Proceedings of the 35th Chinese Control Conference
July 27-29, 2016, Chengdu, China

A Study of Packet-Reordering Integrity Attack on Remote State Estimation

Ziyang Guo[1], Karl Henrik Johansson[2], Ling Shi[1]

1. Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
E-mail: zguoae@ust.hk,eesling@ust.hk

2. ACCESS Linnaeus Centre and School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
E-mail: kallej@kth.se

Abstract: Recent years have witnessed the surge of interest of security issues in cyber-physical systems. In this paper, we consider malicious cyber attacks in a remote state estimation scenario using time division multiple access communication protocol. A gateway collects the local measurement innovation of each sensor every time instant, while only transmits data packets to a remote estimator during specific time slots. It is assumed that a residue-based detection algorithm is used at the remote side to detect data anomalies. We propose a novel packet-reordering attack strategy where the attacker is able to change the order of the transmitted data packets, without being detected, by compromising the gateway. Furthermore, the evolution of the remote estimation error covariance is derived and the degradation of system performance under the proposed attack is analyzed. Finally, we obtain a sufficient condition for the attack policy to be optimal using the terminal estimation error covariance as a performance metric. Simulations are provided to illustrate the theoretical results.

Key Words: Cyber-Physical System Security, Integrity Attack, Remote State Estimation

1 INTRODUCTION

Cyber-Physical Systems (CPS) are the next generation of engineered systems in which sensing, computing, communication, and control technologies are tightly integrated [1]. Public and private infrastructures today, ranging from large-scale power grids and intelligent transportations to environment monitoring and health care systems, are highly dependent on the safe operation of CPS [2, 3].

With an increasing adoption to safety-critical applications, CPS security has attracted considerable interest from both academic and industrial communities in recent years [4]. Since the measurement and control data in CPS are commonly transmitted through unprotected wireless communication networks, such systems are vulnerable to cyber threats. One recent example is Stuxnet [5], targeted industrial controllers and modified the monitoring data. Attacks like this can lead to severe consequences on national economy, social security or even loss of human lives [6].

The cyber-physical attack space is composed by the adversary's system knowledge, disclosure resources and disruption resources [7]. Two major categories of cyber attacks on CPS: denial-of-service (DoS) and deception attacks are studied in [8]. DoS attacks, which aim at blocking the communication channels and preventing the legitimate access to system components, were studied for a resource-constrained attacker in [9]. Li et al. [10] proposed a game-theoretic framework to study the interactive decision-making process with the energy constraints of both the sensor and the attacker. Agah et al. [11] formulated a cooperative game to study the prevention of DoS attacks in wireless sensor networks and showed that more reliable communication can be achieved through the cooperation between different sensor nodes.

Deception attacks compromise data integrity by modifying transmitted data packets. The statistical characteristics change if the data are modified by a malicious agent. Hence, residue-based χ^2 false-data detectors are widely deployed to monitor such system anomalies [12–14]. However, adversaries may bypass the false-data detector by elaborately design the attack strategy. Liu et al. [13] considered false-data injection attacks against state estimation in power grids. Bai et al. [15] analyzed the estimation performance degradation for an ϵ-stealthiness attacker under arbitrary detection algorithms using information-theoretic approach. The consequence and the reachable estimation error under false-data injection attacks have been analyzed in [16]. Furthermore, a data-driven attack on state estimation was proposed and analyzed in [17].

In general, it is easy for a malicious agent to launch a DoS attack since no system knowledge is needed. However, such an attack is usually easy also to detect [18, 19]. On the other hand, although a deception attack is subtler, more difficult to detect, and able to cause more severe damage, it normally requires perfect information of system parameters [13–15]. To identify the underlying model of a system is usually difficult and not all the attackers have capability to do so, which limits the feasibility of deception attacks in many scenarios.

Motivated by these observations, the system architecture considered in this paper is shown in Fig. 1. At each time instant k, a smart sensor locally processes the raw measurements and sends its local innovation to a gateway. The gateway is allowed to communicate with a remote estimator during certain time periods only. We define $T \in \mathbb{N}$ as the communication period between the gateway and the remote estimator. Consequently, all the collected data packets since the last communication, $z_{\tau+1}$ to $z_{\tau+T}$ $(\tau = \alpha T, \alpha \in \mathbb{N})$, are sent through a wireless network. A malicious attacker may intercept and scramble the data sequence. A false-data detector at the remote side monitors the system behavior and identifies the potential existence of the attacker.

The main contributions of this paper are summarized as follows:

The work by Z. Guo and L. Shi is supported by an HKUST Caltech Partnership FP009. The work by K. H. Johansson is supported by the Knut and Alice Wallenberg Foundation and the Swedish Research Council.

