2017 American Control Conference
Sheraton Seattle Hotel
May 24–26, 2017, Seattle, USA

Optimal Innovation-Based Deception Attack on Remote State Estimation

Shuang Wu*, Ziyang Guo*, Dawei Shi†, Karl Henrik Johansson‡, Ling Shi*

Abstract— **The security issue in cyber-physical systems has attracted growing interests in the last decades. This paper considers how false data injection attack can degrade the estimation quality of a remote state estimation system. In this system, smart sensors measure a dynamic process and send preprocessed data through a communication network to a remote estimator to estimate the process. It is assumed that there are malicious attackers in the communication network, who are able to obtain and falsify all the data sent by the sensors. It is common that the remote estimator is equipped with a residue-based detector to detect potential attacks. We propose a class of deception attack and analyze its feasibility. We show that the proposed attack enables the attacker to inject false data into the remote estimator without being detected. We derive a criterion to judge the optimality of performance of this type of attack in the sense of maximizing the estimation error covariance. Furthermore, we find that a simple linear attack strategy, which flips the sign of intercepted signal, satisfies the optimality criterion. We present numerical examples to illustrate our theoretical results.**

Index Terms— **Cyber-physical systems (CPS), deception attack, remote state estimation.**

I. INTRODUCTION

The combination of communication and control system brings about tremendous change in control system design. The emergence of cyber-phyisical systems, which introduces the communication layer to classical feedback control systems, enables possible design and implementation of large and complex systems in a remote operation scenario [1]. A wide range of areas can benefit from the development of the cyber-physical systems (CPSs), including energy, manufacturing, mining and transportation.

The introduction of communication channel, however, also brings more challenges to the design of control systems because the constraints of network effect, such as time delay, packet drop and disorder, would cause new challenges and should be considered. Since signals are transmitted through unprotected communication channels, the vulnerability of the CPSs poses another fundamental challenge to control engineers. Any successful attacks on the cyber-physical systems may cause enormous loss, including leakage of classified information, suspension of industrial process, infrastructure destruction and even casualty in human lives. Recent accidents (e.g. StuxNet malware in [2], [3]) related to security flaws in the cyber-physical systems gradually attracted researchers' interest [4]. It is important to understand what potential attacks are, what effect they might cause, and what the most catastrophic effect would be.

Based on knowledge and resources available to the malicious attackers, they are able to eavesdrop, intercept and falsify signal transmitted in the communication network and thus cause catastrophic effect on the CPSs. In [5], the authors classified several attack patterns in cyber-physical systems from the perspective of attacker's knowledge of the system, disclosure resources and disruption resources. Dennial-of-Service (DoS) attack is one of the basic class of attacks, in which attacker blocks the communication channel for data transmission. Zhang et al. [6] studied the worst-case attack with limited energy budget in the system. Li et al. [7] studied the interactive decisions of sending data by a sensor and jamming channel by an attacker, and adopted a Nash Q-learning algorithm to derive the optimal policy for both parties.

Compared with DoS attacks, depcetion attacks are subtler and thus are more difficult to be detected. Mo and Sinopoli [8] investigated the feasiblity of replay attacks against control systems equipped with a false-data detector. They also designed a watermark based detection scheme to detect such an attack in [9]. Bai et al. [10] analyzed the stealthy attack on Kalman filter. They revealed the trade-off between the degradation of estimation quality due to attacks, versus the stealthiness or detectablity. The authors derived bounds of the degradation under the constraint of avoiding detection by an ergodicity based detector. The results were derived with the assumption that the the system parameters were known for the attacker. In [11], we obtained an optimal linear deception attack against remote state estimation.

