2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Worst-case Analysis of Innovation-based Linear Attack on Remote State Estimation with Resource Constraint

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

Abstract— **In this paper, a security problem in remote estimation scenario is studied. We consider a multi-sensor system where each sensor transmits its local innovation to a remote estimator through a wireless communication network. A centralized residue-based detection criterion is adopted to monitor system anomalies. We propose a linear attack strategy and present the corresponding feasibility constraints to guarantee stealthiness. For a resource-limited attacker, who is able to listen to all the channels while only launches an attack on one sensor at each time instant, we investigate which sensor should be attacked and what strategy should be used such that the remote estimation error covariance is maximized. A closed-form expression of the optimal linear attack strategy is obtained. Simulation examples are provided to illustrate the theoretical results.**

Index Terms— **Cyber-Physical System Security, Remote State Estimation, Integrity Attack**

I. Introduction

Cyber-Physical Systems (CPS) are systems that integrate sensing, communication, control, computation and physical processes [1]. Due to the complex integration of various technologies and components with the cyber information layer, CPS offer a variety of attack surfaces to malicious agents [2]. Attacks may result in severe consequences on national economy, social security or even loss of human lives [3]. Therefore, security is of fundamental importance to ensure the safe operation of CPS.

With the increasing adoption of CPS to safety-critical applications, attack strategies and defense mechanisms have been recently investigated. Depending on their resources and capabilities, the malicious attackers aim to deteriorate the system functionality, while remain undetected for as long as possible [4]. Denial-of-service (DoS) and deception attacks, two major categories of cyber attacks in CPS, were studied in [5]. DoS attacks which attempt to block the communication channels were investigated for resource-constrained attackers [6]–[8]. Besides, Li et al. [9] proposed a game-theoretic framework to study the decision-making process with energy constraints of the sensor and the attacker.

The deception attacks, which aim at compromising the data integrity, have recently received considerable attention. Replay attacks, which are able to access, record, and replay the sensor data, were studied in [10]. False-data injection attacks were first considered for resource-limited attackers against remote state estimation in power grids [11]. The reachable estimation error covariance under such attacks was investigated in [12], [13]. A covert data attack, which misleads the control center to remove useful measurements, was proposed and analyzed for dynamic systems in [14]. Other formulations of cyber attacks on secure estimation problems were investigated in [15], [16].

A type of linear attack that modifies the transmitted innovations without being noticed by the χ^2 detector, was first proposed in our previous work [17]. The evolution of the estimation error covariance and the closed-form optimal attack strategy were obtained for the single sensor system. In this work, we consider a networked system with multiple sensors. If a centralized detection criterion is adopted by the remote estimator, the previous single-senor attack strategy may fail to bypass the false-data detector. Hence, we propose a new linear attack strategy for the multi-sensor scenario and derive the corresponding stealthiness constraints. Moreover, the optimal linear attack which maximizes the remote estimation error covariance is investigated for the resource-limited attacker.

The main contributions of this paper are threefold. First, we extend the linear integrity attack strategy to the multi-sensor framework, which is more general and involved than the single sensor scenario. Second, we propose a centralized false-data detection criterion and analyze two necessary conditions for the malicious agent to guarantee the attack stealthiness. Specifically, one is to modify the feedback information, and the other is to maintain the same statistical characteristics. Last, for a resource-constrained attacker who is only able to attack one sensor at each time instant, we derive an explicit expression of the optimal linear attack strategy, i.e., which sensor should be attacked and what strategy should be adopted to achieve the largest degradation of system estimation performance.

The remainder of the paper is organized as follows. Section II introduces the architecture of the multi-sensor system. Section III presents the linear attack strategy and corresponding feasibility constraints. Section IV derives the optimal linear attack strategy for resource-constrained attackers. Numerical examples are provided in Section V. Some concluding remarks are given in the end.

*: 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).

†: 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 Z. Guo and L. Shi is supported by an RGC General Research Fund 16210015. 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.

978-1-5090-1837-6/16/$31.00 ©2016 IEEE

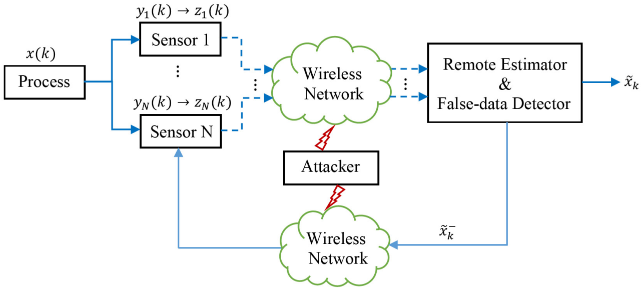

Fig. 1. System block diagram: The attacker tries to intercept and modify sensor data to degrade the remote estimation performance despite the false data detector.