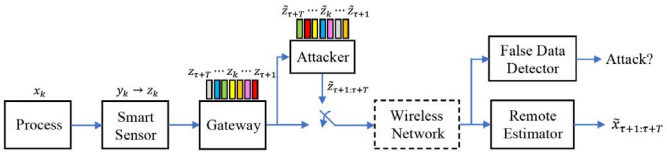

Fig. 1: System Architecture: The attacker is able to reorder the data packets in the buffer of the gateway. The attack degrades the remote estimation performance and might not be detected by the false-data detector.

1) For the above scenario, we propose a novel packet-reordering attack strategy, which does not need to depend on system knowledge and successfully bypasses the false-data detector.
2) We compute the evolution of the estimation error covariance at the remote estimator and analyze the degradation of system performance (**Theorem 1**).
3) When using terminal estimation error covariance as performance metric, the optimal packet-reordering attack strategy is explicitly derived (**Theorem 2**).

The remainder of the paper is organized as follows. Section II presents the system architecture and proposes a novel packet-reordering attack strategy. Section III derives the evolution of remote estimation error covariance. Section IV provides the optimal attack strategy for terminal error as the objective function. Simulation results are provided in Section V. Some concluding remarks are given in the end.

Notations: $\mathbb{N}$ and $\mathbb{R}$ denote the sets of positive integers and real numbers. $\mathbb{R}^n$ is the n-dimensional Euclidean space. $\mathbb{S}_+^n$ and $\mathbb{S}_{++}^n$ are the sets of $n \times n$ positive semi-definite and positive definite matrices. When $X \in \mathbb{S}_+^n$, we simply write $X \geq 0$ (or $X > 0$ if $X \in \mathbb{S}_{++}^n$). $\mathcal{N}(\mu, \Sigma)$ denotes Gaussian distribution with mean μ and covariance matrix Σ. The superscript $'$ and $\mathrm{Tr}(\cdot)$ stand for the transpose and the trace of a matrix. $\mathbb{E}[\cdot]$ denotes the expectation of a random variable. For functions f, f_1, f_2 with appropriate domain, $f_1 \circ f_2(x)$ stands for the function composition $f_1(f_2(x))$, and $f^n(x) \triangleq f(f^{n-1}(x))$. The notation $z_{i:j}$ represents the set $\{z_i, z_{i+1}, \ldots, z_j\}$.

2 Problem Setup

2.1 Process Model

Consider a discrete-time linear time-invariant (LTI) process:

$$x_{k+1} = Ax_k + w_k, \tag{1}$$
$$y_k = Cx_k + v_k, \tag{2}$$

where $k \in \mathbb{N}$ is the time index, $x_k \in \mathbb{R}^n$ the vector of process states, $y_k \in \mathbb{R}^m$ the vector of sensor measurements, $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are zero-mean i.i.d. Gaussian noises with covariances $Q \geq 0$ and $R > 0$, respectively. The initial state x_0 is zero-mean Gaussian with covariance matrix $\Sigma_0 \geq 0$, and is independent of w_k and v_k for all $k \geq 0$. The pair (A, C) is observable and (A, Q) is controllable.

2.2 Smart Sensor

In CPS, sensors are typically equipped with on-board processor and have computation capabilities [20]. At each time k, we assume that the sensor first performs local estimation based on the process measurements and then transmits its innovation to the remote estimator. Let us define $\hat{x}_k^-$ and $\hat{x}_k$ as the *a priori* and the *a posteriori* Minimum Mean Squared Error (MMSE) estimates of the state x_k at the sensor side, respectively, and P_k^- and P_k as the corresponding error covariances. They can be calculated by a Kalman filter:

$$\begin{aligned}
\hat{x}_k^- &= A\hat{x}_{k-1}, \\
P_k^- &= AP_{k-1}A' + Q, \\
K_k &= P_k^- C'(CP_k^- C' + R)^{-1}, \\
\hat{x}_k &= \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-), \\
P_k &= (I - K_k C)P_k^-,
\end{aligned}$$

where the recursion starts from $\hat{x}_0 = 0$ and $P_0 = \Sigma_0 \geq 0$.

To simplify the subsequent discussion, we define the following operators $h, \tilde{g} : \mathbb{S}_+^n \to \mathbb{S}_+^n$ as

$$\begin{aligned}
h(X) &\triangleq AXA' + Q, \\
\tilde{g}(X) &\triangleq X - XC'(CXC' + R)^{-1}CX.
\end{aligned}$$

It is well known that the gain and the error covariance of the Kalman filter converge to steady state values from any initial condition exponentially fast, i.e.,

$$\widehat{P} \triangleq \lim_{k \to +\infty} P_k, \tag{3}$$
$$\overline{P} \triangleq \lim_{k \to +\infty} P_k^-, \tag{4}$$
$$K \triangleq \overline{P}C'(C\overline{P}C' + R)^{-1}, \tag{5}$$

where $\widehat{P}$ and $\overline{P}$ are the unique positive semi-definite solution of $\tilde{g} \circ h(X) = X$ and $h \circ \tilde{g}(X) = X$, respectively.