Note that although an optimal attack policy was derived in [11], the set of the attack policies was restricted to linear form. The linearity of attack policy leaves two questions to be answered. Firstly, the framework of linear attack analysis cannot cover a general form of possible attacks. Therefore, the effect on remote estimator caused by a general form of deception attacks remains unknown. The second is that whether there are any attack policies in a more general form can cause worse effect on the remote estimator. Our goal is to derive the worst-case deception attack, which can maximize the estimation error covariance and evade detection by the χ^2

∗: Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: swuak@ust.hk, zguoae@ust.hk, eesling@ust.hk).

†: State Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, Beijing, 100081, China (e-mail: dawei.shi@outlook.com).

‡: ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: kallej@kth.se).

The work by S. Wu, Z. Guo and L. Shi is supported by an HKUST KTH Partnership FP804. The work by D. Shi is supported by Natural Science Foundation of China (61503027). The work by K. H. Johansson is supported by the Knut and Alice Wallenberg Foundation and the Swedish Research Council.

978-1-5090-5994-2/$31.00 ©2017 AACC

detector. In this work, we extend the linear attack policies to a general form and obtain a criterion for judging the optimality of this type of attacks, which indicates that the optimal attack policy is nonunique.

The main contribution of this work are summarized as follows. Firstly, we extend the linear attack policy of deception attack aimed at innovation sequence to a more general form. We analyze the feasibility of this type of attack and show that attacker can inject false data into the remote estimator while remaining undetected by χ^2 based detector. Secondly, we show how error covariance propagates in the presence of the general deception attack and compare the results with the evolution of Kalman filter iteration. Lastly, a criterion for judging the optimality of this type of attack is derived, and we find that flipping the sign of innovation sequence is capable of meeting this criterion.

The remainder of this work is as follows. We introduce the architecture of the system under cyber-attack and its mathematical formulation in Section II. In Section III, how the cyber-attack is conducted is introduced. The system performance degradation due to the proposed attack along with optimal attack parameter is analyzed in Section IV. Section V presents numerical simulation of the analysis conducted.

Notations: $\mathbb{N}$ is the set of nonnegative integers. $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{R}^n$ is the n-dimensional Euclidean space. $\mathbb{S}_+^n$ is the set of $n \times n$ positive semi-definite matrices. We write $X \geq 0$ for $X \in \mathbb{S}_+^n$ and $X - Y \geq 0$ for $X - Y \in \mathbb{S}_+^n$. X' denotes the transposition of the matrix X. $\mathbb{E}$ stands for the expectation of a random variable.

II. SYSTEM SETUP AND PROBLEM FORMULATION

In this work, the architecture of remote state estimation system under cyber-attacks is presented in Fig. 1. There are six major elements, namely, the process to be estimated, the smart sensor, the communication network, the remote estimator, a malicious attacker and an intrusion detector. The smart sensor is equipped with computation capacity, and it is able to run a local Kalman filter based on measurement y_k from the process and sends innovation sequence, which will be introduced later, through a communication network to the remote estimator. By hijacking the communication network, the attacker is able to intercept the innovation sequence and falsify it so that estimation quality in the remote estimator is degraded. When the innovation sequence reaches the estimator, it is also received by a false data detector. The detector can predict and inform the remote estimator whether the received data has been modified based on testing the statistical characteristics of the sequence.

A. Process Description

Consider the following stochastic discrete-time linear time-invariant process

$$x_{k+1} = Ax_k + w_k, \tag{1}$$

$$y_k = Cx_k + v_k, , \tag{2}$$

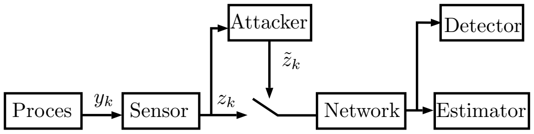

Fig. 1. System architecture. A malicious attacker is able to acquire and falsify data sent by a smart sensor and thus degenerate the performance of remote estimation.

where $k \in \mathbb{N}$ is time index, $x_k \in \mathbb{R}^n$ is the state vector, $y_k \in \mathbb{R}^m$ is the measurement vector of sensor(s), $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are process noise and measurement noise, respectively. The two noise processes are mutually uncorrelated Gaussian stochastic process with covariance $Q \geq 0$ and $R > 0$, respectively. Moreover, the initial condition of system state x_0 is a zero-mean Gaussian distributed random vector uncorrelated with w_k and v_k, and its covariance $\Sigma \geq 0$. We assume (A, C) is detectable and $(A, \sqrt{Q})$ is stabilizable.

