Therefore, it is beneficial for the attacker to design and execute the attack strategy. However, there are various practical application scenarios, where the computational capability is highly limited, e.g., an electronic healthcare system with a body sensor [15], environment monitoring with a temperature or humidity sensor [16], etc. At the same time, most of the existing works consider the attack mechanisms in the absence of the DoS attack detector, e.g., a communication-rate-based detector in [17]. Motivated by these observations, in this article, we focus on designing an intelligent online stealthy attack strategy against RSE, in which sensors send measurement to the remote estimator directly, with an arbitrary communication rate constraint caused by the limited sensor's energy or limited communication bandwidth.

The objective of the attacker is to degrade the estimation quality by cooperatively eavesdropping the measurements, compromising and injecting false feedback data into the sensor node. Since the sensor has limited computation capacity and the remote estimator may detect the attack behavior, if the communication rate is changed, it is challenging to design the attack mechanism to deteriorate the estimation performance as much as possible under the communication rate constraint. Specifically, we are interested in designing an event-based attack with a proper trigger mechanism and an attack threshold to degrade the state estimation of a linear system with Gaussian noises.

The major contributions of this article can be summarized as follows.

1) To the best of our knowledge, it is the first work on designing the online attack against RSE with the arbitrary communication rate constraint.
2) We design an event-based attack mechanism, which leverages real-time measurements and can be implemented through compromising and injecting false feedback data into the sensor.
3) We derive the approximate minimum mean-squared error (MMSE) estimation algorithm from the viewpoint of the attacker under the Gaussian assumption. We further obtain an analytical relation between the attack threshold and the original scheduling

Manuscript received February 4, 2019; revised July 2, 2019; accepted November 11, 2019. Date of publication November 26, 2019; date of current version September 25, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0803501, in part by the National Natural Science Foundation of China under Grant 61833015, and in part by "Research on Lightweight Active Immune Technology for Electric Power Supervisory Control System," a science and technology project of State Grid Co., Ltd. in 2019. This article was presented in part at the 54th IEEE Conference on Decision and Control, 2015 [1]. Recommended by Associate Editor R. M. Jungers. *(Corresponding author: Jiming Chen.)*

P. Cheng, Z. Yang, J. Chen, and Y. Qi are with the State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China (e-mail: pcheng@iipc.zju.edu.cn; zeyuyang@zju.edu.cn; jmchen@ieee.org; yifeiqi1127@gmail.com).

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

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2019.2956021

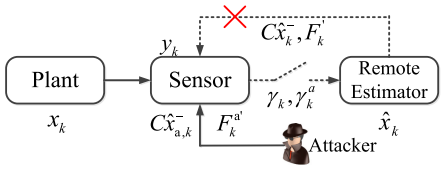

Fig. 1. System architecture.

threshold, which guarantees the invariability of the communication rate.

4) We evaluate the estimation performance degradation under the designed attack, by deriving the evolution of error covariance for the MMSE with attack. It is also proved that the mean-squared stability and the estimation accuracy are both weakened.

The remainder of this article is organized as follows. Section II introduces system architecture and problem formulation. In Section III, we design the intelligent online attack mechanism against RSE, derive the approximate MMSE algorithm from the viewpoint of the attacker and the stealthy attack threshold, and prove the effectiveness of the proposed attack. Some examples are presented in Section IV to illustrate our results. Section V concludes this article.

Notations: $\mathbb{R}^n$ is the n-dimensional Euclidian space. $\mathbb{S}_+^n$ is the set of $n \times n$ positive-semidefinite matrices. When $X \in \mathbb{S}_+^n$, we simply write $X \geq 0$; when X is positive definite, we write $X > 0$. $f_X(x)$ denotes the probability density function (pdf) of the random variable x, and $f_{x|y}(x|y)$ represents the pdf of x conditional on y. $\mathcal{N}(\mu, \Sigma)$ denotes the Gaussian distribution with mean μ and covariance matrix Σ. $\mathbb{E}[X]$ denotes the mathematical expectation of X, and $\mathbb{E}[X|y]$ represents the expectation of X conditional on y. For functions $f_1, f_2 : \mathbb{S}_+^n \to \mathbb{S}_+^n$, $f_1 \circ f_2$ is defined as $f_1 \circ f_2(X) \triangleq f_1(f_2(X))$.

II. Problem Formulation

A. System Model

Consider a discrete linear time-invariant system (see Fig. 1)

$$x_{k+1} = Ax_k + \omega_k \tag{1}$$

$$y_k = Cx_k + \nu_k \tag{2}$$

where $x_k \in \mathbb{R}^n$ is the state system, $y_k \in \mathbb{R}^m$ is the measurement obtained by the sensor, and A and C are known time-invariant real matrices. $\omega_k \in \mathbb{R}^n$ and $\nu_k \in \mathbb{R}^m$ are uncorrelated zero-mean Gaussian random noise with covariances $Q \geq 0$ and $R > 0$, respectively. The initial state x_0 is also a zero-mean Gaussian random vector, which is uncorrelated with ω_k or ν_k and has covariance $P_0 \geq 0$. Assume that the pair $(A, \sqrt{Q})$ is controllable and (C, A) is observable.

Denote $Y_k = \{y_0, y_1, \ldots, y_k\}$ as all the measurement data collected by the sensor from time 0 to time k. However, due to the energy constraint of the battery-powered sensor or the communication bandwidth constraint, the measurements cannot be sent at each time slot, which leads to the sensor-to-estimator communication rate constraint. Thus, the sensor has to decide whether y_k shall be sent or not, which is denoted as $\gamma_k = 1$ or 0, respectively. Then denoting $\mathrm{I}_k = \{\gamma_0 y_0, \ldots, \gamma_k y_k\} \cup \{\gamma_0, \ldots, \gamma_k\}$ with $\mathrm{I}_{-1} = \emptyset$. The remote estimator shall estimate the system state by the MMSE estimator, including the *a priori* and *a posteriori* estimation, which are defined as

$$\hat{x}_k^- \triangleq \mathbb{E}[x_k | \mathrm{I}_{k-1}] \qquad a\,priori \tag{3}$$

$$\hat{x}_k \triangleq \mathbb{E}[x_k | \mathrm{I}_k] \qquad a\,posteriori. \tag{4}$$

The corresponding error and covariance are defined as

$$e_k^- \triangleq x_k - \hat{x}_k^-, \qquad P_k^- \triangleq \mathbb{E}[e_k^- e_k^{-'} | \mathrm{I}_{k-1}] \tag{5}$$

$$e_k \triangleq x_k - \hat{x}_k, \qquad P_k \triangleq \mathbb{E}[e_k e_k' | \mathrm{I}_k]. \tag{6}$$

Furthermore, define the measurement innovation z_k as

$$z_k \triangleq y_k - \mathbb{E}[y_k | \mathrm{I}_{k-1}]. \tag{7}$$

To simplify the notation, we define the operators $h, \widetilde{g}_\lambda$, and $g_\lambda : \mathbb{S}_+^n \to \mathbb{S}_+^n$ as

$$h(X) \triangleq AXA' + Q \tag{8}$$

$$\widetilde{g}_\lambda(X) \triangleq X - \lambda XC'[CXC' + R]^{-1}CX \tag{9}$$

$$g_\lambda(X) \triangleq \widetilde{g}_\lambda \circ h(X) \tag{10}$$

where $\lambda \in [0, 1]$. While $\lambda = 1$, $\widetilde{g}_\lambda$ and g_λ will be, respectively, written as $\widetilde{g}$ and g for brevity. Note that $h(\cdot)$ and $\widetilde{g}$ are the Lyapunov operator and the Riccati operator, respectively.

