53rd IEEE Conference on Decision and Control
December 15-17, 2014. Los Angeles, California, USA

Coding Sensor Outputs for Injection Attacks Detection

Fei Miao Quanyan Zhu Miroslav Pajic George J. Pappas.

Abstract— **This paper considers a method of coding the sensor outputs in order to detect stealthy false data injection attacks. An intelligent attacker can design a sequence of data injection to sensors that pass the state estimator and statistical fault detector, based on knowledge of the system parameters. To stay undetected, the injected data should increase the state estimation errors while keep the estimation residues in a small range. We employ a coding matrix to the original sensor outputs to increase the estimation residues, such that the alarm will be triggered by the detector even under intelligent data injection attacks. This is a low cost method compared with encryption over sensor communication networks. We prove the conditions the coding matrix should satisfy under the assumption that the attacker does not know the coding matrix yet. An iterative optimization algorithm is developed to compute a feasible coding matrix, and, we show that in general, multiple feasible coding matrices exist.**

I. Introduction

Cyber-physical systems (CPSs) integrate computation and communications to interact with physical processes. Many applications are considered as CPSs, including high confidence medical devices, energy conservation, environmental control, and safety critical infrastructures–such as water supply systems, electric power, and communication systems [1]. Therefore, security is a critical aspect of these systems, and CPSs involve additional challenges in control layer. The problem of secure control is defined, and defenses from information security, sensor network security are analyzed in [2]. However, due to the interaction of cyber systems with physical world, these mechanisms alone are not sufficient for the security of CPSs [2].

The reasons that CPSs are vulnerable to cyber attacks and key challenges are summarized in [3]. Novel attack-detection algorithms besides existing information protection methods in cyber security area can be designed, by understanding how attacks affect state estimation and control of the system. Stealthy attacks from an intelligent attacker that can access a partial model of the system are synthesized in [4], and tools to protect state-estimation components in CPSs are developed. Since large numbers of measurements are sent over unencrypted communication channels in power grids [5], the authors developed two algorithms to maximize the utility of encrypted devices placed to increase system security.

Fault detection, isolation and reconfiguration (FDIR) methods have been explored to ensure system safety requirements and robustness [6]. Although active techniques have been designed to tackle various types of attacks, fundamental limitations still exist, as characterized in [7]. Fawzi et al. propose estimation and control schemes of noise free linear systems, when some of the sensors or actuators are corrupted in [8]. Pajic et al. present a robust state estimation method in presence of attacks to no more than half of the sensors for systems with noise and modeling errors [9]. In contrast, we assume a different case where the attacker can inject an arbitrary vector to the communication between sensors and the estimator/detector/controller block, thus no element of the injection vector is constrained to be zero.

In this work, we consider a type of deceptive cyber attack– false data injection attacks to sensor outputs. We assume that there is no physical attack on individual measurements or sensors. The monitoring system can detect malicious behaviors in general. Miao et al. design a stochastic game approach for replay attacks detection when the system is equipped with an active controller, a filter and a statistical detector [10]. However, with knowledge of the system model, an intelligent cyber attacker is able to carefully design a data injection sequence, such that the state estimation error increases without triggering the alarm of the monitor [11], [12]. It is also shown that by only compromising actuators, attackers can never introduce infinite estimation errors that passing a monitor like χ^2 detector [12]. Therefore, we focus on intelligent sensor false data injection attacks and assume actuators are secure in this paper. Regarding the computational overhead of encryptions on embedded architectures [13], we propose an alternative low cost method to code the sensor measurements for detection.

The main contribution of this work is a low cost method of coding sensor outputs to detect stealth sensor false data injection attacks. We assume that the coding matrix is secured, sent to sensors by a side channel before the coding starts. We show that even if the attacker knows the system model without the coding scheme, the system can detect the original stealthy sensor injections by coding the sensor outputs according to certain conditions. We also design an iterative optimization algorithm to compute such coding matrices, and show that in general, multiple feasible coding matrices exist. By encrypting only the coding matrix channel once, the coding approach saves encryption cost compared with encrypting all sensor outputs. Results presented in this

This material is based on research sponsored by DARPA under agreement number FA8750-12-2-0247. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

F. Miao, M. Pajic and G. J. Pappas are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA 19014. Q. Zhu is with Department of Electrical and Computer Engineering, New York University, Brooklyn, NY, USA 11201. Email: {miaofei, pajic, pappasg}@seas.upenn.edu,{quanyan.zhu}@nyu.edu

work also provide a basis for the analysis of situations where the attacker can learn the coding scheme in some steps. In this case, the system can either change a new coding matrix or randomly use a set of coding matrices to fool the attacker.