The innovation transmitted at time k is defined as:

$$z_k = y_k - C\hat{x}_k^-.$$

It has the following properties:

Lemma 1.

1) z_k follows the Gaussian distribution $\mathcal{N}(0, \mathcal{P})$, where $\mathcal{P} = C\overline{P}C' + R$.
2) z_i and z_j are independent $\forall i \neq j$.

Proof: See [21]. ■

Remark 1. *The sensor sends the innovation z_k rather than the measurement y_k or the local estimate $\hat{x}_k$ due to communication efficiency and detection convenience. A steady-state Gaussian distribution of the innovation z_k facilitates the direct detection of abnormal data.*

2.3 False-Data Detector

False-data detectors are widely used to monitor system behavior and detect cyber attacks by checking the statistical characteristics of received data packets [22]. In this paper, a residue-based χ^2 detector is deployed at the remote estimator to detect system anomalies. At time k, the χ^2 detector performs the hypothesis test:

$$g_k = \sum_{i=k-\mathcal{J}+1}^{k} z_i' \mathcal{P}^{-1} z_i \underset{H_1}{\overset{H_0}{\lessgtr}} \delta,$$

where $\mathcal{J}$ is the detection window size, δ is the threshold, the null hypotheses H_0 means that the system is operating

normally, while the hypotheses H_1 is the contrary. Note that the false alarm rate can be easily calculated since g_k is χ^2 distributed with $m\mathcal{J}$ degrees of freedom [23]. If g_k exceeds the threshold, the detector will trigger an alarm.

2.4 Packet-Reordering Attack Strategy

According to the TCP/IP protocol, a unique sequence number (SYN) of each data packet enable the remote estimator to distinguish the order of the received data. We propose a packet-reordering attack strategy where the malicious attacker has the capability to

1) intercept all the transmitted data packets; and
2) modify the sequence number in the TCP header.

Suppose that all the data packets collected by the gateway are sent out during the transmission period. Consequently, T innovations are buffered in the gateway since last communication. The attacker is able to arbitrarily shift these T innovations such that the order of the innovation used by the remote estimator is changed, i.e., at time k, the estimator uses the (wrong) innovation $z_{l(k)}$ rather than z_k, where $l : \mathbb{N} \to \mathbb{N}$ is a function defining the reordering at current time k.

To quantify the estimation performance, we define $\tilde{x}_k^-$ and $\tilde{x}_k$ as the *a priori* and the *a posteriori* MMSE estimates of the state x_k at the remote estimator, and $\tilde{P}_k^-$ and $\tilde{P}_k$ as the corresponding error covariances. The problems we are most interested in contain the estimation error covariance under the proposed packet-reordering attack and the existence of the optimal attack strategy in the sense that the estimation error is maximized, which will be analyzed in the following two sections.

Remark 2. *The proposed attack strategy is quite easy to implement since it does not need the knowledge of system parameters. It suffices to hijack the gateway, which may occur in practice [24, 25].*

Remark 3. *The proposed packet-reordering attack strategy can bypass the χ^2 false-data detector since the distribution of the modified innovation still follows the same Gaussian distribution $\mathcal{N}(0, \mathcal{P})$.*

3 Performance Analysis under Packet-Reordering Attack

To investigate the degradation of system performance under the packet-reordering attack, we focus on the evolution of the remote estimation error covariance in this section.

Consider the LTI process (1)–(2) under the proposed packet-reordering attack. The update of the state estimate at the remote estimator follows

$$\tilde{x}_k^- = A\tilde{x}_{k-1}, \tag{6}$$
$$\tilde{x}_k = \tilde{x}_k^- + Kz_{l(k)}, \tag{7}$$

where the fixed gain K is given in (5), $z_{l(k)}$ represents the shifted innovation used by the remote estimator at time k. The state estimate $\tilde{x}_k$ produced by the remote estimator deviates from the true system state and the evolution of the estimation error covariance under the packet-reordering attack is summarizes in the following theorem.