B. Event-Based RSE

In this article, we consider that the sensor has an *average communication rate* constraint defined as [18]

$$\gamma \triangleq \limsup_{T \to +\infty} \frac{1}{T+1} \sum_{k=0}^{T} \mathbb{E}[\gamma_k]. \tag{11}$$

While feedback is available from the estimator to the sensor, based on the fact that the innovation z_k is a Gaussian variable with zero mean and covariance $CP_k^-C' + R > 0$, Wu *et al.* [18] proposed an event-based scheduling scheme for RSE, which can improve the estimation accuracy under the communication rate constraint.

The event-based sensor data scheduler is described as

$$\gamma_k = \begin{cases} 0, & \text{if } \| \epsilon_k \|_\infty \leq \delta \\ 1, & \text{otherwise} \end{cases} \tag{12}$$

where δ is the event-triggering threshold and

$$\epsilon_k \triangleq F_k' z_k, \; F_k = U_k \Lambda_k^{-\frac{1}{2}}$$

with $\Lambda_k = \mathrm{diag}(\lambda_k^1, \ldots, \lambda_k^m) \in \mathbb{R}^{m \times m}$ and $\lambda_k^1, \ldots, \lambda_k^m \in \mathbb{R}$ are the eigenvalues of $CP_k^-C' + R$, and U_k is a unitary matrix satisfying

$$U_k'(CP_k^-C' + R)U_k = \Lambda_k.$$

Then the approximate MMSE estimator and the communication rate for the remote estimator with (12) are summarized in the following lemma [18].

Lemma 2.1: Consider the RSE with the event-based sensor scheduler (12). Under the Gaussian assumption

$$f_{x_k}(x | \mathrm{I}_{k-1}) = \mathcal{N}(\hat{x}_k^-, P_k^-)$$

the MMSE estimator is given recursively as follows.

1) Time update:

$$\begin{cases} \hat{x}_k^- = A\hat{x}_{k-1} \\ P_k^- = h(P_{k-1}). \end{cases} \tag{13}$$

2) Measurement update:

$$\begin{cases} \hat{x}_k = \hat{x}_k^- + \gamma_k P_k^- C'[CP_k^- C' + R]^{-1} z_k \\ P_k = \gamma_k \widetilde{g}(P_k^-) + (1-\gamma_k)\widetilde{g}_{\beta(\delta)}(P_k^-) \end{cases} \tag{14}$$

where

$$\beta(\delta) = \frac{2}{\sqrt{2\pi}} \delta e^{-\frac{\delta^2}{2}} [1 - 2Q(\delta)]^{-1}$$

and $Q(\cdot)$ is the standard Q-function defined by

$$Q(\delta) \triangleq \int_{\delta}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

From [18, Lemma 3.6], no matter what I_{k-1} takes, the pdf of γ_k on condition I_{k-1} satisfies

$$\Pr(\gamma_k = 0) = [1 - 2Q(\delta)]^m$$
$$\Pr(\gamma_k = 1) = 1 - [1 - 2Q(\delta)]^m.$$

As a result, the sensor data scheduler is an independent identically distributed (i.i.d) progress, according to (11), we have

$$\gamma = 1 - [1 - 2Q(\delta)]^m. \tag{15}$$

Note that for the attack detection mechanism, the remote estimator can compute the average number of the received data at real time, which is the unbiased estimation of γ. At the meantime, the remote estimator can also use sensor data scheduler's statistical characteristics to detect malicious behavior via hypothesis testing [19]. Consequently, the attacking behavior may be detected if the communication rate is changed.

C. Event-Based Attack Model

The energy-efficient state estimation framework proposed in [18] is beneficial for the energy-constrained CPSs, such as personal area network system [20], wireless identification and sensing platform [21], etc. However, as shown in [22], such systems are also vulnerable to be attacked.

In this article, we consider an intelligent attacker that aims at degrading the RSE performance and avoiding the communication-rate-based detection, via modifying the transmission decisions, eavesdroping the measurements and injecting the false feedback information to the sensor. The detailed capacities of the attacker and the model are described as follows.

1) The attacker is aware of the system parameters, including A, C, Q, R, x_0, P_0, and the communication rate γ.
2) The attacker intrudes the sensor and modifies the program to change the transmission decision according to the following event-based attack model:

$$\gamma_k^a = \begin{cases} 0, & \text{if } \mathcal{M}(z_k^a) \geq \delta^a \\ 1, & \text{otherwise} \end{cases} \tag{16}$$

where γ_k^a and δ^a are the transmission decision and the attack threshold, respectively, and z_k^a is the innovation under a specific attack scheduling $\mathcal{M} : \mathbb{R}^m \to \mathbb{R}$, a function which will be designed by the attacker.
3) Since the modified program is executed, the attacker eavesdrops the channel from the sensor to the estimator to obtain the measurements and runs its own estimation algorithm, which will be derived in the next section.
4) At each time slot, the attacker jams the feedback channel and injects the false feedback information, which can be computed by the algorithm presented in the next section.

Remark 2.1: Capacity 1) of the intelligent attacker could be realized by potentially performing system identification [23].

Remark 2.2: The attacker could exploit the sensor's privileges to modify the program [24], compromising the scheduling mechanism.

Remark 2.3: To undermine the estimation accuracy, intuitively and inversely, sensor measurement containing much innovation should not be transmitted. Therefore, the attack mechanism (16) is leveraged, which does not change system characteristics, ensuring the evasion from bad data detection schemes. At the same time, the properly joint-designed trigger function $\mathcal{M}$ and threshold δ^a can maintain communication at a normal rate, achieving the communication rate stealthy.

Remark 2.4: The eavesdropping and estimation algorithm running in 3) and jamming and injecting in 4) can be realized by universal software radio peripheral (USRP) [25], [26]. It is worth noting that all the attack actions in 3) and 4) can be implemented in the compromised sensor, if it has the ability to compute the matrix inversion.

D. Problem of Interest

For executing and evaluating the online attack by the model (16), the following three problems are considered and answered in this article.

1) What is the function $\mathcal{M}$ in the event-based attack model (16)?
2) What is the MMSE estimation algorithm under the designed event-based attack?
3) How to design δ^a to avoid the communication rate detection?
4) How to evaluate the effectiveness of the designed attack?

III. Event-Based Attack Against RSE

In this section, we present the details of a feasible attack mechanism and derive the approximate MMSE estimator under the event-based attack from the viewpoint of the attacker, satisfying the communication rate.