The paper is organized as follows. In Section II we describe the system and attack models. The conditions that a feasible coding matrix should satisfy are presented in Section III, with a proof in Appendix. An iterative optimization algorithm to find a feasible coding matrix is developed in Section IV. Section V shows illustrative examples. Conclusions are given in Section VI.

II. System And Attack Model

We will introduce the normal system model and deception attack model in this section. The system architecture with a discrete-time linear time-invariant (LTI) system and false data injection attack to sensors is shown in Figure 1.

A. Linear system model

Assume the CPS is a discrete time LTI system with the following form:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k, \\ y_k &= Cx_k + v_k, \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the system state vector at time k, $u_k \in \mathbb{R}^m$ is the control input at time k, and $y_k \in \mathbb{R}^p$ is the sensor observation vector. We do not have specific restrictions for the linear control input u_k here, and we will illustrate why the controller does not affect the detection of false data injection later. We assume $w_k \sim N(0, Q)$ and $v_k \sim N(0, R)$, are identical independent Gaussian noises and initial state of the system satisfies $x_0 \sim N(0, \Theta)$.

The optimal Kalman filter used to estimate state $\hat{x}_{k|k}$ is:

$$\begin{aligned} &\hat{x}_{0|-1} = 0, \quad P_{0|-1} = \Theta, \hat{x}_{k+1|k} = A\hat{x}_k + Bu_k, \\ &P_{k+1|k} = AP_kA^T + Q, \\ &K_{k+1} = P_{k+1|k}C^T(CP_{k+1|k}C^T + R)^{-1}, \\ &P_{k+1} = (I - K_{k+1}C)P_{k+1|k}, \\ &z_{k+1} = y_{k+1} - C(A\hat{x}_k + Bu_k), \hat{x}_{k+1} = \hat{x}_{k+1|k} + K_kz_{k+1}. \end{aligned}$$

Under the assumption that (A, B) is stabilizable, (A, C) is detectable, we get a steady state Kalman filter, with the error covariance matrix P and Kalman gain matrix K:

$$P \triangleq \lim_{k\to\infty} P_{k|k-1}, K \triangleq PC^T(CPC^T + R)^{-1}.$$

By the property of Kalman filter, matrix $(A-KCA)$ is stable. Here z_{k+1} is the estimation residue. Without attacks, residue z_k follows a Gaussian distribution $N \sim (0, CPC^T + R)$.

Define the quantities g_k as: $g_k = z_k^T P^{-1} z_k$, where $\mathcal{P}^{-1}$ is the error covariance matrix of Kalman filter, then g_k satisfies a χ^2 distribution with p degrees of freedom. By using g_k, we consider the standardized residue sequence $\eta_k = P^{-\frac{1}{2}} z_k$ for a monitoring system–a χ^2 failure detector, and assume there exists a δ_η such that $\lim_{k\to\infty} \|E\eta_k\| \leq \delta_\eta$. We denote α as the threshold for the alarm, meaning that the alarm is triggered when $g_k > \alpha$.

B. False data injection attack model

In this paper, we assume that actuators are secure and only consider the case of sensor attacks. It has been shown that by only compromising actuators, attackers can never induce infinite estimation error without being detected under monitoring systems like a χ^2 detector [12]. Therefore, we focus on intelligent sensor false data injection in this paper. The system model under attack is described as:

$$x'_{k+1} = Ax'_k + Bu'_k + w_k, y'_k = Cx'_k + y^a_k + v_k, \tag{2}$$

where $y^a_k \in \mathbb{R}^m$ is an arbitrary vector injected by the attacker at time k. Assume the adversary has knowledge of the system model described in Section II-A, and the ability to inject data over communication network between sensors and the estimator/detector/controller.