B. Kalman Filter in Remote Estimation

The sensor module in the remote estimation scheme is deployed to measure the process described above and is able to obtain the y_k in (2). In standard setting, sensors directly send the measurement to an estimator which renders Minimum Mean Squared Error (MMSE) estimate of the state of the process through the following iteration of Kalman filter

$$\hat{x}_k^- = A\hat{x}_{k-1}, \tag{3}$$

$$P_k^- = AP_{k-1}A' + Q, \tag{4}$$

$$K_k = P_k^- C' \Big[C P_k^- C' + R \Big]^{-1}, \tag{5}$$

$$\hat{x}_k = A\hat{x}_{k-1} + K_k \Big(y_k - CA\hat{x}_{k-1} \Big), \tag{6}$$

$$P_k = (I - K_k C) P_k^-, \tag{7}$$

where $\hat{x}_k^-$ and $\hat{x}_k$ are the *a priori* and the *a posteriori* MMSE estimates of the state in the Kalman filter, and P_k^- and P_k are the corresponding estimation error covariances. The iteration starts from $\hat{x}_0 = 0$ and $P_0 = \Sigma_0 > 0$. It is known that the initial condition of iteration does not affect the performance the Kalman filter in the long term in the sense that P_k^- and K_k converge to a unique positive definite matrix for sure if (A, C) is detectable and $(A, \sqrt{Q})$ is stabilizable. The corresponding steady state values are

$$\overline{P} = \lim_{k \to \infty} P_k^-, \tag{8}$$

$$K = \overline{P} C' \left(C \overline{P} C' + R \right)^{-1}. \tag{9}$$

This implies that Kalman filter would become a fixed gain estimator after sufficiently long time, and the corresponding estimation error covariance is $\overline{P}$. This paper, however, studies a modified version.

In our framework, smart sensors are considered instead of standard sensors. Compared with standard sensors, smart sensors are equipped with extra computation unit which can be used to improve the estimation quality. Instead of directly sending measurements to the remote estimator, smart sensors are able to send the innovation sequence which is defined as follows

$$z_k = y_k - CA\hat{x}_{k-1}.$$

The innovation z_k is computed through the iteration of a local Kalman filter running in the smart sensor. Two factors contribute to sending innovation z_k instead of the measurement y_k. The first reason is that the computation load in the remote estimator can be reduced, which will be shown in the later section. The second reason is that z_k's are independent and identically distributed, which facilitates the false data detection.

After the remote estimator receives z_k, it conducts a Kalman filter algorithm similar to (3) to compute the remote estimate of the states. The only difference is that $y_k - CA\hat{x}_{k-1}$ is substituted with z_k.

C. χ^2-Based False Data Detector

When the innovation sequence enters the remote estimator, it is also received by a false data detector. The detector collects data in a monitor windows of appropriate size and computes the statistic of collected data. By comparing the computed statistic with a threshold, the detector is able to judge the authenticity of the received data. If the statistic is greater than the threshold, an alarm will be triggered.

In Kalman filter, the innovation sequnce consists of independent random variables with identical distribution $\mathcal{N}(0, C\overline{P}C' + R)$. The false data detector can utilize the statistical property of the sequence to detect abnormity in the coming sequence. Since the sequence follows a normal distribution with known mean and variance, χ^2 test is an appropriate option. Given a moving windows, we take the summation over the normalized past sequence. By comparing it with a threshold, correctness of the current data in the windows can be decided. The detection algorithm for time k in a moving windows of size J is in the following form:

$$\sum_{i=k-J+1}^{k} z_i' \left(C\overline{P}C' + R\right)^{-1} z_i \underset{H_1}{\overset{H_0}{\lessgtr}} \delta,$$

where δ is a threshold, H_0 is the null hypotheses which suggests that the received sequence in current monitor windows follows the expected probability distribution, and H_1 is the alternative hypotheses which suggests that the received sequence no longer follows the expected probability distribution. This detector is easy to implement and effective in some scenarios.