II. Problem Formulation

A. System Model

Consider a networked system consisting of a group of $N > 1$ wireless sensors and one remote estimator as depicted in Fig. 1. Each sensor $i \in \{1, 2, \ldots, N\}$ measures the output of the same first-order linear time-invariant (LTI) process:

$$x(k+1) = ax(k) + w(k), \tag{1}$$

$$y_i(k) = c_i x(k) + v_i(k), \tag{2}$$

where $a, c \in \mathbb{R}$, $c \neq 0$, $k \in \mathbb{N}$ is the time index, $x(k) \in \mathbb{R}$ is the process state vector, $y_i(k) \in \mathbb{R}$ is the measurement vector obtained by sensor i, $w(k) \in \mathbb{R}$ and $v_i(k) \in \mathbb{R}$ are zero-mean i.i.d. Gaussian noises with $\mathbb{E}[w(k)(w(l))'] = \delta_{kl}q$ $(q \geq 0)$, $\mathbb{E}[v_i(k)(v_j(l))'] = \delta_{ij}\delta_{kl}r_i$ $(r_i > 0)$, $\mathbb{E}[w(k)(v_i(l))'] = 0$, $\forall k, l \in \mathbb{N}$, $i = 1, 2, \ldots, N$. The initial state $x(0)$ is zero-mean Gaussian with covariance $\pi_0 \geq 0$. The pair $(a, [c_1, c_2, \ldots, c_N]')$ is detectable and (a, q) is stabilizable.

B. Remote Estimator

Each sensor sends its data to a remote estimator through a wireless communication network at each time step. To estimate the system state, a centralized Kalman filter is adopted by the remote estimator to fuse the received data:

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

$$P_k^- = a^2 P_{k-1} + q, \tag{4}$$

$$K_k = P_k^- C'(CP_k^- C' + R)^{-1}, \tag{5}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(Y_k - C\hat{x}_k^-), \tag{6}$$

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

where the Kalman gain is

$$K_k \triangleq \begin{bmatrix} k_1(k) & k_2(k) & \ldots & k_N(k) \end{bmatrix},$$

the measurement vector is

$$Y_k \triangleq \begin{bmatrix} y_1(k) & y_2(k) & \ldots & y_N(k) \end{bmatrix}',$$

the measuring matrix is

$$C \triangleq \begin{bmatrix} c_1 & c_2 & \ldots & c_N \end{bmatrix}', \tag{8}$$

the noise covariance matrix

$$R \triangleq \operatorname{diag}\begin{bmatrix} r_1 & r_2 & \ldots & r_N \end{bmatrix}, \tag{9}$$

and $\hat{x}_k^-$, $\hat{x}_k$ are the *a priori* and the *a posteriori* minimum mean squared error (MMSE) estimates of $x(k)$ at the remote estimator and P_k^-, P_k the corresponding error covariances. The recursion starts from $\hat{x}_0^- = 0$ and $P_0^- = \pi_0 \geq 0$.

For notational brevity, we define the Lyapunov and Riccati operators $h, \tilde{g} : \mathbb{R}_+ \to \mathbb{R}_+$ as:

$$h(X) \triangleq a^2 X + q,$$
$$\tilde{g}(X) \triangleq X - XC'(CXC' + R)^{-1}CX.$$

It is well known that the Kalman filter converge from any initial condition exponentially fast [18]. Hence, we assume that the system has already entered the steady state and simplify our subsequent discussion by setting

$$\widehat{P} \triangleq \lim_{k \to +\infty} P_k, \quad \overline{P} \triangleq \lim_{k \to +\infty} P_k^-,$$
$$K \triangleq \overline{P}C'(C\overline{P}C' + R)^{-1}, \tag{10}$$

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.

Similar to the single sensor case [17], to reduce the communication bandwidth and for the security purpose, each sensor will also first locally process the raw measurement data and send its local innovation to the remote estimator. However, in the multi-sensor scenario, each single sensor cannot compute the *a priori* estimate $\hat{x}_k^-$ at the remote side based on its own measurements. Therefore, one efficient way is that the remote estimator broadcasts its $\hat{x}_k^-$ at each time step to reduce the communication costs of the sensors, see Fig. 1. Under such a protocol, the innovation transmitted by sensor i, $i \in \{1, 2, \ldots, N\}$ is defined as

$$z_i(k) = y_i(k) - c_i\hat{x}_k^-, \tag{11}$$

which has the following properties:

Lemma 1 (See [18])

1) $z_i(k)$ has Gaussian distribution $\mathcal{N}(0, p_i)$, where $p_i = c_i^2\overline{P} + r_i$.

2) $z_i(k)$ and $z_i(l)$ are independent $\forall k \neq l$.

Remark 1 *Due to the power asymmetry in many sate estimation applications (the remote side is more powerful than the local sensor), the estimator is able to send feedback information to the sensors. A practical example can based on the IEEE 802.15.4/ZigBee protocol [19].*

Remark 2 *The sensor sends the innovation $z_i(k)$, rather than the measurement $y_i(k)$, due to communication efficiency and detection convenience. The steady-state Gaussian distribution of $z_i(k)$ enables an direct detection of abnormal data modified by malicious attackers.*

C. False-Data Detector

False-data detectors are widely used in CPS to monitor system behavior and detect cyber attacks by checking the statistical characteristics of received data [20]. In this paper, a centralized residue-based detection algorithm is used at the remote estimator.