A. Event-Based Attack Mechanism

Denote $\mathrm{I}_k^a = \{\gamma_0^a y_0, \ldots, \gamma_k^a y_k\} \cup \{\gamma_0^a, \ldots, \gamma_k^a\}$ with $\mathrm{I}_{-1}^a = \emptyset$. Define $\hat{x}_{a,k}^-, \hat{x}_{a,k}, e_{a,k}^-, e_{a,k}, P_{a,k}^-, P_{a,k}$, and z_k^a by changing $\mathrm{I}_{k-1}, \mathrm{I}_k$ to $\mathrm{I}_{k-1}^a, \mathrm{I}_k^a$ in (3)–(7) as the *a priori* estimation and the *a posteriori* estimation and their error from the viewpoint of attacker and the innovation under the online attack. According to Section II-C, the attacker runs the approximate MMSE estimation algorithm, which will be obtained in the next subsection. Thus, the innovation under attack $z_k^a = y_k - \mathbb{E}(y_k|\mathrm{I}_{k-1}^a)$ is a Gaussian random variable with zero mean and covariance $CP_{a,k}^- C' + R > 0$. Then there exists a unitary matrix $U_k^a \in \mathbb{R}^{m\times m}$ satisfying

$$U_k^{a'}(CP_{a,k}^- C' + R)U_k^a = \Lambda_{a,k} \tag{17}$$

where $\Lambda_{a,k} = \mathrm{diag}(\lambda_{a,k}^1, \ldots, \lambda_{a,k}^m) \in \mathbb{R}^{m\times m}$, and $\lambda_{a,k}^1, \ldots, \lambda_{a,k}^m \in \mathbb{R}$ are the eigenvalues of $CP_{a,k}^- C' + R$. Define $F_k^a \in \mathbb{R}^{m\times m}$ as $F_k^a = U_k^a \Lambda_{a,k}^{-\frac{1}{2}}$, and then $F_k^a F_k^{a'} = (CP_{a,k}^- C' + R)^{-1}$.

For ease of computing the attack threshold, the Mahalanobis transformation is first used in the mechanism $\mathcal{M}$ by premultiplying $F_k^{a'}$ to decorrelate the coordinates of z_k^a. In the event-based scheduling mechanism (12), the measurement y_k is decided to be sent when the positive infinity norm of $F_k' z_k$ is larger than the threshold, which means that the innovation is better for estimation. On the opposite, here, the measurements will be sent, while the negative infinity norm of $F_k^{a'} z_k^a$ is smaller than the attack threshold. Mathematically, the attack mechanism $\mathcal{M}$ is defined as

$$\mathcal{M}(z_k^a) = \| F_k^{a'} z_k^a \|_{-\infty} \tag{18}$$

where $\|\cdot\|_{-\infty}$ is the negative infinity norm of a vector defined as $\|V\|_{-\infty}=\min(|V_i|)$, where V is a vector and $i=1,\ldots,n$.

B. Approximate MMSE Estimator Under the Event-Based Attack

In this section, we leverage an approximation technique, which is widely used for nonlinear filtering in the literature, e.g., [18] and [27], to assume that the conditional distribution of x_k given I^a_{k-1} is Gaussian, i.e.,

$$f_{x_k}(x|\mathrm{I}^a_{k-1})=\mathcal{N}(\hat{x}^-_{a,k},P^-_{a,k}). \tag{19}$$

Based on this assumption (19), the MMSE estimator with a simple form from the viewpoint of the attacker can be derived.

Before presenting the estimation and its proof, some preliminary results are stated. First, we prove the *a priori* statistical decorrelation property of $F^{a'}_k z^a_k$ as follows.

Lemma 3.1: Conditioning on I^a_{k-1}, $F^{a'}_k z^a_k$ is a Gaussian random variable with zero mean and covariance I_m.

Proof: See the Appendix. ∎

Then we prove the *a posteriori* statistical properties of $F^{a'}_k z^a_k$ conditioned on $\hat{\mathrm{I}}_k:=\mathrm{I}^a_{k-1}\cup\{\gamma^a_k=0\}$, which are shown in Lemmas 3.2 and 3.3. Namely, we explore the properties of $F^{a'}_k z^a_k$, while the measurement y_k is not sent to the estimator/attacker. From the attack mechanism (16) and function (18), $\gamma^a_k=0$ means that

$$\mathcal{M}(z^a_k)=\|F^{a'}_k z^a_k\|_{-\infty}\geq\delta^a \tag{20}$$

which is known from the viewpoint of the attacker. From Lemma 3.1, given I^a_{k-1}, the ith and jth elements of $F^{a'}_k z^a_k$, denoted as $\epsilon^i_{a,k}$ and $\epsilon^j_{a,k}$, are decorrelated with each other for arbitrary $i\neq j$. Therefore, we first investigate the *a posteriori* statistical properties of a one-dimensional Gaussian variable and prove the results in the following lemma.

Lemma 3.2: Let $x\in\mathbb{R}$ be a Gaussian random variable with zero mean and variance $\mathbb{E}[x^2]=\sigma^2$. Denoting $\Delta=\delta^a\sigma$, then $\mathbb{E}[x^2\big||x|\geq\Delta]=\sigma^2(1+\overline{\beta}(\delta^a))$, where

$$\overline{\beta}(\delta^a)=\frac{1}{\sqrt{2\pi}}\cdot\delta^a\cdot e^{-\frac{(\delta^a)^2}{2}}\cdot[Q(\delta^a)]^{-1}. \tag{21}$$

Proof: See the Appendix. ∎

Leveraging the results in Lemma 3.2, we have the following lemma.

Lemma 3.3: $\mathbb{E}[(F^{a'}_k z^a_k)(F^{a'}_k z^a_k)'\big|\hat{\mathrm{I}}^a_k]=[1+\overline{\beta}(\delta^a)]I_m$.

Proof: See the Appendix. ∎

Furthermore, some equalities are required for deriving the approximate estimation and are given as follows.

Lemma 3.4: The following equalities hold:

$$\mathbb{E}[e^-_{a,k}z^{a'}_k\big|\hat{\mathrm{I}}^a_k]=L^a_k\mathbb{E}[z^a_k z^{a'}_k\big|\hat{\mathrm{I}}^a_k]$$

$$\mathbb{E}[(e^-_{a,k}-L^a_k z^a_k)z^{a'}_k\big|\hat{\mathrm{I}}^a_k]=0$$

$$\mathbb{E}[(e^-_{a,k}-L^a_k z^a_k)(e^-_{a,k}-L^a_k z^a_k)'\big|\mathrm{I}^a_{k-1},z^a_k=z]=\widetilde{g}(P^-_{a,k})$$

$$\mathbb{E}[(e^-_{a,k}-L^a_k z^a_k)(e^-_{a,k}-L^a_k z^a_k)'\big|\hat{\mathrm{I}}^a_k]=\widetilde{g}(P^-_{a,k})$$

where $L^a_k=P^-_{a,k}C'[CP^-_{a,k}C'+R]^{-1}$.

Proof: The proof is similar to that of [18, Lem. 3.3] and is omitted here. ∎

Now, we present the approximate MMSE estimation from the viewpoint of the attacker in the following theorem.