D. Problem of Interest

Based on the setting of the process, sensor, remote estimator, attacker and detector described above, the problems we are interested in are as follows:

1) Are there any feasible attacking strategies such that the attack will not trigger the false data detector?
2) What will be the corresponding estimation performance?
3) Are there any optimal attack strategies which cause maximal estimation error in each time step?

These problems will be elaborated in detail and solved in the following sections.

III. ATTACK CHARACTERIZATION AND FEASIBILITY ANALYSIS

In this section, we propose a type of deception attacks and analyze how an attacker can launch deception attack on the remote estimation system. We first discuss what resources are available to the attacker. Next, the stealthiness of an attack can be achieved so that the false data detector can be bypassed. In other words, feasibility of our proposed attack is addressed.

A. False Data Deception Attack

To run remote estimation based on Kalman filter, the smart sensor sends innovation sequence $z_k = y_k - C\hat{x}_k^-$ to the remote estimator. However, the innovation signal is obtained and modified by a malicious attacker. The attacker is intended to deceit the remote estimator by sending incorrect innovation information. The strategy of an attacker therefore can be described as launching the attack by falsifying z_k without detection by the false data detector described above. At each time step k, a generalized attack strategy is

$$\tilde{z}_k = f_k(z_k), \tag{10}$$

where z_k is the innovation intercepted by the attacker, $\tilde{z}_k$ is the actual innovation provided by the attacker to the remote estimator, and $f_k(\cdot)$ is an arbitrary function. In order to avoid being detected by a χ^2 detector, the falsified innovation $\tilde{z}_k$ should also follow a normal distribution as z_k does, i.e. $\tilde{z}_k \sim \mathcal{N}(0, C\overline{P}C' + R)$. Note that the random variable of innovation sequence is independent and so is their effect on the estimation error covariance. Therefore, it is reasonable to assume the attack is in the form of (10), which only modifies the present innovation.

B. Feasibility of Proposed Attack

According to the analysis presented previously, the only information for the attacker to falsify the innovation data without being detected by χ^2 detector is the variance of z_k, which is $C\overline{P}C' + R$. If the attacker can acquire the parameters of system in (1) and in (2), namely A, C, Q and R, the variance can be calculated by the attacker. By solving the following discrete time infinite horizon Riccati equation

$$\overline{P} = A\overline{P}A' + Q - A\overline{P}C'(C\overline{P}C' + R)C\overline{P}A'$$

to obtain $\overline{P}$, the attacker can obtain the desired variance information about z_k.

The above assumption about the resources available to the attacker is too strong. Simultaneously acquiring knowledge of A, C, Q and R is not an easy task. In the following discussion, we will explain that the attacker can still obtain the variance information after sufficiently long time by adopting interval estimation technique.

Interval estimation is common in statistics to estimate the statistic of a random variable by sampling the random variable for several times. This estimate gives the probability of the random variable being in a region.

Suppose a random variable X following normal distribution with unknown variance σ^2 has been sampled for n times, and let the samples be $X_i,\ i = 1, \cdots, n$. We define the following compound random variables to facilitate the expression followed.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

where $\bar{X}$ is the sample mean, and S^2 is the sample variance. We would like to infer the variance of X from the sampled data. According to knowledge in statistics, we have

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

where $\chi^2(n-1)$ stands for a χ^2 distribution with $n-1$ degrees of freedom. The probability of the variance σ^2 falling into an interval with confidence level $1-\beta$ is given by

$$P\left[\frac{(n-1)S^2}{\chi^2_{\frac{\beta}{2}}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\beta}{2}}(n-1)}\right] = 1-\beta \tag{11}$$

where $\chi^2_\alpha(n)$ is the fractile of χ^2 distribution and is defined as

$$P\left[\chi^2(n) > \chi^2_\alpha(n)\right] = \alpha$$

As we can see in (11), when the size of sampled data grows, the region will shrink if the confidence level $1-\beta$ is fixed. That is to say, if the attacker observes z_k for sufficiently long time, he can pinpoint the variance of z_k in a tight neighborhood with high confidence level.