Specifically, the remote estimator diagnoses the existence of malicious attacks by checking the sum of the normalized innovation sequence for all sensors at the beginning of each time step, i.e., at time k, the remote estimator first collects all the innovations to form a column vector

$$Z_k \triangleq \begin{bmatrix} z_1(k) & z_2(k) & \dots & z_N(k) \end{bmatrix}',$$

which is further checked by the χ^2 false-data detector according to the following criterion:

$$G_k = \sum_{i=k-J+1}^{k} Z_i' P^{-1} Z_i \underset{H_1}{\overset{H_0}{\lessgtr}} \delta, \tag{12}$$

where J is the detection window size, δ the threshold,

$$P = \begin{bmatrix} p_1 & c_1c_2\overline{P} & c_1c_3\overline{P} & \dots & c_1c_N\overline{P} \\ c_1c_2\overline{P} & p_2 & c_2c_3\overline{P} & \dots & c_2c_N\overline{P} \\ c_1c_3\overline{P} & c_2c_3\overline{P} & p_3 & \dots & c_3c_N\overline{P} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1c_N\overline{P} & c_2c_N\overline{P} & c_3c_N\overline{P} & \dots & p_N \end{bmatrix} \tag{13}$$

denotes the covariance matrix of Z_k, and $p_i = c_i^2\overline{P} + r_i$. The null hypothesis H_0 means that the system is operating normally, otherwise the system is under attack. Note that the false alarm rate can be easily calculated since G_k is χ^2 distributed with NJ degrees of freedom. If G_k exceeds the threshold, the detector will trigger an alarm.

D. Problem of Interest

Based on the aforementioned system, we are interested in the following questions:

1) Does the linear attack strategy for the single-sensor case apply to the multi-sensor setup?
2) Under which feasibility constraints do the attacker remain undetectable to the proposed false-data detector?
3) What is the optimal resource-limited attack strategy that yields the largest remote estimation error covariance?

The detailed mathematical formulations and solutions to these problems will be introduced in the following sections.

III. Linear Attack in the Multi-sensor Scenario

In this section, we consider the existence of a malicious agent who intentionally launches cyber attacks to degrade the system performance. We first recall the innovation-based linear attack in a single sensor scenario, based on which the linear attack strategy for the multi-sensor system is proposed. Then, we analyze the feasibility constraints needed for such an attack from being detected by the proposed false-data detector.

It is assumed that the attacker has capability to intercept and modify the innovations transmitted from the sensor to the remote estimator. According to [17], the linear attack strategy for a single-sensor system is defined as

$$\tilde{\mathbf{z}}_k = \mathbf{T}_k \mathbf{z}_k + \mathbf{b}_k, \tag{14}$$

where $\tilde{\mathbf{z}}_k \in \mathbb{R}^m$ is the intercepted innovation, $\mathbf{z}_k \in \mathbb{R}^m$ the corrupted innovation, $\mathbf{T}_k \in \mathbb{R}^{m\times m}$ an arbitrary matrix, and $\mathbf{b}_k \in \mathbb{R}^m$ a zero-mean i.i.d. Gaussian random variable.

For the multi-sensor system (1)–(2) considered in this paper, the single-sensor attack strategy corresponds to $\tilde{z}_i(k) = t_i(k)z_i(k) + b_i(k)$. However, such an attack cannot successfully bypass the false-data detector (12) since the feedback information from the remote estimator has already deviated from the true *priori* estimate $\hat{x}_k^-$ when the system is under attack and the resulting covariance of innovation Z_k is no longer the same with P. Hence, to maintain the stealthiness of the attack, the attacker has to be capable of modifying the feedback information to $\hat{x}_k^-$ at each time step as well. Moreover, for the scenario where the attacker can only compromise a subset of the existing sensors, the attack still cannot avoid being detected since the off-diagonal elements in the covariance matrix of the corrupted innovation are not the same as the original ones. In this case, to guarantee attack stealthiness, we propose to change the attack policy to

$$\tilde{z}_i(k) = \sum_{j=1}^{N} t_{ij}(k) z_j(k) + b_i(k), \tag{15}$$

which thus depends on the information of all sensors. We represent (15) in a matrix expression as

$$\tilde{Z}_k \triangleq T_k Z_k + B_k, \tag{16}$$

where $Z_k \in \mathbb{R}^N$ and $\tilde{Z}_k \in \mathbb{R}^N$ stand for the currently intercepted innovation and the innovation modified by the attacker, respectively. $T_k \in \mathbb{R}^{N\times N}$ is the attack matrix. $B_k \sim \mathcal{N}(0, L)$ is an i.i.d. Gaussian random variable independent of Z_k with $L = \mathrm{diag}(l_1, l_2, \dots, l_N)$. Omitting the time index for simplicity, (16) is equivalent to

$$\begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \\ \vdots \\ \tilde{z}_N \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1N} \\ t_{21} & t_{22} & \dots & t_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ t_{N1} & t_{N2} & \dots & t_{NN} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}.$$