C. The difference between normal and compromised systems

To illustrate how the sensor injection sequence y^a_k will affect the estimation and monitoring system, we examine how the estimation error and residue will change with y^a_k. Let $\hat{x}'_k$ be the state estimation of the compromised system, $z'_k = y'_{k+1} - C(A\hat{x}'_k + Bu'_k)$ be the estimation residue of the compromised system. Then, the difference between the normal and the compromised systems can be captured by:

$$\begin{aligned} &e_k \triangleq x_k - \hat{x}_k, e'_k \triangleq x'_k - \hat{x}'_k, \\ &\Delta e_k \triangleq e'_k - e_k, \Delta z_k \triangleq z'_k - z_k. \end{aligned} \tag{3}$$

The dynamics of the above difference vectors satisfy:

$$\begin{aligned} \Delta e_{k+1} &= (A - KCA)\Delta e_k - Ky^a_{k+1}, \\ \Delta z_{k+1} &= CA\Delta e_k + y^a_k, \end{aligned} \tag{4}$$

Hence the difference vectors between normal and compromised systems – $\Delta z_k(y^a), \Delta e_k(y^a)$ – are functions of the attack sequence $y^a \triangleq (y^a_0, y^a_1, \dots)$. To simplify notations, we concisely denote these vectors as $\Delta z_k, \Delta e_k$.

The objectives of the attacker include maximizing the estimation error e'_k without triggering the alarm, while increasing x'_k to infinity. When the system is secured, $\lim_{k\to\infty} \mathbf{E}[e_k] \to 0$, and z_k is in a small range. Thus the attacker's objective is equivalent to increasing $\|\Delta e_k\|_2$ (the difference between estimation error of the normal and compromised systems) without increasing $\|\Delta z_k\|_2$ much. The

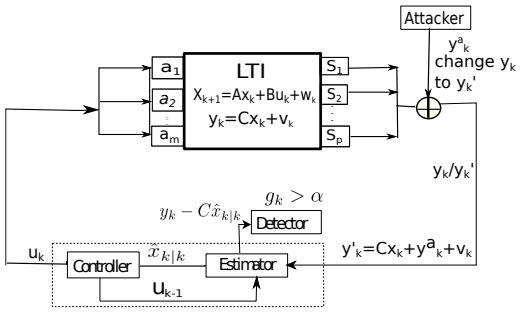

Fig. 1. System diagram, assume the attacker can inject arbitrary false data vector y^a_k to sensor outputs.

probability that y_k^a is detected is $Pr(g'_k = (z'_k)^{'T}\mathcal{P}^{-1}z'_k > \alpha)$. Since computing g'_k is integrating Gaussian on an ellipsoid, the stealthy requirement can be approximated by keeping $\|z'_k\|_2$ small. Residues of the normal system z_k is bounded, and the attacker should keep the change of residues bounded make the injection stealthy. It means the following inequality should hold

$$\|\Delta z_k\|_2 \leq M, \tag{5}$$

where M is a residue norm change threshold designed by the attacker. The compromised estimation residue should be close to that of the normal system, to fool the monitoring system. [1] When y_k^a can be an arbitrary vector, a necessary and sufficient condition for a stealth y_k^a that can increase $\|e'_k\|_2$, $\|x'_k\|_2$ to infinity while keep $\|z'_k\|_2$, $\|\Delta z_k\|_2$ bounded is derived in [12], [11]. One of the conditions that $Cv \in span(I)$, i.e., there exists y^* satisfying $y^* = Cv$ is always satisfied by the attack model (2). Hence, we have the following corollary.

Corollary 1: There exists a stealth sequence $y_k^a, k = 0, 1, \ldots,$ given the attacked system model (2), if and only if matrix A has an unstable eigenvalue λ and the corresponding eigenvector v, such that $v \in span(Q_{oa})$, where Q_{oa} is the controllability matrix associated with the pair $(A - KCA, K)$. □

III. Coding Sensor Outputs For Detecting Stealth Sensor Data Injection

Existing active monitor schemes (design some additive control input u_k^d) and fault detection filters have limitations, that even with secured actuators, they can not detect stealthy data injection attacks. It is necessary to design some inexpensive techniques to compensate for the vulnerability of the system under intelligent sensor data injection attacks.

A. Limitations of existing approaches

The limitation of active monitor approach: Under the assumption that actuators work appropriately for the attacked system (2), the challenge here is whether adding u_k^d to the pre-designed linear control input u_k (like optimal LQG control) can detect stealthy sensor data injections. It is worth noting that active monitor approaches does not help given sensor data injection attacks satisfying Theorem 1. For model (2), the control input does not affect the estimation residue change quantity Δz_{k+1} according to (4), which means there exists no $\tilde{u}_k = u_k + u_k^d, \tilde{u}'_k = u'_k + u_k^d$ that can increase $\|\Delta z_{k+1}\|_2$ under y_k^a. This is because any additional control input will be eliminated by the deduction of z_{k+1} and z'_{k+1} to get Δz_{k+1}. The limitations of active monitors for a unified LTI model are proved in Theorem 4.7 of [7]. [2] In this perspective, different linear controllers are equivalent under stealth sensor data injection attacks, and we do not restrict the controller model for designing our detection techniques.