Theorem 1. *The iteration of the error covariance at the remote estimator under the packet-reordering attack follows the recursion*

$$\begin{aligned}\tilde{P}_k^- &= A\tilde{P}_{k-1}A' + Q - \sum_{i=0}^{k-1} \mathbf{1}_{l(k-1-i)} \\ &\quad \Big\{A^{i+1}KC[A(I-KC)]^{l(k-1-i)-k}Q \\ &\quad + Q[(I-KC)'A']^{l(k-1-i)-k}C'K'(A')^{i+1}\Big\}, \end{aligned} \tag{8}$$

$$\begin{aligned}\tilde{P}_k &= \tilde{P}_k^- + \Delta - \mathbf{1}_{l(k)}\Big\{\overline{P}[(I-KC)'A']^{l(k)-k}C'K' \\ &\quad + KC[A(I-KC)]^{l(k)-k}\overline{P}\Big\} - \big(1-\mathbf{1}_{l(k)}\big) \\ &\quad \Big\{A^{k-l(k)}\overline{P}C'K' + KC\overline{P}(A')^{k-l(k)}\Big\}, \end{aligned} \tag{9}$$

where $\Delta = \overline{P}C'(C\overline{P}C' + R)^{-1}C\overline{P} \geq 0$, and

$$\mathbf{1}_{l(k)} = \begin{cases} 1 & \text{if } l(k) > k \\ 0 & \text{if } l(k) \leq k. \end{cases}$$

Proof: According to the process model (1)–(2) and the iteration of the state estimate (6)– (7), the error covariances at the remote estimator side can be obtained as

$$\begin{aligned}\tilde{P}_k^- &= \mathbb{E}[(x_k - \tilde{x}_k^-)(x_k - \tilde{x}_k^-)'] \\ &= A\tilde{P}_{k-1}A' + Q + \mathbb{E}[A(x_{k-1} - \tilde{x}_{k-1})w_{k-1}'] \\ &\quad + \mathbb{E}[w_{k-1}(x_{k-1} - \tilde{x}_{k-1})'A'], \end{aligned} \tag{10}$$

$$\begin{aligned}\tilde{P}_k &= \mathbb{E}[(x_k - \tilde{x}_k)(x_k - \tilde{x}_k)'] \\ &= \tilde{P}_k^- + K(C\overline{P}C' + R)K' - \mathbb{E}[(x_k - \tilde{x}_k^-)z_{l(k)}'K'] \\ &\quad - \mathbb{E}[Kz_{l(k)}(x_k - \tilde{x}_k^-)']. \end{aligned} \tag{11}$$

In order to obtain the iteration of the remote estimation error covariance, we need to calculate the last two terms of (10) and (11). We first focus on the *a posteriori* estimation error covariance $\tilde{P}_k$ and evaluate

$$\begin{aligned}&x_k - \tilde{x}_k^- \\ &= A^k(x_0 - \hat{x}_0^-) + \sum_{i=0}^{k-1} A^i w_{k-1-i} - \sum_{i=0}^{k-1} A^{i+1}Kz_{l(k-1-i)}, \end{aligned} \tag{12}$$

based on the assumption $\hat{x}_0^- = \tilde{x}_0^-$. Since z_k is an i.i.d. Gaussian random variable and $z_{l(k)}$ can be regarded as a sequence of z_k out of order, we obtain that $\mathbb{E}[z_{l(i)}z_{l(j)}'] = 0$, $\forall i \neq j$. Thus, we only concern about the correlation between the first two terms of (12) and $z_{l(k)}$. Then, according to

$$x_k - \hat{x}_k^- = A(I-KC)(x_{k-1} - \hat{x}_{k-1}^-) + w_{k-1} - AKv_{k-1},$$

we further represent $z_{l(k)}$ in the form of

$$\begin{aligned}z_{l(k)} &= C[A(I-KC)]^{l(k)}(x_0 - \hat{x}_0^-) \\ &\quad + \sum_{i=0}^{l(k)-1} C[A(I-KC)]^i w_{l(k)-1-i} + V, \end{aligned} \tag{13}$$

where $V = v_{l(k)} - \sum_{i=0}^{l(k)-1} C[A(I-KC)]^i AKv_{l(k)-1-i}$ is independent of $x_k - \tilde{x}_k^-$. It now follows that the second last term of (11) can be written as

$$\begin{aligned}
&\mathbb{E}\left[(x_k - \tilde{x}_k^-)z'_{l(k)}K'\right] \\
&= \mathbb{E}\left[\left\{A^k(x_0 - \hat{x}_0^-) + \sum_{i=0}^{k-1} A^i w_{k-1-i}\right\}\left\{C[A(I-KC)]^{l(k)}\right.\right. \\
&\left.\left.\times (x_0 - \hat{x}_0^-) + \sum_{i=0}^{l(k)-1} C[A(I-KC)]^i w_{l(k)-1-i}\right\}' K'\right],
\end{aligned} \tag{14}$$

from which we can obviously see that the error covariance generated by the process noise w depends on the relationship between $l(k)$ and k.