Theorem 3.1: Under the Gaussian assumption (19), the approximate MMSE estimator with the attack mechanism (16) and function (18) is given recursively as follows:

1) Time update:

$$\begin{cases}\hat{x}^-_k=A\hat{x}_{k-1}\\ P^-_{a,k}=h(P_{a,k-1}).\end{cases} \tag{22}$$

2) Measurement update:

$$\begin{cases}\hat{x}_k=\hat{x}^-_k+\gamma^a_k L^a_k z^a_k\\ P_{a,k}=\gamma^a_k\widetilde{g}(P^-_{a,k})+(1-\gamma^a_k)\widetilde{g}_{[-\overline{\beta}(\delta^a)]}(P^-_{a,k}).\end{cases} \tag{23}$$

Proof: Based on the results in Lemmas 3.1 and 3.4, it is easy to derive the time update and the measurement update when $\gamma^a_k=1$. Thus, for saving space, we only prove the measurement update when $\gamma^a_k=0$. Since y_k is not transmitted, from the viewpoint of the attacker, the state is estimated as

$$\begin{aligned}\hat{x}_{a,k}&=\mathbb{E}[x_k|\hat{\mathrm{I}}^a_k]=\mathbb{E}[\mathbb{E}[x_k|z^a_k=(F^{a'}_k)^{-1}\epsilon,\mathrm{I}^a_{k-1}]]\\&=\frac{1}{p_{\delta^a}}\int_\Omega\mathbb{E}[x_k|z^a_k=(F^{a'}_k)^{-1}\epsilon,\mathrm{I}^a_{k-1}]f_{F^{a'}_k z^a_k}(\epsilon|\mathrm{I}^a_{k-1})d\epsilon\\&=\int_\Omega[\hat{x}^-_{a,k}+L^a_k(F^{a'}_k)^{-1}\epsilon]\cdot p^{-1}_{\delta^a}\cdot f_{F^{a'}_k z^a_k}(\epsilon|\mathrm{I}^a_{k-1})d\epsilon\\&=\hat{x}^-_{a,k}\int_\Omega p^{-1}_{\delta^a}\cdot f_{F^{a'}_k z^a_k}(\epsilon|\mathrm{I}^a_{k-1})d\epsilon\\&\quad+L^a_k(F^{a'}_k)^{-1}\int_\Omega p^{-1}_{\delta^a}\cdot\epsilon f_{F^{a'}_k z^a_k}(\epsilon|\mathrm{I}^a_{k-1})d\epsilon\end{aligned}$$

where $p_{\delta^a}\triangleq Pr(\|F^{a'}_k z^a_k\|_{-\infty}\geq\delta^a|\mathrm{I}^a_{k-1})$ and $\Omega=\{F^{a'}_k z^a_k\in\mathbb{R}^m:\|F^{a'}_k z^a_k\|_{-\infty}\geq\delta^a\}$. Note that, from Lemma 3.1, we have

$$f_{F^{a'}_k z^a_k}(\epsilon|\hat{\mathrm{I}}^a_k)=\begin{cases}\frac{f_{F^{a'}_k z^a_k}(\epsilon|\mathrm{I}^a_{k-1})}{p_{\delta^a}},&\text{if }\|F^{a'}_k z^a_k\|_{-\infty}\geq\delta^a\\0,&\text{otherwise}\end{cases}$$

which leads to $\int_\Omega p^{-1}_{\delta^a}\cdot f_{F^{a'}_k z^a_k}(\epsilon|\mathrm{I}^a_{k-1})d\epsilon=1$. Furthermore, $f_{F^{a'}_k z^a_k}(\epsilon|\mathrm{I}^a_{k-1})$ is an even function, ϵ is an odd function, and Ω is a symmetric subset with the center $[0,0,\ldots,0]$ in $\mathbb{R}^m$, which leads to $\int_\Omega p^{-1}_{\delta^a}\cdot\epsilon f_{F^{a'}_k z^a_k}(\epsilon|\mathrm{I}^a_{k-1})d\epsilon=0$. Therefore, we have

$$\hat{x}_{a,k}=\mathbb{E}[x_k|\hat{\mathrm{I}}^a_k]=\hat{x}^-_{a,k}.$$

Then from Lemmas 3.2–3.4, the corresponding error covariance $P_{a,k}$ can be computed as

$$\begin{aligned}P_{a,k}&=\mathbb{E}\big[(x_k-\hat{x}_{a,k})(x_k-\hat{x}_{a,k})'\big|\hat{\mathrm{I}}^a_k\big]\\&=\mathbb{E}\big[(x_k-\hat{x}^-_{a,k})(x_k-\hat{x}^-_{a,k})'\big|\hat{\mathrm{I}}^a_k\big]\\&=\mathbb{E}\big[\{(e^-_{a,k}-L^a_k z^a_k)+L^a_k z^a_k\}\cdot\{(e^-_{a,k}-L^a_k z^a_k)+L^a_k z^a_k\}'\big|\hat{\mathrm{I}}^a_k\big]\\&=\mathbb{E}\big[(e^-_{a,k}-L^a_k z^a_k)(e^-_{a,k}-L^a_k z^a_k)'+L^a_k z^a_k z^{a'}_k L^{a'}_k\\&\quad+(e^-_{a,k}-L^a_k z^a_k)z^{a'}_k L^{a'}_k+L^a_k z^a_k(e^-_{a,k}-L^a_k z^a_k)'\big|\hat{\mathrm{I}}^a_k\big]\\&=\widetilde{g}(P^-_{a,k})+L^a_k\mathbb{E}[z^a_k z^{a'}_k|\hat{\mathrm{I}}^a_k]L^{a'}_k\\&=\widetilde{g}(P^-_{a,k})+[1+\overline{\beta}(\delta^a)]L^a_k(F^a_k F^{a'}_k)^{-1}L^{a'}_k\\&=\widetilde{g}(P^-_{a,k})+[1+\overline{\beta}(\delta^a)]L^a_k(CP^-_{a,k}C'+R)L^{a'}_k\\&=\widetilde{g}_{[-\overline{\beta}(\delta^a)]}(P^-_{a,k})\end{aligned}$$

which leads to (23). ∎

C. Design of the Attack Threshold

To avoid being detected by the estimator, the attack threshold should be designed properly, namely it must be computed to guarantee the communication rate. First, a basic property is derived.

Lemma 3.5: While $\delta^a > 0$, we have

$$Pr(\mathcal{M}(z_k^a) \geq \delta^a | \mathrm{I}_{k-1}^a) = [2Q(\delta^a)]^m. \tag{24}$$

Proof: The proof is similar to that of [18, Lemma 3.6] and is omitted here. ■

Then we show the relation between the attack threshold and the communication rate, as well as the relation between the attack threshold and the scheduling threshold, in the following theorem.

Theorem 3.2: For the approximate MMSE estimation under attack, the average communication rate

$$\gamma^a \triangleq \limsup_{T\to+\infty} \frac{1}{T+1} \sum_{k=0}^{T} \mathbb{E}[\gamma_k^a] \tag{25}$$

satisfies

$$\gamma^a = 1 - [2Q(\delta^a)]^m \tag{26}$$

and to avoid being detected, the attack threshold δ^a must satisfies

$$Q(\delta^a) + Q(\delta) = \frac{1}{2} \tag{27}$$

where δ is the event-based scheduling threshold without attack.

Proof: Since γ_k^a is a random variable taking value in $\{0, 1\}$ with $\Pr(\gamma_k^a = 0|\mathrm{I}_{k-1}^a) = \Pr(\mathcal{M}(z_k^a) \geq \delta^a|\mathrm{I}_{k-1}^a)$. From Lemma 3.5, $\Pr(\mathcal{M}(z_k^a) \geq \delta^a|\mathrm{I}_{k-1}^a) = [2Q(\delta^a)]^m$ whatever value I_{k-1}^a takes. Therefore, the distribution of γ_k^a is only dependent on δ^a and can be described by

$$\Pr(\gamma_k^a = 0) = [2Q(\delta^a)]^m,\ \Pr(\gamma_k^a = 1) = 1 - [2Q(\delta^a)]^m.$$

Then the average communication rate is given as

$$\gamma^a = \mathbb{E}\big[\gamma_k^a\big] = 1 - [2Q(\delta^a)]^m.$$

Furthermore, to avoid being detected, the communication rate under the attack should be equal to that without attack (15), namely

$$1 - [2Q(\delta^a)]^m = 1 - [1 - 2Q(\delta)]^m$$

which leads to (27). ■

Remark 3.1: γ_k^a is an i.i.d process with probability γ^a, because γ_k^a is independent with I_{k-1}^a, which contains historical scheduler states $\{\gamma_0, \ldots, \gamma_k\}$.

