Proceedings of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Online Deception Attack Against Remote State Estimation

Heng Zhang [*] **Peng Cheng** [*] **Junfeng Wu** [**] **Ling Shi** [**]
Jiming Chen [*]

[*] *State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, P.R. China (e-mail: ezhangheng@gmail.com, pcheng@iipc.zju.edu.cn, jmchen@ieee.org).*
[**] *Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, P.R. China (e-mail: jfwu@ust.hk,eesling@ust.hk)*

Abstract: Security issue has become a new hotspot in cyber-physical systems (CPS) research field in recent years due to the vulnerability of CPS to security threats. This paper focuses on stealthy deception attack in remote state estimation, which is one typical attack in CPS. From the standpoint of deception attacker, we investigates how to design proper deception attack strategy to degrade the state estimation quality with communication rate constraint. We design an online attack strategy and prove that the proposed attack strategy can degrade the estimation quality. To study the effectiveness of the proposed strategy, we analyze the cost deviation, which depicts the difference between the estimation quality with and without the proposed attack strategy. Our results show that the cost deviation will be maximum when the communication rate is 0.75. A numerical example is presented to demonstrate the main results.

Keywords: Cyber-Physical Systems; Secure; State estimation.

1. INTRODUCTION

Cyber-physical systems (CPS), which smoothly integrate information and physical elements, have a large spectrum of applications, including smart grid (Bitar et al., 2011), smart building (Novak and Gerstinger, 2010), intelligent transportation (Qu et al., 2010), public health (Sarwate and Chaudhuri, 2013), etc. Due to its importance, it is of great research interests to investigate the vulnerability of CPS under various threats launched in either cyber or physical space (Zhang et al., 2014). A well-known example is the Stuxnet worm, which attacked Iran's nuclear facilities and resulted in more than 1000 centrifuges (10 percent) breakdown between November 2009 and late January 2010 (Wilson, 2013).

In this paper we focus on stealthy deception attack, which compromises sensor nodes, aiming at degrading the system performance without being detected (Cardenas et al., 2009). A typical deception attacker can capture the sensor nodes, exploit its unauthorized privileges to inject malicious code or modify the program, and then deteriorate the system performance stealthily (Bryant et al., 2004; Song et al., 2007; Kavitha and Sridharan, 2010).

One basic issue in CPS security is to study the consequence of attack actions (Shoukry et al., 2013). Zhang et al. (2013) has studied an optimal offline DoS attack strategy against state estimation, where, subject to an energy constraint in a finite time horizon, the attacker jams the transmission channel without being detected. In (Zhang et al., 2013), it was assumed that the sensor can always send the data to the estimator and every data can be received by the estimator with a certain probability. However, if the sensor has energy constraint or communication bandwidth constraint, it cannot send the data in every time slot. Therefore needs to design its transmission schedule to improve the estimation quality. Wu et al. (2013b) designed an online transmission schedule under communication rate constraint to minimize the state estimation error. It is interesting and challenging to design an attack strategy to maximize the attack effect under such communication rate constraint.

Since the remote estimator may detect the attack behavior if the communication rate constraint is violated, the basic research direction is whether and how the attacker can exploit the online information to degrade the system performance as much as possible under the communication rate constraint. Motivated by this, we focus on the online attack strategy design in order to degrade the estimation quality. Specifically, we consider deception attack strategy against state estimation of a linear system with Gaussian noises. In the viewpoint of attacker, we are interested in design proper online deception attack strategy to degrade the state estimation quality.

Our main contributions can be summarized as follows:

⋆ The work was supported in part by NSFC under Grants 61222305, 61290325, the SRFDP under Grant 20120101110139, NCET-11-0445, and National Program for Special Support of Top-Notch Young Professionals. The work by L. Shi is supported by an HKUST Caltech Partnership FP004.

(1) We propose an online attack strategy against state estimation and prove that the proposed strategy can degrade the estimation quality.
(2) We study the cost deviation under the proposed attack strategy, and prove that there exists a sensor-to-estimator rate to maximize this deviation.
(3) We obtain a closed-form expression of the sensor-to-estimator rate which maximizes the cost deviation.

The remainder of the paper is organized as follows: Section 2 formulates the problem. Section 3 proposes an online attack strategy and then evaluates the impact of this strategy. Section 4 illustrates the effectiveness of our proposed attack strategy. Section 5 concludes the whole paper.

Notations: $\mathbb{S}_+^n$ is the set of $n \times n$ positive semi-definite matrices. $\mathbb{R}^r$ is the r dimensional Euclidean space. $\mathbb{E}[X]$ and $\mathbb{D}[X]$ stand for the mean and variance of random variable X, respectively. $\mathbb{E}[X|Y]$ stands for the the mean of random variable X conditioned on Y. $\phi(\cdot)$ is the probability density function of Gaussian distribution $\mathcal{N}(0,1)$. $Tr(\cdot)$ is the trace operation of matrix. $\|\xi\|$ stands for Euclidean norm of a vector ξ. I_r represents $r \times r$ identity matrix. $\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$ stands for the diagonal matrix with the diagonal elements $\lambda_i, i = 1, 2, \ldots, r$. $\mathrm{rank}(\cdot)$ is the rank of a matrix. $(\cdot)'$ is the transpose operation of a matrix.