In the above discussion, we analyzed the feasibility of the proposed deception attack. Furthermore, we show that the attack can be launched even if the system information, i.e. A, C, Q and R is absent. The next question is that what the optimal attack in the set of all feasible attacks is. The optimality discussed here originates from the perspective of the attacker. Namely, the optimal attack should maximize the estimation error covariance of the remote estimator. The next section gives solution to the optimality problem.

IV. PERFORMANCE ANALYSIS

In this section, we consider the remote estimator under the proposed attack (10). We first derive the estimation error covariance propagation equation in the presence of deception attack. A criterion of whether a strategy can maximize the estimation error is obtained. This criterion only reflects the covariance evolution under an attack. Motivated by this observation, we finally obtain a simple attack strategy, which simply flips the sign of intercepted innovation z_k, and is capable of providing optimal attack. It is proven that this simple strategy is powerful enough to meet the goal of maximizing destruction on the remote estimation.

A. Estimation Error Propagation

The previous section discussed the feasibility of our proposed attack. The remote estimator would receive maliciously modified innovation $\tilde{z}_k$ and update its estimate of the system process based on the following equations

$$\tilde{x}_k^- = A\tilde{x}_{k-1}, \tag{12}$$
$$\tilde{x}_k = \tilde{x}_k^- + K\tilde{z}_k, \tag{13}$$

where K is the steady state Kalman filter gain defined in (9).

Based on the analysis in previous section, it is possible for an attacker to falsify innovation used to update the remote estimation and bypass a χ^2 detector. Therefore, although the received innovation may not be the original one, the remote estimator can still update its estimate according to (12) and (13). As a result, the estimation gradually deviate from the real system states. The following lemma illustrates how the estimation error covariance evolve under the deception attack of the form (10).

Lemma 1 *Given the attack $\tilde{z}_k$ on the remote estimation system described in* (10)*, the recursion of the estimation error covariance at the remote estimator is as follows*

$$\begin{aligned}\tilde{P}_k^- &= A\tilde{P}_{k-1}^-A' + Q + A\overline{P}C'\Sigma C\overline{P}A' \\ &\quad - A\Psi_k\Sigma C\overline{P}A' - A\overline{P}C'\Sigma{\Psi_k}'A',\end{aligned} \tag{14}$$

where $\Sigma = (C\overline{P}C' + R)^{-1}$, $\Psi_k = \mathbb{E}\left[(x_k - \tilde{x}_k^-)\tilde{z}_k'\right]$.

Proof: Let us assume that the concerned Kalman filter has entered in steady state before the attack begins.

Denote the *a posterior* estimation error and the *a prior* estimation error as $\tilde{e}_k = x_k - \tilde{x}_k$ and $\tilde{e}_k^- = x_k - \tilde{x}_k^-$, then

$$\begin{aligned}\tilde{e}_k &= x_k - \tilde{x}_k \\ &= x_k - \tilde{x}_k^- - K\tilde{z}_k \\ &= \tilde{e}_k^- - K\tilde{z}_k,\end{aligned}$$

and

$$\begin{aligned}\tilde{e}_k^- &= x_k - \tilde{x}_k^- \\ &= Ax_{k-1} + w_{k-1} - A\tilde{x}_{k-1} \\ &= A\tilde{e}_{k-1} + w_{k-1},\end{aligned}$$

where $K = \overline{P}C'(C\hat{P}C' + R)^{-1}$ is the steady state Kalman filter gain.