D. Effectiveness of the Event-Based Attack

In this section, we analyze the effectiveness of the designed attack indirectly from the viewpoint of attack, via comparing the asymptotic expected prediction error covariance of MMSE state estimation with and without attack. First, the lemma needed is given as follows.

Lemma 3.6: Two functions $\gamma + (1-\gamma)\beta(\delta)$ and $\gamma(\delta)$ are strictly decreasing with respect to δ. Moreover, two functions $\gamma^a - (1-\gamma^a)\overline{\beta}(\delta^a)$ and $\gamma^a(\delta^a)$ are strictly increasing with respect to δ^a.

Proof: See the Appendix. ■

We will show that the performance of estimation becomes weakened under the attack in the following theorem, both in the mean-squared stability condition and the asymptotic prediction accuracy.

Theorem 3.3: For the system in Fig. 1, when the system is unstable, there exists $\gamma_c^a > \gamma_c$, where γ_c^a and γ_c and are critical communication rates for the RSE with and without attack, respectively, below which the estimation will diverge and above which the estimation will converge.

Furthermore, when $\gamma = \gamma^a > \gamma_c^a$ or the system is stable, we have

$$\lim_{T\to+\infty} \mathbb{E}[P_{a,k}^-] > \lim_{T\to+\infty} \mathbb{E}[P_k^-]. \tag{28}$$

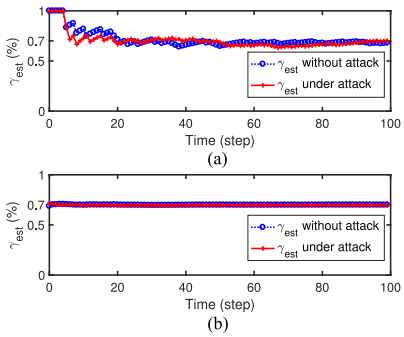

Fig. 2. Unbiased estimation method. Comparison of the communication rate estimation when $\gamma = 0.7$. (a) Communication rate in one-time simulation. (b) Average communication rate over 1000-times simulation.

Proof: From the MMSE estimation (13), (14) under the event-based scheduler, the prediction error covariance without attack is updated as

$$\begin{aligned} P_{k+1}^- &= \gamma_k h \circ \widetilde{g}(P_k^-) + (1-\gamma_k) h \circ \widetilde{g}_{\beta(\delta)}(P_k^-) \\ &= AP_k^-A' + Q - [\gamma_k + (1-\gamma_k)\beta(\delta)] \\ &\quad \cdot AP_k^-C'[CP_k^-C' + R]^{-1}CP_k^-A'. \end{aligned} \tag{29}$$

Similarly, from (22) and (23), the prediction error covariance under attack, from the viewpoint of the attacker, is updated as

$$\begin{aligned} P_{a,k+1}^- &= \gamma_k^a h \circ \widetilde{g}(P_{a,k}^-) + (1-\gamma_k^a) h \circ \widetilde{g}_{[-\overline{\beta}(\delta)]}(P_{a,k}^-) \\ &= AP_{a,k}^-A' + Q - [\gamma_k^a - (1-\gamma_k^a)\overline{\beta}(\delta^a)] \\ &\quad \cdot AP_{a,k}^-C'[CP_{a,k}^-C' + R]^{-1}CP_{a,k}^-A'. \end{aligned} \tag{30}$$

On the other hand, by the Gaussian approximation, it follows that γ_k and γ_k^a are i.i.d processes with the same expectation, namely, $\mathbb{E}[\gamma_k] = \gamma = \gamma^a = \mathbb{E}[\gamma_k^a]$, which is due to the same communication rate. Thus, $\gamma_k + (1-\gamma_k)\beta(\delta)$ and $\gamma_k^a - (1-\gamma_k^a)\overline{\beta}(\delta^a)$ are i.i.d processes with the expectation $\gamma + (1-\gamma)\beta(\delta)$ and $\gamma^a - (1-\gamma^a)\overline{\beta}(\delta^a)$, respectively. According to [28, Th. 2], there exists a $\zeta_c \in [0, 1)$ such that

$$\begin{aligned} &\mathbb{E}[P_k^-] < +\infty, \forall k \ \text{ for } \zeta_c < \gamma + (1-\gamma)\beta(\delta) \leq 1 \\ &\lim_{k\to\infty} \mathbb{E}[P_k^-] = +\infty, \ \text{ for } 0 \leq \gamma + (1-\gamma)\beta(\delta) \leq \zeta_c \\ &\mathbb{E}[P_{a,k}^-] < +\infty, \forall k \ \text{ for } \zeta_c < \gamma^a - (1-\gamma^a)\overline{\beta}(\delta^a) \leq 1 \\ &\lim_{k\to\infty} \mathbb{E}[P_{a,k}^-] = +\infty, \ \text{ for } 0 \leq \gamma^a - (1-\gamma^a)\overline{\beta}(\delta^a) \leq \zeta_c. \end{aligned}$$

From Lemma 3.6, we know there exist a critical event-based threshold δ_c and a critical attack threshold δ_c^a, as well as a critical event-based communication rate γ_c and a critical attack communication rate γ_c^a, satisfying

$$\gamma_c + (1-\gamma_c)\beta(\delta_c) = \zeta_c = \gamma_c^a - (1-\gamma_c^a)\overline{\beta}(\delta_c^a).$$

Thus, combining with the fact that $\zeta_c \in [0, 1)$, we have

$$\gamma_c < \zeta_c < \gamma_c^a$$

which proves that the mean-squared stability condition become weakened under the designed attack.

Now, we consider the situation where the communication rate $\gamma = \gamma^a$ is larger than γ_c^a or the system is stable, namely, both $\mathbb{E}[P_k^-]$ and $\mathbb{E}[P_{a,k}^-]$ will converge. Taking expectation with respect to P_k^- and $P_{a,k}^-$ on both sides of (29) and (30), we get

$$\begin{aligned} \mathbb{E}[P_{k+1}^-] &= \mathbb{E}[h \circ \widetilde{g}_{[\gamma+(1-\gamma)\beta(\delta)]}(P_k^-)] \\ \mathbb{E}[P_{a,k+1}^-] &= \mathbb{E}[h \circ \widetilde{g}_{[\gamma^a-(1-\gamma^a)\overline{\beta}(\delta^a)]}(P_{a,k}^-)] \end{aligned}$$