Thus, in the case $l(k) > k$, (14) becomes

$$\begin{aligned}
&\mathbb{E}\left[(x_k - \tilde{x}_k^-)z'_{l(k)}K'\right] \\
&= \left\{A^k\mathbb{E}[(x_0 - \hat{x}_0^-)(x_0 - \hat{x}_0^-)'][(I-KC)'A']^{l(k)}\right. \\
&\left. + \sum_{i=0}^{k-1} A^i\mathbb{E}[w_{k-1-i}w'_{k-1-i}][(I-KC)'A']^{l(k)-k+i}\right\}C'K' \\
&= \left\{A^k\overline{P}[(I-KC)'A']^k + \sum_{i=0}^{k-1} A^i Q[(I-KC)'A']^i\right\} \\
&\times [(I-KC)'A']^{l(k)-k}C'K' \\
&= \overline{P}[(I-KC)'A']^{l(k)-k}C'K',
\end{aligned} \tag{15}$$

where the last equality is due to the fact that $\overline{P}$ is the unique positive semi-definite fixed point of $h \circ \tilde{g}$, i.e., $(h \circ \tilde{g})^n(\overline{P}) = \overline{P}$. Similarly, we obtain

$$\mathbb{E}\left[Kz_{l(k)}(x_k - \tilde{x}_k^-)'\right] = KC[A(I-KC)]^{l(k)-k}\overline{P}. \tag{16}$$

In the case $l(k) \leq k$, (14) becomes

$$\begin{aligned}
&\mathbb{E}\left[(x_k - \tilde{x}_k^-)z'_{l(k)}K'\right] \\
&= \left\{\sum_{i=0}^{l(k)-1} A^{k-l(k)+i}\mathbb{E}[w_{l(k)-1-i}w'_{l(k)-1-i}][(I-KC)'A']^i\right. \\
&\left. + A^k\mathbb{E}[(x_0 - \hat{x}_0^-)(x_0 - \hat{x}_0^-)'][(I-KC)'A']^{l(k)}\right\}C'K' \\
&= A^{k-l(k)}\left\{\sum_{i=0}^{l(k)-1} A^i Q[(I-KC)'A']^i\right. \\
&\left. + A^{l(k)}\overline{P}[(I-KC)'A']^{l(k)}\right\}C'K' \\
&= A^{k-l(k)}\overline{P}C'K'.
\end{aligned} \tag{17}$$

Similarly, we obtain

$$\mathbb{E}\left[Kz_{l(k)}(x_k - \tilde{x}_k^-)'\right] = KC\overline{P}(A')^{k-l(k)}. \tag{18}$$

Then the iteration of the *a posteriori* estimation error covariance $\tilde{P}_k$ can be obtained by substituting (15)–(18) into (11).

Now we move to analyze the *a priori* estimation error covariance $\tilde{P}_k^-$ and first evaluate

$$\begin{aligned}
&x_{k-1} - \tilde{x}_{k-1} \\
&= A^{k-1}(x_0 - \hat{x}_0^-) + \sum_{i=0}^{k-2} A^i w_{k-2-i} - \sum_{i=0}^{k-1} A^i K z_{l(k-1-i)},
\end{aligned}$$

which follows from (12). Since w_k is an i.i.d Gaussian noise, we obtain that $\mathbb{E}[w_i w'_j] = 0$, $\forall i \neq j$. According to (13), when $l_{k-1-i} < k$, $i \in \{0, 1, \ldots, k-1\}$, one has

$$\mathbb{E}[z_{l(k-1-i)})w'_{k-1}] = 0,$$

however, when $l_{k-1-i} > k$, $i \in \{0, 1, \ldots, k-1\}$,

$$\mathbb{E}[z_{l(k-1-i)})w'_{k-1}] = C[A(I-KC)]^{l(k-1-i)-k}Q.$$

It now follows that the second last term of (10) is written as

$$\begin{aligned}
&\mathbb{E}[A(x_{k-1} - \tilde{x}_{k-1}^-)w'_{k-1}] \\
&= \mathbb{E}[A(-\sum_{i=0}^{k-1} A^i K z_{l(k-1-i)})w'_{k-1}] \\
&= -\sum_{i=0}^{k-1} \mathbf{1}_{l(k-1-i)} A^{i+1} KC[A(I-KC)]^{l(k-1-i)-k}Q.
\end{aligned} \tag{19}$$

Similarly, we obtain

$$\begin{aligned}
&\mathbb{E}[w_{k-1}(x_{k-1} - \tilde{x}_{k-1}^-)'A'] \\
&= -\sum_{i=0}^{k-1} \mathbf{1}_{l(k-1-i)} Q[(I-KC)'A']^{l(k-1-i)-k}C'K'(A')^{i+1}.
\end{aligned} \tag{20}$$

Furthermore, the iteration of the *a priori* estimation error covariance $\tilde{P}_k^-$ can be calculated by substituting (19)– (20) into (10), which completes the proof. ■

4 Optimal Packet-reordering Attack Strategies

Based on the aforementioned attack policy and the evolution of the estimation error covariance, we find out the optimal packet-reordering attack strategy and potential attack consequences in this section.