[1] The relation between the scale or norm of the injection sequence and the alarm trigger threshold α is shown in Theorem 1 in [12].

[2] A different case when adding exogenous Gaussian distribution control input can detect replay attacks is discussed in [14].

The limitation of fault detection filter: Besides Kalman filter, observer-based fault detection filters for LTI systems with unknown error have been developed. The design requirements usually include robustness to unknown inputs and sensitivity to faults. Such filters generate a different residue from z_k of Kalman filter. Consider the following form of residual generator and residual evaluator (including a threshold and a decision logic unit, see [15] for details) [15]:

$$\begin{aligned} \hat{x}_{k+1} &= A\hat{x}_k + Bu_k + H(y_k - \hat{y}_k), \\ \hat{y}_k &= C\hat{x}_k, r_k = V(y_k - \hat{y}_k), \end{aligned} \tag{6}$$

where $\hat{x}_k \in \mathbb{R}^n$ and $\hat{y}_k \in \mathbb{R}^p$ represent the state and output estimation vectors, respectively, and r_k is the residual signal. This fault detector shares the same limitation with Kalman filter, i.e., the intelligent sensor data injection attack is stealth for the filter described as (6), since the residue is still observer based difference between y_k and $\hat{y}_k$.

B. Coding sensor outputs to detect stealth data injection

Since existing monitoring system can not detect intelligent false data injection attacks, and encryption method has a constraint of significant computation overhead, we propose a design of *coding the sensor outputs* to detect stealth sensor data injection attacks. An intelligent attacker designs the sequence y_k^a carefully to keep the change of residue $\|\Delta z_k\|_2 \leq M$, where M is a constant. Thus the objective of a detection approach is equivalent to increasing $\|\Delta z_k\|_2$ to infinity as time goes to infinity.

The necessary and sufficient conditions for stealth false sensor data injection in Corollary 1 assume that the attacker knows (A, B, C, K). Parameters A and B are related to physical dynamics that may not be altered, while C is related to the sensor measurements, corresponding specific physical states. Without changing the physical setup, we still can manipulate the sensor outputs. To violate the attacker's design, we consider the method of transforming sensor outputs as shown in Figure 2–instead of sending the output vector $\mathbf{y}_k = Cx_k + v_k$ to the estimator/controller/detector, sensors are transmitting the value:

$$Y_k = \Sigma(Cx_k + v_k), C \in \mathbb{R}^{p\times n}, \tag{7}$$

where $\Sigma \in \mathbb{R}^{p\times p}$ is an invertible matrix. One can think of Σ as an inexpensive code. We assume that the attacker does not know the matrix Σ, and designed a sequence of stealth attack signal y_k^a with parameters (A, B, C, K). One can compare Σ with a secured key. By encrypting only the coding matrix channel once, the coding approach saves encryption cost

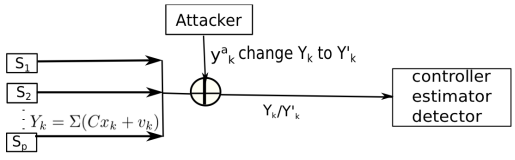

Fig. 2. System diagram when coding sensor outputs with a matrix Σ that satisfies the conditions of Theorem 1. Assume the attacker can inject arbitrary false data vector y_k^a to sensor outputs.

compared with encrypting all sensor outputs. In this case, the false sensor value after transforming changes to:

$$Y'_k = \Sigma C x_k + y^a_k + \Sigma v_k. \tag{8}$$

Assume the corresponding Kalman filter with sensor output Y_k has a steady state estimation error covariance matrix P' and a Kalman gain matrix K' that satisfy

$$K' = P'C^T\Sigma^T(\Sigma C P' C^T \Sigma^T + \Sigma R)^{-1}.$$

Similarly as (4), define $\Delta e'_k, \Delta z'_k$ as the change of state estimation and residue for the sensor output (7), (8). With Σ, a stealth data injection designed for (1) (with parameters (A, B, C, K)), $\|\Delta z'_k\|_2$ increases to infinity as $k \to \infty$ under certain conditions. In the following theorem, we show the sufficient conditions that Σ should satisfy for any stealth sequence of $y^a_k, k = 0, 1, \dots$ that satisfies Theorem 1.