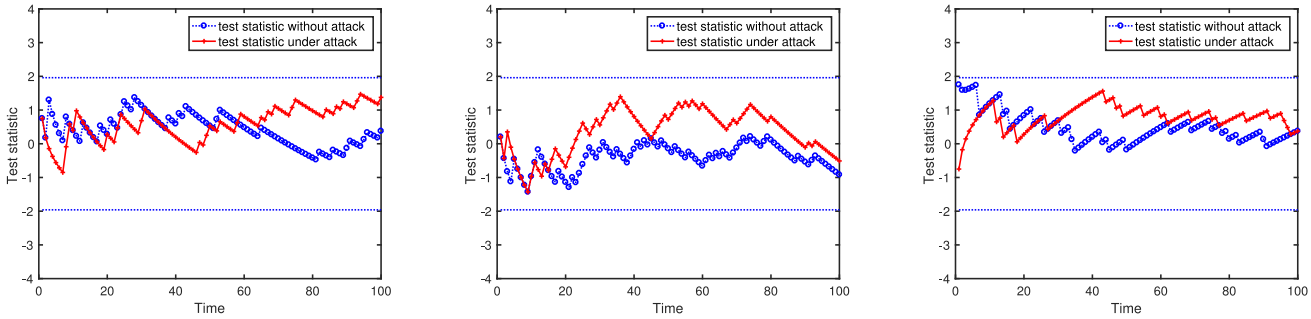

Fig. 3. Hypothesis testing method. Comparison of test statistics and the quantiles under different communication rates. (a) $\gamma = 0.2$. (b) $\gamma = 0.4$. (c) $\gamma = 0.8$.

where $h \circ \widetilde{g}_\lambda(X)$ is strictly increasing in X and strictly decreasing in λ according to the Lemma 1(c) and (d) in [28], respectively. Note that, for the given communication rate, $\gamma + (1-\gamma)\beta(\delta) > \gamma^a - (1-\gamma^a)\overline{\beta}(\delta^a)$. Combining with the fact that $P_0^- = P_{a,0}^-$, we can obtain (28) by induction. ■

Results in the above theorem show that the MMSE estimation quality under attack, derived from the viewpoint of the attacker, becomes worse than that without attack. Since the remote estimator of the system infers wrong information about the not received measurements, its estimation algorithm is no longer the MMSE estimation and, thus, leads to a larger expected prediction error covariance. The exact theoretical gap will be studied in the future work.

Remark 3.2: It is worth noting that the designed attack only changes the scheduling mechanism of sensor measurement, but does not modify its value. Under this attack, system dynamics still holds, which means that the iteration of P_k^- will not change. Then the convergence value of $\mathbb{E}[P_k^-]$ only depends on the random sequence $\{\gamma_k\}_0^\infty$, whose distribution remains the same whether the designed attack is launched or not. In this way, the system estimator cannot observe the performance degradation caused by the attack.

Remark 3.3: It is an interesting but challenging question to analyze the stealthiness of the proposed attack against data integrity detectors [6], [10], which will be further investigated as future work. The false data detector in [6] is designed based on the system model (e.g., autoregressive models $\hat{y}_{k+1} = \sum_{i=k-N}^{k} \alpha_i y_i + \alpha_0$, where α is a time-invariant parameter) and the prediction error $y_{k+1} - \hat{y}_{k+1}$. Intuitively, the proposed attack may be stealthy, because the received measurements under attack have a same distribution as measurements without attack. However, for the false data detector in [10] relying on innovation $z_k = y_k - C\hat{x}_k^-$, it is hard to analyze the stealthiness due to difficulties in comparing the difference between detection metric in normal and compromised systems. First, the analytical evolution of z_k^a at the estimator is hard to derive because of the nonlinear scheduling attack. Besides, it is not direct to build a proper detector based on properties of z_k because of intermittent observations.

IV. Examples

Two examples are presented in this section to illustrate the effectiveness and stealthiness of the proposed attack mechanism.

Example 4.1: Consider a two-dimensional stable system with the following parameters:

$$A = \begin{bmatrix} 0.8 & 0.2 \\ 0 & 0.8 \end{bmatrix}, Q = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \end{bmatrix}, R = 2.$$

First, we verify that the proposed attack mechanism can evade two detection methods based on the statistic of communication: unbiased estimation and hypothesis testing.

On the one hand, for the real-time communication rate unbiased estimation method, which is computing the average number of packets at every iteration step, we do the simulation test for one time and 1000 times in the case of the communication rate $\gamma = 0.7$. We get simulation results by taking the average packet arrival frequency as the estimation and expected estimation of the communication rate for normal and attacking cases, respectively. For one-time simulation in Fig. 2(a), we can see that the communication rates with attack and without attack are close to each other. For 1000-times simulation in Fig. 2(b), the attack does not change the average communication rates.

On the other hand, the real event-triggering process is i.i.d with probability γ. Considering the following hypothesis problem:

$$\mathcal{H}_0 : \gamma_{\text{est}} = \gamma;\ \mathcal{H}_1 : \gamma_{\text{est}} \neq \gamma$$

where γ_{est} is the estimation of the real communication rate, $\mathcal{H}_0$ is the null hypothesis, which means no attack exists, and $\mathcal{H}_1$ is the alternative hypothesis. From [19, Lem. 3], given time horizon k and significance level α, if

$$\frac{\tau - k\gamma + 0.5}{\sqrt{k\gamma(1-\gamma)}} > \mu_{1-\alpha/2}$$

or

$$\frac{\tau - k\gamma + 0.5}{\sqrt{k\gamma(1-\gamma)}} < \mu_{\alpha/2}$$

where τ is the packet arrival frequency, and $\mu_{1-\alpha/2}$ and $\mu_{\alpha/2}$ are $1-\alpha/2$ quantile and $\alpha/2$ quantile of the standard Gaussian distribution, respectively, we accept the null hypothesis. Otherwise, we reject the null hypothesis, believing that there exists an attack (16). In our test, we take $k = 100$ and $\alpha = 0.05$, which means that we reject the null hypothesis with 5% probability when there is no attack. Fig. 3(a)–(c) shows hypothesis testing results when $\gamma = 0.2$, $\gamma = 0.4$, and $\gamma = 0.8$, respectively. They show that the test statistics are both in quantiles with or without attack, so the detector cannot recognize the proposed attack mechanism.

Furthermore, we verify the estimation performance under the attack. We compare the steady expected prediction error covariance $\lim_{k\to\infty} \text{tr}(\mathbb{E}[P_k^-])$ and $\lim_{k\to\infty} \text{tr}(\mathbb{E}[P_{a,k}^-])$ of the estimation without attack and under attack for the communication ranging from 0.1 to 0.9. The curves in Fig. 4 show that the estimation quality is always degraded for any communication rate, which verifies the effectiveness of the designed attack mechanism.

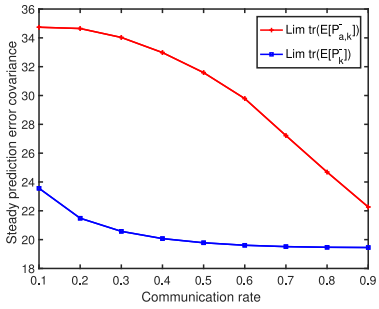

Fig. 4. Comparison of steady prediction error covariance in *Example 1*.

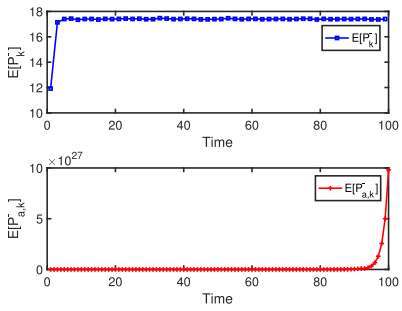

Fig. 5. Expectation of the prediction error covariance for *Example 2* (communication rate $\gamma = \gamma^a = 0.5$).