We use the terminal estimation error covariance as an objective function to measure the system performance, i.e.,

$$J(T) \triangleq \mathrm{Tr}(\tilde{P}_T),$$

where we recall that T is the communication period of the gateway communication protocol (as well as the attack period of packet-reordering attack). The optimal attack strategy, in the sense of maximum objective function, is investigated next, but let us first introduce the following lemma which illustrates the inherent consequence of shifting a sequence of innovations.

Lemma 2. *For a sequence of innovations within the time interval* $[\tau+1, \tau+T], \tau = \alpha T, \alpha \in \mathbb{N}$*, switching any two innovations, e.g.,* $z_{l(k_1)}$ *and* $z_{l(k_2)}$ *with* $k_1, k_2, l(k_1), l(k_2) \in [1, T]$*, leads to a larger terminal estimation error covariance if at least one of the following conditions is satisfied:*

1) $A > 0$ *is a scalar, and* $k_2 > k_1$, $l(k_2) > l(k_1)$*; or*
2) $\mathrm{Tr}[(A^{2T-k-j} - A^{2T-k-t-j})(A^{j-i} - I)\Delta] \geq 0$ $\forall i, j, k, k+t \in [1, T]$, $k+t > k$, $j > i$, *and* $k_2 > k_1$, $l(k_2) > l(k_1)$.

Proof: Suppose that attacks start from steady state. Without loss of generality, we let $\tau = 0$ and consider time interval $\{1, 2, \ldots, k-1, k, k+1, \ldots, k+t-1, k+t, k+t+1, \ldots, T-1, T\}$ with innovations $\{z_{l(1)}, z_{l(2)}, \ldots, z_{l(k-1)}, z_i, z_{l(k+1)}, \ldots, z_{l(k+t-1)}, z_j, z_{l(k+t+1)}, \ldots, z_{l(T-1)}, z_{l(T)}\}$. What we need to prove is that a larger $J(T)$ will be obtained if we change the order of z_i and z_j for any $1 \leq k < k+t \leq T$ and $i < j$. The following six situations are needed to be taken into consideration: (a) $i < j \leq k < k+t$; (b) $i \leq k < j \leq k+t$; (c) $i \leq k < k+t < j$; (d) $k < i < j \leq k+t$; (e) $k < i \leq k+t < j$; (f) $k < k+t < i < j$. The proof is straightforward and omitted, but the results of above situations are the same and given by

$$\mathrm{Tr}[2(A^{2T-k-j} - A^{2T-k-t-j})(A^{j-i} - I)\Delta]. \quad (21)$$

For the scalar case, no matter the system is stable ($0 < A < 1$) or not ($A \geq 1$), the terminal estimation error covariance after switching is always larger than the original one. For the matrix case, the same result holds if (21) ≥ 0 $\forall i, j, k, k+t \in [1, T]$, $k+t > k$, $j > i$, which completes the proof. ■

According to the switching principle stated in Lemma 2, we can now easily find the optimal attack strategy, which is summarized in the following theorem.

Theorem 2. *For the terminal estimation error covariance as the objective function, i.e.,* $\gamma = 1$*, the optimal packet-reordering attack strategy for the time interval* $[\tau+1, \tau+T], \tau = \alpha T, \alpha \in \mathbb{N}$*, is to rearrange the innovations in a reverse chronological order if*

1) $A > 0$ *is a scalar; or*
2) $\forall i, j, p, q \in [1, T]$, $q > p$, $j > i$,

$$\mathrm{Tr}[(A^{2T-p-j} - A^{2T-q-j})(A^{j-i} - I)\Delta] \geq 0. \quad (22)$$

Proof: Consider any time horizon $[\tau+1, \tau+T]$, $\forall \tau = \alpha T$, $\alpha \in \mathbb{N}$. According to Lemma 2, no matter we consider a scalar $A > 0$ or a matrix A satisfying (22), if the sequence of the modified innovations $z_{l(\tau+1)}, z_{l(\tau+2)}, \ldots, z_{l(\tau+T-1)}, z_{l(\tau+T)}$ is not in the order $z_{\tau+T}, z_{\tau+T-1}, \ldots, z_{\tau+2}, z_{\tau+1}$, one can always find at least one attack strategy which yields a larger terminal estimation error by switching the innovation $z_{l(\tau+t_1)}$ and $z_{l(\tau+t_2)}$, where $l(\tau+t_2) > l(\tau+t_1)$, $t_2 > t_1$. Therefore, a reverse chronological ordered sequence of innovations leads to the worst estimation performance, which is thus the optimal packet-reordering attack strategy. ■

For any given process, satisfaction of the second condition in Lemma 2 can be checked. However, it is not easy to check for large T, which motivates the sufficient condition in the following corollary.