Theorem 1: Given an attacked system model (2), assume that the attacker designs a sequence of sensor data injection $y^a_k, k = 0, 1, \dots$, based on one unstable eigenvector v of A that satisfies Corollary 1. If there exists an invertible matrix Σ such that $(A, \Sigma C)$ is detectable, and the direction of $\Sigma C v$ is not the same with that of Cv, i.e.,

$$\frac{(Cv)'\Sigma C v}{\|\Sigma C v\|_2 \|C v\|_2} \neq 1, \tag{9}$$

then after injecting y^a_k the estimation residue change $\|\Delta z'_k\|_2$ is increased , by coding sensor outputs (7) with Σ. □

Proof: See Appendix. ■

We call a matrix Σ that satisfies the conditions of Theorem 1 a feasible coding matrix. Theorem 1 proves that even the attacker knows system parameters (A, B, C, K), without changing the physical structure or altering A, B, we can utilize the sensor data to get different residues for detecting. Leveraging sensor outputs is the key reason to detect a stealth sensor data injection. It is worth noting that here we do not constrain specific structure of the matrix Σ besides conditions in Theorem 1. For an LTI system, ΣC is simply a linear transform of the original sensor measurement. When A has several unstable eigenvectors satisfying Corollary 1, the following lemma extends the result of Theorem 1.

Lemma 1: Given an attacked system (2) with a set of unstable eigenvectors $v_1, \dots, v_u$ satisfying Corollary 1, if Σ is an invertible matrix such that $(A, \Sigma C)$ is detectable, and

$$\frac{(C\tilde{v})'\Sigma C \tilde{v}}{\|\Sigma C \tilde{v}\|_2 \|C \tilde{v}\|_2} \neq 1, \tag{10}$$

for any linear combination of $v_1, \dots, v_u$ – $\tilde{v}$, then Σ is a feasible coding matrix to increase $\|\Delta z'_k\|_2$ for any stealth data injection to attacked system (2). □

Remark 1: When the attacker is able to learn Σ by analyzing sensor outputs and actuator inputs, the system can send a new Σ before the attacker figures out the current applied coding matrix. This will be an avenue for future work. □

IV. Algorithm to Compute A Coding Matrix

To compute a set of feasible coding matrices, we first analyze requirements of an optimal coding matrix, and then decouple the requirements to an approximate optimization problem and a corresponding system parameter design problem. The constraints are relaxed step by step to achieve an iterative algorithm, such that each iteration step of the algorithm is a convex optimization problem.

The transformed observer data should maximize the difference between estimation residue of the normal and attacked system – $\|\Delta z'_k\|_2$, which is equivalent to maximizing $\|\Sigma C v - C v\|_2$ (or $\tilde{v}$), by the proof of Theorem 1. We do not want to sacrifice the state estimation performance of the coded the system, and $\Delta e'_k$ should not increase fast compared with Δe_k, thus we require that $(A, \Sigma C)$ is detectable for a steady state Kalman filter. With transformed sensor values Y_k in (7), we compute a new gain matrix K' for the steady state Kalman filter, such that $\tilde{A} = A - K'\Sigma C A$ is stable. Hence the estimation error of the normal system with sensor output $Y_k = \Sigma C x_k + v_k$ satisfies

$$\lim_{k\to\infty} E[\tilde{e}_{k+1}] = (A - K'\Sigma C A) E[\tilde{e}_k] \to 0. \tag{11}$$

The norm of Σ should be bounded, because computable values of the estimator, controller, and detector are bounded. These requirements and objective are described as

$$\begin{aligned} &\text{maximize } \|\Sigma C v - C v\|_2 \\ &\text{subject to } \|\Sigma\|_2 \leq \gamma, \Sigma \text{ invertible}, \\ &\qquad (A, \Sigma C) \text{ detectable}, \\ &\qquad (A - K'\Sigma C A) \text{ stable}. \end{aligned} \tag{12}$$

The above problem (12) is not convex, since maximizing a norm is a concave function of Σ. There is no closed form equation between K' and Σ for the constraint related to K'. To decouple the design process, we ignore the constraint related to K' and detectability of $(A, \Sigma C)$, compute an optimal Σ satisfying other constraints first, and check the ignored constraints later.