It is also worth noticing that $\tilde{Z}_k$ is Gaussian distributed with zero mean and covariance $T_k P T_k' + L$ since $Z_k \sim \mathcal{N}(0, P)$. To bypass the false-data detector, the modified innovation $\tilde{Z}_k$ needs to keep the same distribution as the original innovation Z_k, i.e.,

$$T_k P T_k' + L = P. \tag{17}$$

Based on the proposed attack strategy for the multi-sensor scenario and corresponding feasibility constraints, the problem we are interested in is to find the optimal linear attack under attacker resource constraints. Optimality is quantified in terms of the estimation performance, where we define $\tilde{P}_k^-$ and $\tilde{P}_k$ as the *a priori* and the *a posteriori* MMSE estimation error covariance of the remote estimator under attack.

IV. Optimal Attack Strategy under Resource Constraint

In practical applications, the malicious attacker may not be able to intercept and modify the transmitted data packet of each sensor all the time due to resource constraints. We thus study such a scenario in this section. As a first and important step, we assume that the attacker can only attack one of N sensors at each time instant. An explicit expression of the optimal linear attack strategy is obtained for this case, which yields the largest estimation error covariance.

A. Optimal Linear Attack in Single-sensor Scenario

Before we derive the optimal attack policy for multi-sensor system, we recall the results of the single-sensor case, which are summarized in the following proposition.

Proposition 1 (See [17]) *Consider a single-sensor system under linear attack $\tilde{\mathbf{z}}_k = \mathbf{T}_k\mathbf{z}_k + \mathbf{b}_k$.*

1) *The remote estimation error covariance follows*
$$\tilde{P}_k = A\tilde{P}_{k-1}A' + Q + \overline{P}C'(\Sigma - \mathbf{T}_k'\Sigma - \Sigma\mathbf{T}_k)C\overline{P},$$
where $\Sigma = (C\overline{P}C' + R)^{-1}$.
2) *The optimal linear attack, which maximizes the remote estimation error covariance, is achieved when $\mathbf{T}_k = -\mathbf{I}$ and $\mathbf{b}_k = \mathbf{0}$.*

B. Optimal Linear Attack in Multi-sensor Scenario with Resource Constraint

In this subsection, we consider the networked system with detection criterion (12) under linear attack (15). For a resource-constrained attacker, the optimal linear attack strategy when one sensor is under attack is first investigated. Then, the best choice for the attacker of which sensor should be attacked is obtained in the sense of maximum estimation error covariance. We now introduce the following lemmas which will be used in the subsequent derivation.

Lemma 2 *At steady state, the gain of the Kalman filter at the remote estimator is given as $K = [k_1, k_2, \ldots, k_N]$, where $k_i = \widehat{P}c_i/r_i$.*

Proof: According to (10), the steady-state value of the Kalman gain at the remote estimator satisfies $K(C\overline{P}C' + R) = \overline{P}C'$, which is equivalent to
$$KR = (I - KC)\overline{P}C' = \widehat{P}C'. \tag{18}$$
Substituting (8) and (9) into (18), one has $K = \widehat{P}C'R^{-1} = [\widehat{P}c_1/r_1, \widehat{P}c_2/r_2, \ldots, \widehat{P}c_N/r_N]$. ■

Lemma 3 (See [21]) *For matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times k}$, $C \in \mathbb{R}^{k\times k}$, $D \in \mathbb{R}^{k\times n}$, if A and C are invertible, then*
$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}.$$

Lemma 4 (See [21]) *For a block matrix $X = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times(m_1+m_2)}$, if D and its Schur complement $D_S = A - BD^{-1}C$ are invertible, then*
$$X^{-1} = \begin{bmatrix} D_S^{-1} & -D_S^{-1}BD^{-1} \\ -D^{-1}CD_S^{-1} & D^{-1} + D^{-1}CD_S^{-1}BD^{-1} \end{bmatrix}.$$

Theorem 1 *Consider the multi-sensor system* (1)–(2) *with centralized detection criterion* (12) *under linear attack* (15). *When sensor i is under attack, the optimal linear attack strategy, which yields the largest estimation error covariance, is given by $t_{ii} = -1, b_i = 0$.*