The *a posterior* estimation error covariance and the *a prior* estimation error covariance are given as

$$\tilde{P}_k = \tilde{P}_k^- + K(C\overline{P}C' + R)K' - K\Psi_k' - \Psi_kK', \tag{15}$$
$$\tilde{P}_k^- = A\tilde{P}_{k-1}A' + Q, \tag{16}$$

where $\Psi_k = \mathbb{E}\left[(x_k - \tilde{x}_k^-)\tilde{z}_k'\right]$.

Plug (15) in (16), we obtain

$$\begin{aligned}\tilde{P}_k^- &= A\tilde{P}_{k-1}^-A' + Q + AK(C\tilde{P}C' + R)K'A' \\ &\quad - AK\Psi_k'A' - A\Psi_kK'A' \\ &= A\tilde{P}_{k-1}^-A' + Q + A\overline{P}C'\Sigma C\overline{P}A' \\ &\quad - A\Psi_k\Sigma C\overline{P}A' - A\overline{P}C'\Sigma{\Psi_k}'A',\end{aligned}$$

where $\Sigma = (C\overline{P}C' + R)^{-1}$. ■

The closed form of error propagation is derived in this lemma. In the perspective of an attacker, who is interested in degrading the performance of the estimation system as much as possible, the next question is how to maximize the estimation error in each time step. In next subsection, we will answer this question. Furthermore, the effect of such an attack is also be analyzed.

B. Optimal Attack

In this subsection, we first present a lemma in which a criterion is provided to test whether an attacker can maximize the effect of its attack. Based on obtained result, we find the gap between optimal estimation and worst estimation in presence of a deception attack.

Lemma 2 *Consider an attack $\tilde{z}_k$ on the system described in* (10)*, the attack can maximize the prior estimation error covariance $\tilde{P}_k^-$ in each time step if the following condition is satisfied*

$$\Psi_k = \mathbb{E}[\tilde{e}_k^- \tilde{z}_k] = -\overline{P}C'. \tag{17}$$

Proof: Note that when the Kalman filter is applied, the prior estimation error covariance is minimized [12]. According to the principle of orthogonality,

$$\mathbb{E}[(\tilde{e}_k^- - K\tilde{z}_k)\tilde{z}_k'] = 0. \tag{18}$$

By expanding (18), we obtain

$$\Psi_k = \mathbb{E}[e_k^- \tilde{z}_k'] = \mathbb{E}[K\tilde{z}_k\tilde{z}_k'] = K(C\overline{P}C' + R) = \overline{P}C'. \tag{19}$$

Consequently, when $\Psi_k = \overline{P}C'$, the *a prior* estimation error covariance $\tilde{P}_k^-$ in (14) is minimized, which is equivalent to the following function being minimized.

$$f(\Psi_k) = A\overline{P}C'\Sigma C\overline{P}A' - A\Psi_k\Sigma C\overline{P}A' - A\overline{P}C'\Sigma\Psi_k'A'. \tag{20}$$

Therefore, for any matrix Δ compatible in dimention, we have

$$f(\overline{P}C' + \Delta) - f(\overline{P}C') = A\Delta\Sigma C\overline{P}A' + A\overline{P}C'\Sigma\Delta'A' \geq 0.$$

Following a similar argument, we have

$$\begin{aligned} &f(-\overline{P}C') - f(-\overline{P}C' - \Delta) \\ &= A\Delta\Sigma C\overline{P}A' + A\overline{P}C'\Sigma\Delta'A' \geq 0, \end{aligned}$$

which suggests that $f(\Psi_k)$ is maximized when $\Psi_k = -\overline{P}C'$. This is equivalent to (14) being maximized. Thereupon, the optimal attack should satisfy (17). ■