Given the system parameter matrices C, A, ignoring the requirement about K', the objective of (12) is approximately to find an invertible, bounded Σ that maximizes $\|\Sigma C v - C v\|_2$. In general, when Σ is unbounded, $\|\Sigma C v - C v\|_2 \to \infty$. Considering the direction change of vector Cv after transformed by $\Sigma C v$, the optimal direction to maximize the difference between $\Sigma C v$ and Cv is the orthogonal direction. Thus, we have the following Lemma 2.

Lemma 2: Assume the attacker designs a sequence of injections that satisfies Corollary 1, based on an unstable eigenvector v. If there exists a feasible solution of Σ that satisfies the following constraints (13),

$$(Cv)'\Sigma C v = 0, \|\Sigma\|_2 \leq \gamma, \ \Sigma \text{ is invertible}, \tag{13}$$

and $(A, \Sigma C)$ is detectable, then we have an optimal direction coding matrix Σ that satisfies Theorem 1. □

Remark 2: When Σ should work for any linear combination of multiple unstable eigenvectors $\tilde{v}$, we start from an invertible orthogonal rotation matrix Σ that rotates any vector through $\frac{\pi}{2}$ angle. Then check other constraints, if they are violated, we will find a heuristic algorithm. This will be an avenue for future work. □

For system with only one unstable eigenvector that satisfies Corollary 1, solving (13) directly may not return a feasible Σ satisfying Theorem 1. The following convex optimization formulation presents a relaxed problem as one iteration. When $\Sigma = \mathbf{0}$, Σ is not invertible, though $(Cv)'\Sigma Cv = 0$. So we use a conservative constraint $\Sigma \succ 0$ to replace the invertible constraint. Log determinant function of $\Sigma \succ 0$ will drive Σ away from an all zero elements matrix, and the objective function is concave. There always exists a solution for (13) (an orthogonal rotation matrix), however, when we restrict the solution space to be positive-semidefinite (SDP), and $(A, \Sigma C)$ to be detectable, $(Cv)'\Sigma Cv = 0$ is a strict constraint. When this is the case, we relax (13) to an inequality constraint with a small ϵ, and get the formulation:

$$\begin{aligned} \text{maximize} \quad & \log\det\Sigma \\ \text{subject to} \quad & |(Cv)'\Sigma Cv| \le \epsilon \|Cv\|_2^2, \\ & \|\Sigma\|_2 \le \gamma, \Sigma \succ 0. \end{aligned} \tag{14}$$

Thus, we have the following iteration algorithm. Algorithm 1 starts from the constraint that $(Cv)'\Sigma Cv = 0$, i.e., the orthogonal transform. During each iteration, the algorithm relaxes the constraint, increases ϵ, till there is a feasible Σ satisfying $(A, \Sigma C)$ is detectable. With such a Σ and a corresponding K', (11) is guaranteed. The change of estimation error of the coded system under a data injection attack– $\Delta e'_k$ should not increase too fast compared with Δe_k. It means we do not need to sacrifice the estimation performance to detect stealth data injection attacks. This result is also shown in Section V.

Algorithm 1 : Compute a feasible coding matrix Σ

Input: System model parameters A, B, C, K, an unstable eigenvalue and eigenvector λ, v of A, and a stealth sensor data injection sequence y^a_k.
Initialization: Set the total iteration step number T, $\epsilon = 0$, and $\Delta\epsilon$–the increase step size of ϵ.
Iteration: For $t = 1, 2 \ldots, T$, compute problem (14):
If there is no feasible Σ, let $\epsilon = \epsilon + \Delta\epsilon$, $t = t + 1$.
If there exists a Σ, check the detectability of $(A, \Sigma C)$:
if $(A, \Sigma C)$ is not detectable, let $\epsilon = \epsilon + \Delta\epsilon$, $t = t + 1$;
else if $(A, \Sigma C)$ is detectable, compute a steady state Kalman filter gain matrix K' corresponding to the new sensor outputs $Y_k = \Sigma C x_k + v_k$, and stop the iteration, return the result.
Return: A feasible transform matrix Σ and a corresponding steady state kalman filter gain matrix K'.

Algorithm 1 is terminated once a feasible Σ is computed.In the worst case, when $\epsilon \to \gamma$, solution $\Sigma = \gamma I$ satisfies all the constraints, and $(A, \gamma IC)$ is detectable, Algorithm 1 will terminate. It is worth noting that Theorem 1 is a sufficient condition. Even the direction of ΣCv is the same with that of Cv, with $\Sigma Cv \neq Cv$ we still increase $\|\Delta z\|'_2$. The norm increasing speed is relatively slow though. This is explained in the proof of Theorem 1.