Proof: Without loss of generality, we investigate the optimal linear attack strategy when sensor $i = 1$ is under attack. We partition K, C, P, R, Z_k, B_k, T_k into block matrices as
$$K = \begin{bmatrix} k_1 & \mathbf{K} \end{bmatrix}, C = \begin{bmatrix} c_1 \\ \mathbf{C} \end{bmatrix}, Z_k = \begin{bmatrix} z_1 \\ \mathbf{Z} \end{bmatrix}, B_k = \begin{bmatrix} b_1 \\ \mathbf{B} \end{bmatrix},$$
$$P = \begin{bmatrix} p_1 & \mathbf{M} \\ \mathbf{M}' & \mathbf{P} \end{bmatrix}, T_k = \begin{bmatrix} t_{11} & \mathbf{N} \\ \mathbf{N}' & \mathbf{T} \end{bmatrix}, R = \begin{bmatrix} r_1 & \\ & \mathbf{R} \end{bmatrix},$$
where $k_1, c_1, z_1, b_1, p_1, t_{11}, r_1$ are scalars. $\mathbf{K}'$, $\mathbf{C}$, $\mathbf{Z}$, $\mathbf{B}$, $\mathbf{M}'$, $\mathbf{N}' \in \mathbb{R}^{N-1}$ and $\mathbf{P}$, $\mathbf{T}$, $\mathbf{R} \in \mathbb{R}^{(N-1)\times(N-1)}$. Then, the modified innovation can be represented as
$$\tilde{Z}_k = \begin{bmatrix} t_{11} & \mathbf{N} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} z_1 \\ \mathbf{Z} \end{bmatrix} + \begin{bmatrix} b_1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} t_{11}z_1 + \mathbf{NZ} + b_1 \\ \mathbf{z} \end{bmatrix}.$$
To bypass the false-data detector, the feasibility constraint (17) needs to be satisfied, i.e.,
$$\begin{bmatrix} t_{11}^2 p_1 + \mathbf{NPN}' + l_1 + t_{11}(\mathbf{MN}' + \mathbf{NM}') & t_{11}\mathbf{M} + \mathbf{NP} \\ (t_{11}\mathbf{M} + \mathbf{NP})' & \mathbf{P} \end{bmatrix}$$
$$= \begin{bmatrix} p_1 & \mathbf{M} \\ \mathbf{M}' & \mathbf{P} \end{bmatrix},$$
based on which one has
$$\begin{cases} t_{11}^2 p_1 + \mathbf{NPN}' + l_1 + t_{11}(\mathbf{MN}' + \mathbf{NM}') = p_1 \\ t_{11}\mathbf{M} + \mathbf{NP} = \mathbf{M}. \end{cases}$$
Consequently,
$$\begin{cases} \mathbf{N} = (1 - t_{11})\mathbf{MP}^{-1} \\ t_{11}^2 = \dfrac{p_1 - \mathbf{MPM}' - l_1}{p_1 - \mathbf{MPM}'}. \end{cases} \tag{19}$$
According to the evolution of the remote estimation error covariance in Proposition 1, to maximize the error covariance
$$\begin{aligned} \tilde{P}_k &= \tilde{P}_k^- + \Delta - \overline{P}C'T_k'K' - KT_kC\overline{P} \\ &= \tilde{P}_k^- + \Delta - \overline{P}(c_1t_{11}k_1 + \mathbf{C}'\mathbf{N}'k_1 + \mathbf{C}'\mathbf{K}') \\ &\quad - (k_1t_{11}c_1 + k_1\mathbf{NC} + \mathbf{KC})\overline{P}, \end{aligned} \tag{20}$$
where $\Delta = \overline{P}C'(C\overline{P}C' + R)^{-1}C\overline{P} \geq 0$, is equivalent to minimize
$$\begin{aligned} &k_1t_{11}c_1 + k_1\mathbf{NC} + \mathbf{KC} \\ &\quad = k_1(c_1 - \mathbf{MP}^{-1}\mathbf{C})t_{11} + k_1\mathbf{MP}^{-1}\mathbf{C} + \mathbf{KC}, \end{aligned} \tag{21}$$
which is an affine function in t_{11}. According to Lemma 2 and $\mathbf{M} = c_1\overline{P}\mathbf{C}'$, the coefficient of t_{11} in (21) can be further simplified as
$$\begin{aligned} k_1(c_1 - \mathbf{MP}^{-1}\mathbf{C}) &= \widehat{P}\frac{c_1^2}{r_1}[1 - \mathbf{C}'(\frac{1}{\overline{P}}\mathbf{P})^{-1}\mathbf{C}] \\ &= \widehat{P}\frac{c_1^2}{r_1}[1 - \mathbf{C}'(\frac{1}{\overline{P}}\mathbf{R} + \mathbf{CC}')^{-1}\mathbf{C}], \end{aligned}$$
where
$$\frac{1}{\overline{P}}\mathbf{P} = \begin{bmatrix} c_2^2 + \frac{r_2}{\overline{P}} & c_2c_3 & \ldots & c_2c_N \\ c_2c_3 & c_3^2 + \frac{r_3}{\overline{P}} & \ldots & c_3c_N \\ \vdots & \vdots & \ddots & \vdots \\ c_2c_N & c_3c_N & \ldots & c_N^2 + \frac{r_N}{\overline{P}} \end{bmatrix}. \tag{22}$$

According to Lemma 3, one has

$$\begin{aligned}
&1 - \mathbf{C}'(\frac{1}{\overline{P}}\mathbf{R} + \mathbf{C}\mathbf{C}')^{-1}\mathbf{C} \\
&= 1 - \mathbf{C}'[\overline{P}\mathbf{R}^{-1} - \overline{P}\mathbf{R}^{-1}\mathbf{C}(1 + \mathbf{C}'\overline{P}\mathbf{R}^{-1}\mathbf{C})^{-1}\mathbf{C}'\overline{P}\mathbf{R}^{-1}]\mathbf{C} \\
&= \frac{1}{1+V} > 0, \qquad (23)
\end{aligned}$$

where $V = \overline{P}\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} = \sum_{i=2}^{N} \overline{P}c_i^2/r_i \geq 0$. Hence, the coefficient of t_{11} in (21) is always positive and the maximum value of $\tilde{P}_k$ is achieved when t_{11} is minimized, i.e., when $l_1 = 0$, $t_{11} = -1$. ■

Based on the above optimal linear attack strategy when one sensor is under attack, the best decision for the malicious agent of which sensor should be attacked is summarized in the following theorem.