Corollary 1. *Condition 2 of Lemma 2 and Theorem 2 holds if* $A \geq 0$.

Proof: According to Schur decomposition [26], for each $A \in \mathbb{R}^{n \times n}$ with eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_n$ in any prescribed order, there exists an orthogonal matrix U satisfying $U'U = UU' = I$ such that $U'AU = D$, where D is an upper triangular matrix with diagonal entries $d_{ii} = \lambda_i$. Then for $A \geq 0$, it is easy to obtain $U'A^\ell U = D^\ell$ and $(D^\ell)' = (U'A^\ell U)' = U'A^\ell U = D^\ell$. In this case, D is a diagonal matrix with $d_{ii} = \lambda_i \geq 0$ such that

$$D^{2T-k-t-j}(D^t - I)(D^{j-i} - I) \geq 0$$

always holds, which is equivalent to

$$A^{2T-k-i} + A^{2T-k-t-j} - A^{2T-k-j} - A^{2T-k-t-i} \geq 0.$$

Since $\Delta = \overline{P}C'(C\overline{P}C' + R)^{-1}C\overline{P} \geq 0$,

$$\mathrm{Tr}[(A^{2T-k-j} - A^{2T-k-t-j})(A^{j-i} - I)\Delta] \geq 0$$

is true. ■

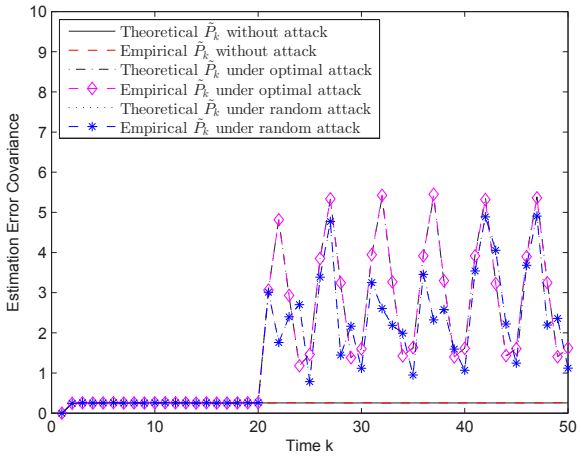

Fig. 2: Estimation error covariance for stable process.

Remark 4. *When the attacker launches attacks repeatedly, the objective becomes* $J(\mathcal{K}) \triangleq \mathrm{Tr}(\tilde{P}_{\mathcal{K}})$*, where* $\mathcal{K} = \beta T, \beta \in \mathbb{N}$*. The optimal attack in this case is still the reverse chronological one since the remote estimation error covariances* $\tilde{P}_{iT+1}, \tilde{P}_{iT+2}, \ldots \tilde{P}_{(i+1)T}$ *can be represented as a linear increasing function of* $\tilde{P}_{iT}$*,* $\forall i \in \{0, 1, \ldots, \beta-1\}$*. Therefore, the terminal error during* β *attack periods is maximized when terminal error in each period is maximized.*

5 Simulation Examples

To demonstrate the analytical results, we present some numerical simulations in this section.

We first consider a stable process with parameters $A = 0.8, C = 2, Q = 1.5, R = 1.2$. The simulation result is shown in Fig. 2. It is assumed that the sensor is able to communicate with the remote estimator once every five time slots, i.e., $T = 5$. During time interval $[0, 20]$, the remote estimator runs a Kalman filter and enters steady state. At $k = 21$, a malicious attacker launches a packet-reordering attack. The optimal attack strategy which rearranges the innovations in a reverse chronological order corresponds to the magenta diamond line in the figure. The blue star line represents that the attacker randomly switches the data sequence. The red dashed line stands for the estimation error covariance without attack. In this case, the terminal estimation error of each shifting period under the optimal packet-reordering attack achieves the largest. Moreover, the error covariance converges when the process is stable.