By substituting Ψ_k in (14) with (17), the estimation error covariance propagates under optimal deception attack is derived as follows

$$\tilde{P}_k^- = A\tilde{P}_{k-1}^- A' + Q + 3A\overline{P}C'(C\overline{P}C' + R)^{-1}C\overline{P}A'.$$

Meanwhile, recall that if $\tilde{z}_k = z_k$, i.e., the optimal estimation is used for update in the remote estimator, we will have the following error covariance propagation instead.

$$\tilde{P}_k^- = A\tilde{P}_{k-1}^- A' + Q - A\overline{P}C'(C\overline{P}C' + R)^{-1}C\overline{P}A'.$$

As the two iterations differ by $4A\overline{P}C'(C\overline{P}C' + R)^{-1}C\overline{P}A'$, if the system is stable, the gap between the two iterations will converge to a constant value. If the system is unstable, the gap will diverge to infinity.

With Lemma 2, the set of the optimal attack policies could be characterized. However, we would not derive an optimal attack policy directly from this criterion. Instead, we demonstrate that a linear attack strategy is powerful enough to meet (17) and thus renders the optimal attack.

C. Optimal Attack in a Linear Form

Note that (17) in Lemma 2 is concerned with the covariance of two random variables. The covariance only reflects the linear properties in random variables. This lead us to wonder whether a linear attack is capable of providing such an optimal attack which maximizes the estimation error covariance in each time step. Consider the set linear attack strategies of the following form

$$\tilde{z}_k = T_k z_k.$$

Then we compute $\mathbb{E}\left[(x_k - \tilde{x}_k^-)\tilde{z}_k'\right]$ by comparing it with equation (25) in [11]. The expression is obtained as follows

$$\mathbb{E}\left[(x_k - \tilde{x}_k^-)\tilde{z}_k'\right] = \overline{P}C'T_k'.$$

Theorem 1 *The linear attack policy $\tilde{z}_k = -z_k$ can maximize the estimation error covariance in each time step.*

Proof: Under the attack $\tilde{z}_k = -z_k$, we have

$$T_k = -I.$$

Therefore,

$$\Psi_k = \mathbb{E}\left[(x_k - \tilde{x}_k^-)\tilde{z}_k'\right] = -\overline{P}C',$$

which meets the sufficient condition of the optimal attack. ■

V. Numerical Examples

To validate the theoretical results obtained in the previous sections, several numerical simulations are presented in this section. We compare the impact of the optimal deception attack with the case when the communication network is blocked, i.e., the innovation is not available to the remote estimator. This scenario can also be taken as a DoS attack, in which the innovation is zero. As a consequence, the *a posterior* state estimate and its error covariance is equal to those of the prior estimate. The optimal attack is implemented by flipping the sign of the innovation. The process we consider is a stable first-order process whose parameters are $A = 0.8$, $C = 1$, $Q = 1$ and $R = 1$.

The simulation starts from the time $k = 0$ and ends at $k = 100$. During the time slot between 20 and 40, the communication is under the DoS attack. During the time slot between 60 and 80, the optimal deception attack is launched. Fig. 2 and 3 present how the remote state estimate and its corresponding error covariance evolve.

As it is shown in Fig. 2, during the time interval$[0, 20]$, the system is in normal operation and the remote state estimate perfectly follows the estimate of Kalman filter. From the time slot $k = 20$, the communication network is blocked and the remote estimate simply projects its estimate in each time step. After $k = 40$, the communication channel is recovered and the remote estimator gradually bring the state estimate to the optimal value. During $[60, 80]$, the deception attack is launched and the remote estimate moves to the opposite direction with respect to the estimate of the Kalman filter. After the attack ends, the state estimate gradually recovers as it does during $[40, 60]$.

In Fig. 3, we observe that the error covariance does not diverge but shifts to a value much greater than P_k, which is consistent with our analysis on the gap between Kalman filter and optimal deception attack. Since $\tilde{P}_k$ is bounded, $\tilde{x}_k$ is also bounded. It is obvious that the impact caused by optimal deception attack is greater than that of the DoS attack.