Example 4.2: Consider an unstable scalar system, whose parameters are $A = 1.4, C = 1, Q = 8,$ and $R = 5$.

From the results in Theorem 3.3 and its proof, we can obtain that the critical communication rates without attack and under attack are $\gamma_c = 0.1198$ and $\gamma_c^a = 0.8715$, respectively, and the corresponding critical thresholds are $\delta_c = 1.5556$ and $\delta_c^a = 1.52$. In Fig. 5, we fix the communication rate as $\gamma = \gamma^a = 0.5$ and plot the expected prediction error covariance $\mathbb{E}[P_k^-]$ without attack and $\mathbb{E}[P_{a,k}^-]$ under attack from the viewpoint of the attacker. It can be found that $\mathbb{E}[P_k^-]$ converges since $\gamma > \gamma_c$ and $\mathbb{E}[P_{a,k}^-]$ diverges since $\gamma^a < \gamma_c^a$, which verifies the validity of Theorem 3.3.

V. Conclusion

In this article, we investigate an event-based attack design against RSE with an arbitrary communication rate constraint. We design an intelligent attack strategy that can degrade the estimation performance by leveraging the online measurement information. The approximate MMSE estimation algorithm is derived with a simple form from the viewpoint of the attacker. The choice of the attack threshold to avoid being detected is presented with an analytical form. We analyze the stability condition for the remote estimation and asymptotic estimation performance under the proposed attack strategy.

Appendix

Proof of Lemma 3.1. From the definition of z_k^a and the assumption (19), z_k^a is zero-mean Gaussian conditioned on I_{k-1}^a, and z_k^a is jointly Gaussian with x_k conditioned on I_{k-1}^a. Furthermore, we have

$$\mathbb{E}[z_k^a z_k^{a'}\big|\mathrm{I}_{k-1}^a] = \mathbb{E}[(Ce_{a,k}^- \nu_k)(Ce_{a,k}^- + \nu_k)'\big|\mathrm{I}_{k-1}^a]$$
$$= C\mathbb{E}[e_{a,k}^- e_{a,k}^{-'}\big|\mathrm{I}_{k-1}^a]C' + R = CP_{a,k}^- C' + R$$
$$\mathbb{E}[(F_k^{a'} z_k^a)(F_k^{a'} z_k^a)'\big|\mathrm{I}_{k-1}^a] = F_k^{a'}\mathbb{E}[z_k^a z_k^{a'}\big|\mathrm{I}_{k-1}^a]F_k^a = I_m$$

which completes the proof. ∎

Proof of Lemma 3.2. The conditional pdf is

$$f_x(x\big||x| \geq \Delta) = \begin{cases} \frac{f_x(x)}{2\int_\Delta^{+\infty} f_t(t)dt}, & |x| \geq \Delta \\ 0, & \text{otherwise} \end{cases}$$

and the conditional covariance is

$$\mathbb{E}[x^2\big||x| \geq \Delta] = 2\int_\Delta^{+\infty} x^2 f_x(x\big||x| \geq \Delta)dx$$
$$= \frac{1}{\int_\Delta^{+\infty} f_x(x)dx}\int_\Delta^{+\infty} x^2 f_x(x)dx.$$

Let $y = x/\sigma$; then we have

$$\mathbb{E}[x^2\big||x| \geq \Delta] = \frac{1}{\int_\Delta^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx}\int_\Delta^{+\infty}\frac{x^2}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx$$
$$= \frac{1}{\int_{\delta^a}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\sigma^2y^2}{2\sigma^2}}d(\sigma y)}\int_{\delta^a}^{+\infty}\frac{\sigma^2y^2}{\sqrt{2\pi}\sigma}e^{-\frac{\sigma^2y^2}{2\sigma^2}}d(\sigma y)$$
$$= \frac{1}{\int_{\delta^a}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dx}\int_{\delta^a}^{+\infty}\frac{\sigma^2y^2}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$$
$$= \frac{\sigma^2}{Q(\delta^a)}\int_{\delta^a}^{+\infty}\frac{y^2}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy.$$

By integrating by parts, we have

$$\int_{\delta^a}^{+\infty}\frac{y^2}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy = Q(\delta^a) + \frac{1}{\sqrt{2\pi}}\delta^a e^{-\frac{(\delta^a)^2}{2}}.$$

Define $\overline{\beta}(\delta^a)$ as in (21); then we obtain

$$\mathbb{E}[x^2\big||x| \geq \Delta] = \frac{\sigma^2}{Q(\delta^a)}\left[Q(\delta^a) + \frac{1}{\sqrt{2\pi}}\delta^a e^{-\frac{(\delta^a)^2}{2}}\right]$$
$$= \sigma^2[1 + \overline{\beta}(\delta^a)]$$

which completes the proof. ∎

Proof of Lemma 3.3. Given I_{k-1}^a, due to the decorrelation of $\epsilon_{a,k}^i$ and $\epsilon_{a,k}^j$, from Lemma 3.2, we have

$$\mathbb{E}[(\epsilon_{a,k}^i)^2\big|\hat{\mathrm{I}}_k^a] = \mathbb{E}[(\epsilon_{a,k}^i)^2\big|\mathrm{I}_{k-1}^a, \| F_k^{a'} z_k^a \|_{-\infty} \geq \delta^a]$$
$$= \mathbb{E}[(\epsilon_{a,k}^i)^2\big|\mathrm{I}_{k-1}^a, |\epsilon_{a,k}^i| \geq \delta^a] = 1 + \overline{\beta}(\delta^a)$$

and

$$\mathbb{E}[\epsilon_{a,k}^i\epsilon_{a,k}^i\big|\hat{\mathrm{I}}_k^a] = \mathbb{E}[\epsilon_{a,k}^i\epsilon_{a,k}^i\big|\mathrm{I}_{k-1}^a, |\epsilon_{a,k}^i| \geq \delta^a, |\epsilon_{a,k}^j| \geq \delta^a] = 0.$$

Thus, we have

$$\mathbb{E}[(F_k^{a'} z_k^a)(F_k^{a'} z_k^a)'\big|\hat{\mathrm{I}}_k^a] = [1 + \overline{\beta}(\delta^a)]I_m$$

which completes the proof. ∎

Proof of Lemma 3.6. First, it is easy to know the monotonicity of $\gamma = 1 - [1 - 2Q(\delta)]^m$ and $\gamma^a = 1 - [2Q(\delta^a)]^m$.