V. Illustrative Examples

We show the effects of coding sensor outputs by examples of two-dimensional LTI systems. Consider a detectable 2-dimensional linear system with parameters:

$$\mathbf{A} = \begin{bmatrix} 0.8 & 0 \\ 0.5 & 1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 2 & 0.5 \\ 0 & 1 \end{bmatrix}, \mathbf{D} = 0,$$

where A has an unstable eigenvalue $\lambda = 1$ and eigenvector $v = [0\ 1]^T$. One stealth attack sequence is: $y^a_0 = [0.0588\ 0.0588]^T$, $y^a_1 = [0.1286\ -0.9706]^T$, $y_k = y^a_{k-2} - y^a_0, k \ge 2$. By Algorithm 1, we get a feasible coding matrix Σ_1. Note that Theorem 1 does not require Σ to be an SDP matrix. Another feasible matrix Σ_2 that satisfies Theorem 1, but not an SDP calculated by Algorithm 1 is shown.

$$\Sigma_1 = \begin{bmatrix} 2 & -0.5 \\ -0.5 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}.$$

Figure 3 shows the comparison result of Δz_k, $\Delta z'_k$, and $\Delta z'_k$ increases with time k after coded by Σ_1, while without coding Δz_k is bounded. Figure 4 shows that for the sensor outputs transformed by Σ_2, $\Delta z'_k$ increases with time k, while

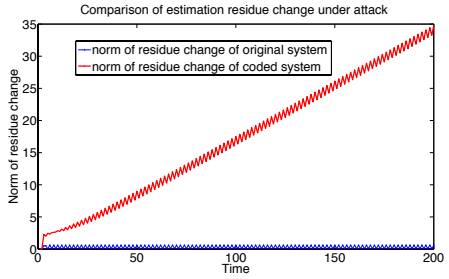

Fig. 3. Comparison of norms of Δz_k, $\Delta z'_k$ for Σ_1

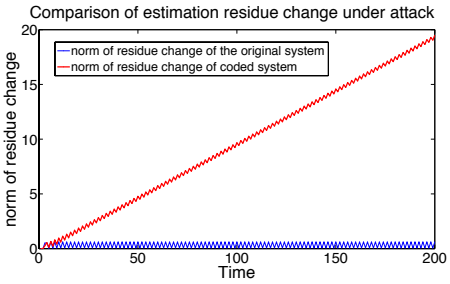

Fig. 4. Comparison of norm of residue change between the original system and coded system, Δz_k and $\Delta z'_k$, for Σ_2 that satisfies Theorem 1 but is not an SDP matrix.

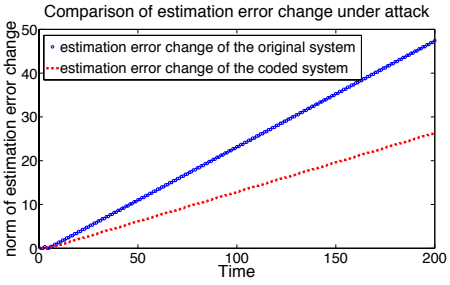

Fig. 5. Comparison of norm of estimation error change between the original system and coded system, Δe_k and $\Delta e'_k$, for Σ_2 that satisfies Theorem 1 but is not an SDP matrix.

the original system Δz_k stays inside a bounded range. For the transformed sensor outputs, change of the estimation error $\Delta e'_k$ increases even slower than Δe_k under data injection attack as shown in Figure 5. By comparing the change of estimation error Δe_k and $\Delta e'_k$, we show that estimation error of a coded system does not necessarily increase faster than the original system.

VI. Conclusion

In this work, we have proposed a method of coding sensor outputs to detect stealth sensor data injection attacks, designed by an intelligent attacker with system model knowledge. Without changing physical setup, we transform the sensor outputs and provide conditions when a linear combination of original sensor outputs can help to detect a stealthy injection sequence. An iterative optimization algorithm is developed to compute a feasible transform matrix efficiently, and examples show detection effects after coding sensor values. In the future, we will explore scenarios where the attacker is capable to learn the coding matrix.