Theorem 2 *Consider the multi-sensor system* (1)–(2) *with centralized detection criterion* (12) *under linear attack* (15). *If the attacker can only attack one sensor at each time instant, the optimal linear attack strategy, which maximizes the estimation error covariance, is to attack the sensor i having the largest ratio c_i^2/r_i.*

Proof: Without loss of generality, we assume that $c_1^2/r_1 \geq c_2^2/r_2 \geq \cdots \geq c_N^2/r_N$. What we need to prove is that a larger estimation error covariance will be obtained if the first sensor is under attack rather than any other sensor.

We use subscript $\mathbf{_i}$ in the subsequent proof to represent variables corresponding to sensor $\mathbf{i} \in \{1, 2, \ldots, N\}$. $\mathbf{K_i}$ stands for the matrix K without the i-th column. $\mathbf{C_i}$ refers to the matrix C without the i-th row and $\mathbf{M_i} = c_i\overline{P}\mathbf{C_i'}$. $\mathbf{P_i}$ is the matrix P removing the i-th column and i-th row.

The estimation error covariance when the first sensor is under optimal attack $t_{11} = -1, \mathbf{N_1} = 2\mathbf{M_1}\mathbf{P_1}^{-1}$ is

$$\begin{aligned}
\tilde{P}_k^1 = \tilde{P}_k^- &+ \Delta - \overline{P}(-c_1k_1 + 2\mathbf{C_1'}\mathbf{P_1}^{-1}\mathbf{M_1'}k_1 + \mathbf{C_1'}\mathbf{K_1'}) \\
&- (-k_1c_1 + 2k_1\mathbf{M_1}\mathbf{P_1}^{-1}\mathbf{C_1} + \mathbf{K_1}\mathbf{C_1})\overline{P}. \qquad (24)
\end{aligned}$$

Similarly, when sensor $j \neq 1$ is under attack, the error covariance is

$$\begin{aligned}
\tilde{P}_k^j = \tilde{P}_k^- &+ \Delta - \overline{P}(-c_jk_j + 2\mathbf{C_j'}\mathbf{P_j}^{-1}\mathbf{M_j'}k_j + \mathbf{C_j'}\mathbf{K_j'}) \\
&- (-k_jc_j + 2k_j\mathbf{M_j}\mathbf{P_j}^{-1}\mathbf{C_j} + \mathbf{K_j}\mathbf{C_j})\overline{P}. \qquad (25)
\end{aligned}$$

By substituting $\mathbf{C_1}, \mathbf{K_1}, \mathbf{C_j}, \mathbf{K_j}$, the difference between (24) and (25) can be calculated as

$$\begin{aligned}
&\tilde{P}_k^1 - \tilde{P}_k^j \\
&= 4\left(k_1c_1\overline{P} - k_1\mathbf{M_1}\mathbf{P_1}^{-1}\mathbf{C_1}\overline{P} - k_jc_j\overline{P} + k_j\mathbf{M_j}\mathbf{P_j}^{-1}\mathbf{C_j}\overline{P}\right) \\
&= 4\overline{P}\widehat{P}\left[\frac{c_1^2}{r_1}[1 - \mathbf{C_1'}(\frac{1}{\overline{P}}\mathbf{P_1})^{-1}\mathbf{C_1}] - \frac{c_j^2}{r_j}[1 - \mathbf{C_j'}(\frac{1}{\overline{P}}\mathbf{P_j})^{-1}\mathbf{C_j}]\right] \\
&\qquad (26)
\end{aligned}$$

where the last equality is due to Lemma 2 and $\mathbf{M_i} = c_i\overline{P}\mathbf{C_i'}$.