Second, transfer $\gamma^a - (1-\gamma^a)\overline{\beta}(\delta^a)$ to the type of $\gamma^a(1 + \overline{\beta}(\delta^a)) - \overline{\beta}(\delta^a)$. In the following of proof, we differentiate the latter one. The derivative of $\gamma^a(\delta^a)$ in δ^a is

$$\frac{d(\gamma^a)}{d\delta^a} = -m2^m[Q(\delta^a)]^{m-1}Q'(\delta^a)$$
$$= m2^m[Q(\delta^a)]^{m-1}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}$$
$$= \frac{m2^m}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{m-1}$$

and the derivative of $\overline{\beta}(\delta^a) = \frac{1}{\sqrt{2\pi}}\delta^a e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{-1}$ in δ^a is

$$\begin{aligned}\frac{d\overline{\beta}(\delta^a)}{d\delta^a} &= \frac{1}{\sqrt{2\pi}}\Big[e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{-1} - (\delta^a)^2 e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{-1} \\ &\qquad + \delta^a e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{-2}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}\Big] \\ &= \frac{1}{\sqrt{2\pi}[Q(\delta^a)]^2}e^{-\frac{(\delta^a)^2}{2}}\Big[Q(\delta^a) - (\delta^a)^2 Q(\delta^a) \\ &\qquad + \delta^a\frac{1}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}\Big].\end{aligned}$$

Thus, the whole derivative of the function $\gamma^a(1+\overline{\beta}(\delta^a)) - \overline{\beta}(\delta^a)$ is

$$\begin{aligned}&\frac{d(\gamma^a(1+\overline{\beta}(\delta^a)) - \overline{\beta}(\delta^a))}{d\delta^a} \\ &= \gamma'(\delta^a)(1+\bar{\beta}(\delta^a)) + \gamma(\delta^a)\overline{\beta}'(\delta^a) - \bar{\beta}'(\delta^a) \\ &= \frac{m2^m}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{m-1} \\ &\quad + \frac{m2^m}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{m-1}\frac{\delta^a}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{-1} \\ &\quad - \frac{2^m}{\sqrt{2\pi}}[Q(\delta^a)]^m e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{-2}[Q(\delta^a) - (\delta^a)^2 Q(\delta^a) \\ &\quad + \delta^a\frac{1}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}] \\ &= \frac{m2^m}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{m-1} - \frac{2^m}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{m-1} \\ &\quad + \frac{2^m(\delta^a)^2}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{m-1} \\ &\quad + \frac{m2^m\delta^a}{2\pi}e^{-(\delta^a)^2}[Q(\delta^a)]^{m-2} - \frac{2^m\delta^a}{2\pi}e^{-(\delta^a)^2}[Q(\delta^a)]^{m-2} \\ &= \frac{2^m}{\sqrt{2\pi}}e^{-\frac{(\delta^a)^2}{2}}[Q(\delta^a)]^{m-1}(m-1+(\delta^a)^2) \\ &\quad + \frac{2^m\delta^a}{2\pi}e^{-(\delta^a)^2}[Q(\delta^a)]^{m-2}(m-1).\end{aligned}$$

Note that the derivative is larger than 0 when $m \geq 1$, namely, $\gamma^a - (1-\gamma^a)\overline{\beta}(\delta^a)$ is strictly increasing with respect to δ^a. Similarly, we can get the monotonicity of $\gamma + (1-\gamma)\beta(\delta)$. ■

REFERENCES

[1] Y. Qi, P. Cheng, L. Shi, and J. Chen, "Event-based attack against remote state estimation," in *Proc. IEEE Conf. Decis. Control*, 2015, pp. 6844–6849.

[2] K. Sharma and L. M. Saini, "Performance analysis of smart metering for smart grid: An overview," *Renew. Sustain. Energy Rev.*, vol. 49, pp. 720–735, 2015.

[3] X. Koutsoukos *et al.*, "Performance evaluation of secure industrial control system design: A railway control system case study," in *Proc. Resilience Week*, 2016, pp. 101–108.

[4] A. Riel, C. Kreiner, G. Macher, and R. Messnarz, "Integrated design for tackling safety and security challenges of smart products and digital manufacturing," *CIRP Ann.*, vol. 66, no. 1, pp. 177–180, 2017.

[5] S. McLaughlin and P. McDaniel, "SABOT: Specification-based payload generation for programmable logic controllers," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2012, pp. 439–449.

[6] D. I. Urbina *et al.*, "Limiting the impact of stealthy attacks on industrial control systems," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2016, pp. 1092–1105.

[7] Y. Chen, C. M. Poskitt, and J. Sun, "Learning from mutants: Using code mutation to learn and monitor invariants of a cyber-physical system," in *Proc. IEEE Symp. Secur. Privacy*, 2018, pp. 240–252.

[8] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Worst-case stealthy innovation-based linear attack on remote state estimation," *Automatica*, vol. 89, pp. 117–124, 2018.

[9] Y. Guan and X. Ge, "Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 1, pp. 48–59, Mar. 2018.

[10] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 1, pp. 4–13, Mar. 2017.

[11] 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.

[12] 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.

[13] L. Peng, L. Shi, X. Cao, and C. Sun, "Optimal attack energy allocation against remote state estimation," *IEEE Trans. Autom. Control*, vol. 63, no. 7, pp. 2199–2205, Jul. 2018.

[14] J. Qin, M. Li, L. Shi, and X. Yu, "Optimal denial-of-service attack scheduling with energy constraint over packet-dropping networks," *IEEE Trans. Autom. Control*, vol. 63, no. 6, pp. 1648–1663, Jun. 2018.

[15] G.-Z. Yang, *Body Sensor Networks*. Berlin, Germany: Springer, 2006.

[16] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. ACM Int. Workshop Wireless Sens. Netw. Appl.*, 2002, pp. 88–97.

[17] J. Jung, V. Paxson, A. W. Berger, and H. Balakrishnan, "Fast portscan detection using sequential hypothesis testing," in *Proc. IEEE Symp. Secur. Privacy*, 2004, pp. 211–225.

[18] J. Wu, Q.-S. Jia, K. H. Johansson, and L. Shi, "Event-based sensor data scheduling: Trade-off between communication rate and estimation quality," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 1041–1046, Apr. 2013.

[19] H. Zhang, Y. Qi, H. Zhou, J. Zhang, and J. Sun, "Testing and defending methods against dos attack in state estimation," *Asian J. Control*, vol. 19, no. 4, pp. 1295–1305, 2017.

[20] E. Callaway *et al.*, "Home networking with IEEE 802.15.4: A developing standard for low-rate wireless personal area networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 70–77, Aug. 2002.

[21] J. R. Smith, A. P. Sample, P. S. Powledge, S. Roy, and A. Mamishev, "A wirelessly-powered platform for sensing and computation," in *Proc. Int. Conf. Ubiquitous Comput.*, 2006, pp. 495–506.

[22] S. Lim, T. H. Oh, Y. B. Choi, and T. Lakshman, "Security issues on wireless body area network for remote healthcare monitoring," in *Proc. IEEE Int. Conf. Sensor Netw., Ubiquitous, Trustworthy Comput.*, 2010, pp. 327–332.

[23] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 93–109, Feb. 2015.

[24] L. Garcia, F. Brasser, M. H. Cintuglu, A.-R. Sadeghi, O. Mohammed, and S. A. Zonouz, "Hey, my malware knows physics! Attacking PLCs with physical model aware rootkit," in *Proc. Annu. Netw. Distrib. Syst. Secur. Symp.*, 2017, pp. 26–28.

[25] E. Bayraktaroglu, C. King, X. Liu, G. Noubir, R. Rajaraman, and B. Thapa, "Performance of IEEE 802.11 under jamming," *Mobile Netw. Appl.*, vol. 18, no. 5, pp. 678–696, 2013.

[26] J. Staggs, "Adventures in attacking wind farm control networks," in *Proc. Black Hat*, 2017. [Online]. Available: https://www.blackhat.com/docs/us-17/wednesday/us-17-Staggs-Adventures-In-Attacking-Wind-Farm-Control-Networks.pdf

[27] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis, "SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4782–4795, Dec. 2006.

[28] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.