2. PROBLEM FORMULATION

Consider the following linear system (Fig. 1)

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

where $x_k \in \mathbb{R}^n$ is the state variable with $n \in \mathbb{N}$, $y_k \in \mathbb{R}^m$ is the measurement variable with $m \in \mathbb{N}$, $w_k \in \mathbb{R}^n$ is the process noise, $v_k \in \mathbb{R}^m$ is the measurement noise, w_k and v_k are uncorrelated zero mean Gaussian noises with covariance Σ_w and Σ_v, respectively. The pair $(A, \Sigma_w^{\frac{1}{2}})$ is stabilizable and (A, C) is assumed to be detectable .

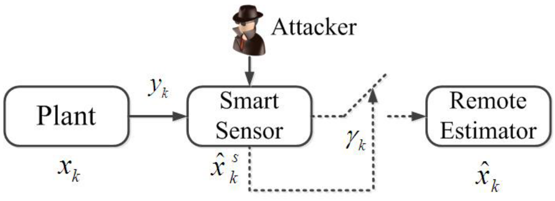

Fig. 1. System architecture

2.1 System architecture

The sensors, which have sufficient computational capability to estimate the system state x_k after reading the measurement y_k, are referred to as smart sensors. We assume a smart sensor is used. Its local estimate is calculated by a Kalman filter, i.e., $\widehat{x}_k^s = \mathbb{E}[x_k|y_0, \ldots, y_k]$. Sensor's estimation error is defined as $e_k^s = x_k - \widehat{x}_k^s$, and its error covariance matrix is defined as $P_k^s = \mathbb{E}[(e_k^s)(e_k^s)'|y_0, \ldots, y_k]$.

It is assumed that $\widehat{x}_0^s = 0$ and $P_0^s = \Pi_0 \geq 0$. From (Anderson and Moore, 1981), one can see that error covariance matrix P_k^s converges to a steady-state value P exponentially. We shall ignore the transient period and assume that $\Pi_0 = P$.

The sensor then decides whether or not to send this state estimate to the remote estimator. We denote $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_N)$ as the sensor's decision vector in a finite time horizon $[1, N]$, i.e., $\gamma_k = 1$ if the sensor sends $\widehat{x}_k^s$, and $\gamma_k = 0$ otherwise.

Denote the data set at remote estimator as $\mathcal{D}(\gamma)$. Then, its state estimate and corresponding error covariance are given by

$$\hat{x}_k(\gamma) = \mathbb{E}[x_k|\mathcal{D}(\gamma)]$$

and

$$P_k(\gamma) = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)'|\mathcal{D}(\gamma)].$$

For simplicity, we write $\hat{x}_k(\gamma)$ as $\hat{x}_k$, etc., when the schedule γ is given.

From (Shi et al., 2011b), it can be seen that

$$\widehat{x}_k = \begin{cases} \widehat{x}_k^s, & \text{if } \gamma_k = 1; \\ A\widehat{x}_{k-1}, & \text{otherwise.} \end{cases} \tag{2}$$

The estimation quality is measured by the cost

$$J(\gamma) = \limsup_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} Tr\{\mathbb{E}[P_k]\}.$$

From the sensor's point of view, it aims to find transmission strategy which minimizes J for a given average sensor-to-estimator communication rate [1] $\overline{\gamma}$, where

$$\overline{\gamma} = \limsup_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}(\gamma_k).$$

It is assumed that the sensor runs an online scheduler θ^s (cf.(Wu et al., 2013a)) in the sensor as follows:

$$\gamma_k^s = \begin{cases} 0, \text{ if } k \text{ is even and } \|E'\epsilon_k\| < \delta; \\ 1, \text{ otherwise.} \end{cases} \tag{3}$$

where $\epsilon_k = \widehat{x}_k^s - A\widehat{x}_{k-1}$, and δ is event-triggering threshold which is determined by the given average sensor-to-estimator communication rate $\overline{\gamma} \in [\frac{1}{2}, 1]$. The matrix E will be defined shortly.

This schedule can improve the estimation quality with the sensor-to-estimator communication rate constraint (cf. (Wu et al., 2013b)).

2.2 The objective of attacker

In our scenario, the deception attacker intrudes the sensor, stealing the compromised sensor's codes to learn its online transmission strategy. It then tampers the sensor's program by implanting its designed codes to the sensor.

Here we assume that the estimator knows that the sensor's online schedule is of the form given by (3). It means that the attacker can only recompose the transmission schedule in even time since no transmission at odd time can be easily detected by estimator. The estimator can estimate the communication rate from its prior knowledge

[1] The communication rate is defined due to sensor's energy constraint or the limitation of communication bandwidth (cf.(Wu et al., 2013a)).