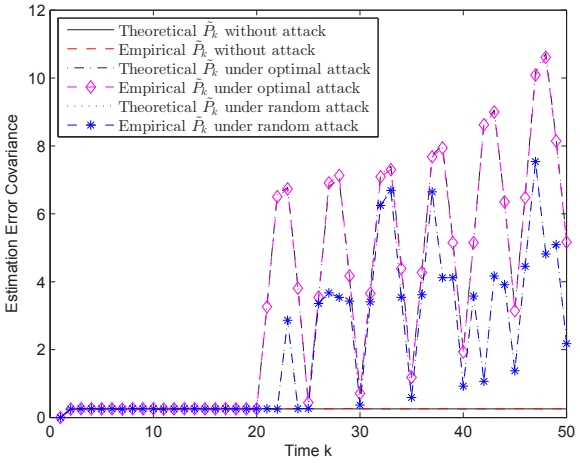

Fig. 3: Estimation error covariance for unstable process.

Fig. 3 shows the error covariances at the remote estimator under the optimal packet-reordering attack, random attack strategies for an unstable process with parameters $A = 1.05, C = 2, Q = 1, R = 1$. It can be observed that the terminal estimation error is maximized by the optimal attack strategy. It is also worth noticing that the error covariance diverges as time goes to infinity in this case.

6 Conclusion

In this paper, we proposed a novel packet-reordering attack strategy which can successfully bypass a χ^2 false-data detector in a remote state estimation scenario. We investigated the evolution of the remote estimation error covariance under the attack and analyzed the degradation of system performance. Furthermore, we proved that rearranging the innovation sequence in a reverse chronological order is the optimal attack strategy, which yields the largest terminal estimation error covariance, if $A \geq 0$. Simulation examples demonstrated the analytical results. Future work includes the analysis of optimal attack strategies for other objective functions and the development of detection criteria to prevent these attacks.

References

[1] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, 2012.

[2] H. Sandberg, S. Amin, and K. Johansson, "Cyberphysical security in networked control systems: An introduction to the issue," *IEEE Control Systems*, vol. 35, no. 1, pp. 20–23, 2015.

[3] R. Poovendran, K. Sampigethaya, S. K. S. Gupta, I. Lee, K. V. Prasad, D. Corman, and J. L. Paunicka, "Special issue on cyber–physical systems," *Proceedings of the IEEE*, vol. 1, no. 100, pp. 1–12, 2012.

[4] E. Lee *et al.*, "Cyber physical systems: Design challenges," in *11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing*, 2008, pp. 363–369.

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

[6] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber–physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[7] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 24–45, 2015.

[8] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *The 28th International Conference on Distributed Computing Systems Workshops*, 2008, pp. 495–500.

[9] A. Gupta, C. Langbort, and T. Basar, "Optimal control in the presence of an intelligent jammer with limited actions," in *49th IEEE Conference on Decision and Control*, 2010, pp. 1096–1101.

[10] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, "Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2831–2836, 2015.

[11] A. Agah and S. K. Das, "Preventing dos attacks in wireless sensor networks: A repeated game theory approach," *Network Security*, vol. 5, no. 2, pp. 145–153, 2007.

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

[13] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, p. 13, 2011.

[14] Y. Mo and B. Sinopoli, "Integrity attacks on cyber-physical systems," in *Proceedings of the 1st international conference on High Confidence Networked Systems*, 2012, pp. 47–54.

[15] C.-Z. Bai, F. Pasqualetti, and V. Gupta, "Security in stochastic control systems: Fundamental limitations and performance bounds."

[16] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, 2011.

[17] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: A data driven approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1102–1114, 2015.

[18] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *Proceedings of ACM International Symposium on Mobile Ad hoc Networking and Computing*, 2005, pp. 46–57.

[19] M. Li, I. Koutsopoulos, and R. Poovendran, "Optimal jamming attacks and network defense policies in wireless sensor networks," in *26th IEEE International Conference on Computer Communications*, 2007, pp. 1307–1315.

[20] J.-M. Favennec, "Smart sensors in industry," *Journal of Physics E: Scientific Instruments*, vol. 20, no. 9, p. 1087, 1987.

[21] B. D. Anderson and J. B. Moore, *Optimal Filtering*. Courier Corporation, 2012.

[22] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, no. 6, pp. 601–611, 1976.

[23] B. Brumback and M. D. Srinath, "A chi-square test for fault-detection in kalman filters," *IEEE Transactions on Automatic Control*, vol. 32, no. 6, pp. 552–554, 1987.

[24] W. R. Cheswick, S. M. Bellovin, and A. D. Rubin, *Firewalls and Internet Security: Repelling the Wily Hacker*. Addison Wesley Longman Publishing Co., Inc., 2003.

[25] X. Hu and Z. M. Mao, "Accurate real-time identification of IP prefix hijacking," in *IEEE Symposium on Security and Privacy*, 2007, pp. 3–17.

[26] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge university press, 2012.