We partition $\mathbf{C_1}, \mathbf{C_j}, \frac{1}{\overline{P}}\mathbf{P_1}, \frac{1}{\overline{P}}\mathbf{P_j}$ as

$$\mathbf{C_1} = \begin{bmatrix} c_j \\ \mathbb{C} \end{bmatrix}, \frac{1}{\overline{P}}\mathbf{P_1} = \begin{bmatrix} x_1 & \mathbf{Y_1} \\ \mathbf{Y_1'} & \mathbf{Z_1} \end{bmatrix},$$

$$\mathbf{C_j} = \begin{bmatrix} c_1 \\ \mathbb{C} \end{bmatrix}, \frac{1}{\overline{P}}\mathbf{P_j} = \begin{bmatrix} x_j & \mathbf{Y_j} \\ \mathbf{Y_j'} & \mathbf{Z_j} \end{bmatrix},$$

where $\mathbb{C} = [c_2, \ldots, c_{j-1}, c_{j+1}, \ldots, c_N]'$, $x_1 = c_j^2 + r_j/\overline{P}$, $\mathbf{Y_1} = c_j\mathbb{C}'$, $x_j = c_1^2 + r_1/\overline{P}$, $\mathbf{Y_j} = c_1\mathbb{C}'$, and $\mathbf{Z_1} = \mathbf{Z_j} = \mathbf{Z}$ is the matrix in (22) removing the $(j-1)$-th row and column.

According to Lemma 4, one has

$$(\frac{1}{\overline{P}}\mathbf{P_1})^{-1} = \begin{bmatrix} s_1 & -s_1c_j\mathbb{C}'\mathbf{Z}^{-1} \\ -s_1c_j\mathbf{Z}^{-1}\mathbb{C} & s_1c_j^2\mathbf{Z}^{-1}\mathbb{C}\mathbb{C}'\mathbf{Z}^{-1} + \mathbf{Z}^{-1} \end{bmatrix},$$

where $s_1 = (x_1 - \mathbf{Y_1}\mathbf{Z}^{-1}\mathbf{Y_1'})^{-1} = \frac{\overline{P}}{c_j^2\overline{P}(1-\mathbb{C}'\mathbf{Z}^{-1}\mathbb{C})+r_j}$, based on which

$$\begin{aligned}
&1 - \mathbf{C_1'}(\frac{1}{\overline{P}}\mathbf{P_1})^{-1}\mathbf{C_1} \\
&= \frac{r_j}{c_j^2\overline{P}(1 - \mathbb{C}'\mathbf{Z}^{-1}\mathbb{C}) + r_j}(1 - \mathbb{C}'\mathbf{Z}^{-1}\mathbb{C}). \qquad (27)
\end{aligned}$$

Similarly, we can easily obtain that

$$\begin{aligned}
&1 - \mathbf{C_j'}(\frac{1}{\overline{P}}\mathbf{P_j})^{-1}\mathbf{C_j} \\
&= \frac{r_1}{c_1^2\overline{P}(1 - \mathbb{C}'\mathbf{Z}^{-1}\mathbb{C}) + r_1}(1 - \mathbb{C}'\mathbf{Z}^{-1}\mathbb{C}). \qquad (28)
\end{aligned}$$

Substituting (27) and (28) into (26), it is obvious that

$$\frac{c_1^2}{r_1}[1 - \mathbf{C_1'}(\frac{1}{\overline{P}}\mathbf{P_1})^{-1}\mathbf{C_1}] - \frac{c_j^2}{r_j}[1 - \mathbf{C_j'}(\frac{1}{\overline{P}}\mathbf{P_j})^{-1}\mathbf{C_j}] \geq 0$$

is equivalent to

$$\frac{c_1^2r_j^2}{c_j^2\overline{P}(1 - \mathbb{C}'\mathbf{Z}^{-1}\mathbb{C}) + r_j} - \frac{c_j^2r_1^2}{c_1^2\overline{P}(1 - \mathbb{C}'\mathbf{Z}^{-1}\mathbb{C}) + r_1} \geq 0.$$

According to the assumption $c_1^2/r_1 > c_j^2/r_j$ and the fact that $1 - \mathbb{C}'\mathbf{Z}^{-1}\mathbb{C} > 0$,

$$(c_1^4r_j^2 - c_j^4r_1^2)\overline{P}(1 - \mathbb{C}'\mathbf{Z}^{-1}\mathbb{C}) + (c_1^2r_j - c_j^2r_1)r_1r_j \geq 0$$

holds, which completes the proof. ■

Remark 3 *Note that the optimal linear attack strategy is to attack the sensor having the largest ratio c_i^2/r_i. This is consistent with the intuition that tampering the most accurate data leads to the worst estimation quality.*

Remark 4 *The optimal linear attack strategy for a single-sensor system obtained in [17] can be viewed as a special case of the results obtained in this paper when there is no resource constraint for the malicious attacker.*

V. Simulation Example

In this section, we provide numerical examples to demonstrate the analytical results. We consider a system with parameters $a = 0.8$, $q = 1.5$, $C = [c_1\ c_2\ c_3]' = [1.5\ 1.1\ 0.8]'$, $R = \text{diag}\,(r_1, r_2, r_3) = \text{diag}\,(0.7, 0.8, 0.9)$.