VI. Conclusion

We propose a novel class of deception attacks against remote estimation systems. This type of attack can degrade the performance of the Kalman filter in a remote estimator by falsifying the innovation z_k sent by sensors. This attack is able to bypass a χ^2-based false data detector by keeping certain statistical property unchanged. After analyzing the feasibility of this attack, we study the propagation of the error covariance of estimation when the innovation-based attack is launched in the system. A criterion for maximizing the error covariance is derived from the propogation. We further provide a simple linear attack which simply flips the sign of the innovation and is able to meet the criterion obtained. Future work on this subject might include on how to detect this type of attacks. Moreover, if the remote estimator is aware of the potential existence of such attacks, the decision of how to minimize its estimation error covariance is another interesting topic.

References

[1] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 138–162, 2007.

[2] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *Security & Privacy, IEEE*, vol. 9, no. 3, pp. 49–51, 2011.

[3] T. Rid, "Cyber war will not take place," *Journal of strategic studies*, vol. 35, no. 1, pp. 5–32, 2012.

[4] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[5] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st international conference on High Confidence Networked Systems*. ACM, 2012, pp. 55–64.

[6] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 3023–3028, 2015.

[7] Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "SINR-based dos attack on remote state estimation: A game-theoretic approach," *IEEE Transactions on Control of Network Systems*, 2016.

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

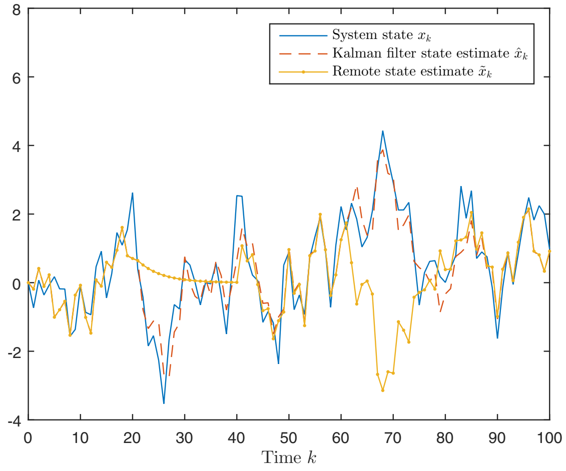

Fig. 2. Remote state estimate. Under DoS attack, which happens during $[20, 40]$, the remote estimate simply projects its estimate. During $[60, 80]$, the remote estimate is under the optimal deception attack and moves to the optimal direction of the true Kalman filter.

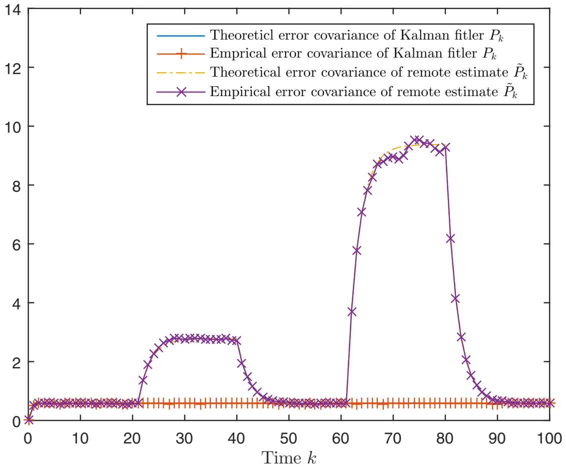

Fig. 3. Remote estimation error covariance. In $[60, 80]$, optimal deception attack is performed and the estimation quality degrades to a larger extent than in $[20, 40]$, during which the communication channel is blocked.

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

[10] C.-Z. Bai and V. Gupta, "On Kalman filtering in the presence of a compromised sensor: Fundamental performance bounds," in *2014 American Control Conference*. IEEE, 2014, pp. 3029–3034.

[11] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Transactions on Control of Network Systems*, 2016.

[12] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.