Appendix

Proof of Theorem 1

Proof: Given a system under data injection attacks as (2), we assume that the system has one unstable eigenvector v with corresponding eigenvalue λ. According to the definition in equation (3), the dynamics of $\Delta e_k, \Delta z_k$ satisfy

$$\begin{aligned} \Delta e_{k+1} &= (A - KCA)\Delta e_k - K y^a_{k+1}, \\ \Delta z_{k+1} &= CA\Delta e_k + y^a_k. \end{aligned} \tag{15}$$

Similarly, for coded sensor outputs (8),

$$\begin{aligned} \Delta e'_{k+1} &= (A - K'\Sigma CA)\Delta e'_k - K' y^a_{k+1}, \\ \Delta z'_{k+1} &= \Sigma CA\Delta e'_k + y^a_k, \end{aligned} \tag{16}$$

Based on the proof of *Theorem* 1 in [12], the only component of Δe_k that goes to infinity eventually is

$$c_k v, \lim_{k\to\infty} c_k = \infty, \tag{17}$$

and Δe_k can be decomposed as

$$\Delta e_k = c_k v + \epsilon_{1k}, \|\epsilon_{1k}\|_2 \le M_1. \tag{18}$$

To keep Δz_k bounded as $k \to \infty$, any stealth injection sequence y^a_k must satisfy

$$y^a_k = -c_k \lambda C v + \epsilon_{2k}, \|\epsilon_{2k}\|_2 \le M_2, k = 0, 1, 2, \ldots, \tag{19}$$

where M_2 is a constant such that $\|\Delta z_k\|_2 \le M$ for all k.

We assume that the attacker does not know Σ, and designs a injection sequence for the original system (1) as described in (19). Similarly as Δe_k, the only component of $\Delta e'_k$ that can goes to infinity is $c_k\lambda$, since matrix A is not changed by the coding matrix Σ. However, with any y^a_k in (19), $\Delta z'_k$ can be decomposed as

$$\Delta z'_k = c_k \lambda(\Sigma C v - C v) + \epsilon_{3k}, k = 0, 1, 2, \ldots, \tag{20}$$

where ϵ_{3k} is a bounded vector components of $\Delta z'_k$. When Σ satisfies equation (9), $\Sigma C v - C v \neq 0$. With $c_k \to \infty$, $\|\Delta z'_k\| \to \infty$ as $k \to \infty$.

When there are a set of unstable eigenvectors $v_1, \ldots, v_u$ and eigenvalues $\lambda_1, \ldots, \lambda_u$, the above proof still holds after replacing $c_k v$ with $\sum_{i=1}^{u} c_{ik} v_i$, $c_k \lambda C v$ with $\sum_{i=1}^{u} c_{ik}\lambda_i C v_i$.

When $(A, \Sigma C)$ is detectable, there exists a steady state Kalman filter with parameter K' for the coded system, and the corresponding fault detector. Hence, when the attacker designs a stealth injection sequence without knowledge of Σ, the system can detect it by increasing $\|\Delta z'_k\|_2$ with Σ. ■

References

[1] K.-D. Kim and P. R. Kumar, "Cyber-physical systems: A perspective at the centennial," *Proceedings of the IEEE*, pp. 1287–1308, 2012.

[2] 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.

[3] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, 2009.

[4] A. Teixeira, S. Amin, H. Sandberg, K. Johansson, and S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *2010 49th IEEE Conference on Decision and Control (CDC)*, 2010, pp. 5991–5998.

[5] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2010, pp. 214–219.

[6] I. Hwang, S. Kim, Y. Kim, and C. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *Control Systems Technology, IEEE Transactions on*, vol. 18, no. 3, pp. 636–653, 2010.

[7] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, Nov 2013.

[8] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[9] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas, "Robustness of attack-resilient state estimators," in *2014 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, 2014, pp. 163–174.

[10] F. Miao, M. Pajic, and G. Pappas, "Stochastic game approach for replay attack detection," in *IEEE 52nd Annual Conference on Decision and Control (CDC)*, Dec 2013, pp. 1854–1859.

[11] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *First Workshop on Secure Control Systems, CPS Week*, 2010.

[12] W. C.Kwon and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *Amerian Control Conference (ACC)*, June 2013.

[13] P. Ganesan, R. Venugopalan, P. Peddabachagari, A. Dean, F. Mueller, and M. Sichitiu, "Analyzing and modeling encryption overhead for sensor network nodes," in *2nd ACM International Conference on Wireless Sensor Networks and Applications*, 2003, pp. 151–159.

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

[15] M. Zhong, S. X. Ding, J. Lam, and H. Wang, "An LMI approach to design robust fault detection filter for uncertain LTI systems," *Automatica*, vol. 39, no. 3, pp. 543 – 550, 2003.