For a resource-constrained attacker, Fig. 2 illustrates the optimal linear attack strategy when one sensor is under attack. Without loss of generality, we assume that sensor 1 is under attack. During time period $[0, 9]$, the remote estimator runs a Kalman filter and enters steady state. The blue star line stands for the estimation error covariance when sensor 1 is under the optimal linear attack $t_{11} = -1$, $b_1 = 0$, while the blue diamond line represents that the malicious attacker randomly launches linear attacks on sensor 1 at each time

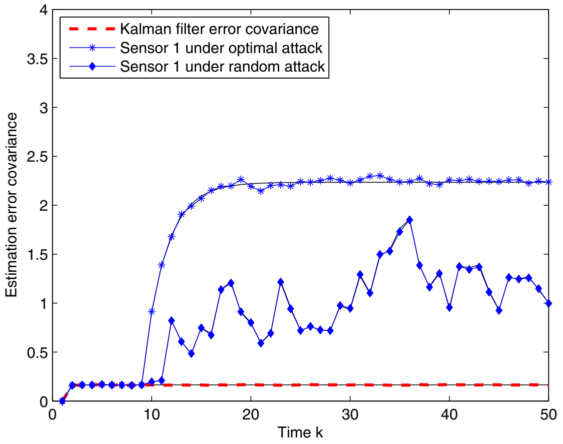

Fig. 2. Remote state estimation error covariances when sensor 1 is under the optimal linear attack $t_{11} = -1, b_1 = 0$ and random attack strategy.

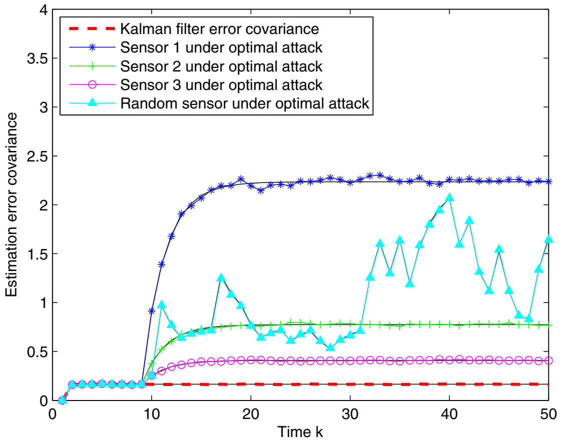

Fig. 3. Remote state estimation error covariances when each sensor $i \in \{1, 2, 3\}$ is under the optimal linear attack $t_{ii} = -1, b_i = 0$ and random sensor is under the optimal linear attack.

instant. It can be easily observed that the optimal linear attack strategy is worse.

Fig. 3 shows the best choice for the resource-constrained attacker, i.e., which sensor should be attacked such that the largest estimation performance degradation is achieved. Besides the estimation error covariance when each sensor $i \in \{1, 2, 3\}$ is under optimal linear attack $t_{ii} = -1$, $b_i = 0$, the case when random senor under optimal attack is also shown in Fig. 3. Note that $c_1^2/r_1 = 3.2 > c_2^2/r_2 = 1.5 > c_3^2/r_3 = 0.7$. According to Theorem 2, the optimal linear attack strategy which maximizes the error covariance is to attack the sensor with largest ratio c_i^2/r_i, i.e., to attack sensor 1. This is consistent with the result observed from the figure. It is also worth noticing that the estimation error covariance converges for a stable system, while it diverges exponentially fast when the system is unstable.

VI. Conclusion

In this paper, we have considered a multi-sensor system in a remote state estimation scenario. We have proposed a linear attack strategy and analyzed the corresponding necessary conditions to bypass the proposed centralized false-data detector. For a resource-constrained attacker, who is able to listen to all the channels but only launches an attack on one sensor at each time instant, we have proved that the optimal linear attack strategy is to attack the sensor with largest ratio c_i^2/r_i using the strategy $t_{ii} = -1$, $b_i = 0$. Simulations are provided to demonstrate the analytical results.

References

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

[3] S. H. Ahmed, G. Kim, and D. Kim, "Cyber physical system: Architecture, applications and research challenges," in *Wireless Days*, 2013, pp. 1–5.

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

[6] 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.

[7] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Hybrid Systems: Computation and Control*. Springer, 2009, pp. 31–45.

[8] 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.

[9] 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.

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

[11] 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.

[12] A. Teixeira, H. Sandberg, G. Dán, and K. H. Johansson, "Optimal power flow: Closing the loop over corrupted data," in *American Control Conference*, 2012, pp. 3534–3540.

[13] 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.

[14] 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.

[15] D. Shi, T. Chen, and M. Darouach, "Event-based state estimation of linear dynamic systems with unknown exogenous inputs," *Automatica*, vol. 69, pp. 275–288, 2016.

[16] D. Shi, R. J. Elliott, and T. Chen, "On finite-state stochastic modeling and secure estimation of cyber-physical systems," *IEEE Transactions on Automatic Control*, 2016.

[17] 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: Special Issue on Secure Control of Cyber Physical Systems*, 2016.

[18] B. D. Anderson and J. B. Moore, *Optimal filtering*. Courier Corporation, 2012.

[19] S. C. Ergen, "ZigBee/IEEE 802.15. 4 summary," *UC Berkeley, September*, vol. 10, p. 17, 2004.

[20] R. K. Mehra and J. Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems," *Automatica*, vol. 7, no. 5, pp. 637–640, 1971.

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