Presented as late breaking results contribution
21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

On Sensor Attack Detection in Control Systems Using Moving Horizon Estimation and Control Performance ⋆

Kei Isono * **Koichi Kobayashi** * **Ryosuke Adachi** ** **Yuh Yamashita** *

* *Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan (e-mail: {kei2017@stl.,k-kobaya@,yuhyama@}ssi.ist.hokudai.ac.jp)*
** *Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Ube, Japan (e-mail: r-adachi@yamaguchi-u.ac.jp)*

Abstract: The sensor attack detection problem in control systems is important in the field of cybersecurity. In this paper, we propose a sensor attack detection method based on both moving horizon estimation and control performance. In the existing methods, the signal from an attacker is regarded as the unknown input or the error in state estimation. In the proposed method, we suppose that the closed-loop system is composed of the plant, the state estimator, and the controller by the linear quadratic regulator. We utilize moving horizon estimation for linear singular systems. Then, a sensor attack is detected based on the control performance. By a numerical example, the effectiveness of the proposed method is presented.

Keywords: Linear quadratic regulator, Moving horizon estimation, Sensor attack detection

1. INTRODUCTION

In the last two decades, various security incidents have been reported (see, e.g., Farwell and Rohozinski (2011)). Thus, cybersecurity has attracted much attention from the viewpoint of control theory (see, .e.g, Fawzi et al. (2014); Pasqualetti et al. (2015); Teixeira et al. (2015)). In particular, the attack detection problem in control systems is one of the typical problems in cybersecurity (see, e.g, Pasqualetti et al. (2015)).

In control theoretic approach to attack detection, unknown input observers are frequently utilized (see, e.g., Negash et al. (2017)). In unknown input observers, not only the state but also the unknown input (i.e., the disturbance) can be estimated. When unknown input observers are applied to cybersecurity, the unknown input is regarded as an attack. As other approaches, a detection method using the Kalman filter has been proposed in e.g., Shinohara and Namerikawa (2017). In this method, an attack is detected by using the estimation error. On the other hand, a control system is designed based on a certain performance index. Hence, it is important to develop an attack detection method based on degradation of the control performance. From this viewpoint, the authors have proposed an attack detection method (Isono et al. (2019)). In this method, an attack is detected based on the control performance by the linear quadratic regulator (LQR). However, we have considered only an attack to actuators. When we consider a large-scale system, it is important to consider an attack to sensors in a sensor network.

In this paper, we propose a method of sensor attack detection based on both moving horizon estimation and control performance. By evaluating degradation of the control performance, we can estimate an attack to sensors more precisely. Here, we use the performance of the LQR as a performance index. Furthermore, in the method of moving horizon estimation in Boulkroune et al. (2010), only an attack to actuators is considered. In this paper, to realize sensor attack detection, we improve the method in Boulkroune et al. (2010). The effectiveness of the proposed method is presented by a numerical example.

Notation: Let $\mathcal{R}$ denote the set of real numbers. Let I_n and $0_{m\times n}$ denote the $n \times n$ identity matrix and the $m \times n$ zero matrix, respectively. Let $M > 0$ ($M \geq 0$) denote that the matrix M is positive-definite (positive-semi-definite). For the vector x, let $\|x\|$ denote the Euclidean norm of x. For the vector x, let $x^{(i)}$ denote the i-th element of x. For the matrix M, let $M^\top$ denote the transpose matrix of M. For the vector x and the positive-semidefinite matrix M, we define $\|x\|_M^2 := x^\top M x$.

2. PRELIMINARIES

In this section, some preparations are given.

2.1 Linear Quadratic Regulator

First, we explain the outline of the linear quadratic regulator (LQR). See, e.g., Hespanha (2018) for further details.

As a plant, consider the following discrete-time linear system:

$$x_{k+1} = Ax_k + Bu_k,$$

where $k \in \{0, 1, \dots\}$ is the discrete time, $x_k \in \mathcal{R}^n$ is the state at time k, and $u_k \in \mathcal{R}^m$ is the control input at time k. Matrices A, B are constant with appropriate dimensions. We assume that the pair (A, B) is stabilizable. Consider finding a state-feedback controller minimizing the following cost function:

$$J = \sum_{k=0}^{\infty} \left\{ x_k^\top Q x_k + u_k^\top R u_k \right\},$$

⋆ This work was supported by JSPS KAKENHI Grant Numbers JP17K06486, JP19H02157, JP19H02158.

Presented as late breaking results contribution
21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

where $Q > 0$ and $R > 0$ are weighting matrices. We assume that the pair $(Q^{1/2}, A)$ is detectable. Then, the optimal state-feedback controller can be derived as $u_k^* = -(R + B^\top PB)^{-1}B^\top PAx_k$, where P is the symmetric positive definite solution of the discrete-time algebraic Riccati equation $A^\top PA - P - A^\top PB(R + B^\top PB)^{-1}B^\top PA + Q = 0$. The optimal value of J can be derived as

$$J^* = x_0^\top Px_0. \tag{1}$$

We remark that by $x_k^\top Px_k$, we can evaluate the control performance in the time interval $[k, \infty]$.

2.2 Moving Horizon Estimation

We explain the outline of unknown input observers based on an optimization approach (see, e.g., Boulkroune et al. (2010)). In Boulkroune et al. (2010), the unknown input is added to the state equation. In this paper, to model a sensor attack, the unknown input is added to the output equation.

Consider the following discrete-time linear system:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k, \\ y_k &= Cx_k + C_d d_k + v_k, \end{aligned} \tag{2}$$

where $k \in \{0, 1, \dots\}$ is the discrete time, $x_k \in \mathcal{R}^n$ is the state at time k, $u_k \in \mathcal{R}^m$ is the control input at time k. $d_k \in \mathcal{R}^q$ is the unknown input at time k, $w_k \in \mathcal{R}^n$ is the system noise, $y_k \in \mathcal{R}^p$ is the measured output, and $v_k \in \mathcal{R}^p$ is the measurement noise at time k. Matrices A, B, C, C_d are constant with appropriate dimensions. We introduce C_d as a coefficient of the attacks.

Consider estimating both the state x_k and the unknown input d_k by using an observer. First, we introduce the state transformation to treat above estimation problem as a state estimation problem. Defining $z_k := [x_k^\top \ d_k^\top]^\top$, the system (2) is transformed into the following singular system:

$$\begin{aligned} Ez_{k+1} &= Fz_k + Bu_k + w_k, \\ y_k &= Hz_k + v_k, \end{aligned} \tag{3}$$

where

$$E = [I_n \ 0_{n\times q}], \ F = [A \ 0_{n\times q}], \ H = [C \ C_d].$$

Next, we consider the optimization problem to minimize the estimated weighted error of the state and measurement. See Boulkroune et al. (2010) for details of derivation. In moving horizon estimation, the following optimization problem is solved at each time k:

$$\begin{aligned} &\text{given} \ \ \bar{z}_{k-N}, \{u_i\}_{i=k-N}^k, \{y_i\}_{i=k-N}^k \\ &\text{find} \ \ \{z_i\}_{i=k-N}^k \\ &\text{minimize} \ \ J_k = \|e_{k-N}\|^2_{P_{k-N}^{-1}} + \sum_{i=k-N}^{k-1} \|w_i\|^2_{W^{-1}} \\ &\qquad\qquad + \sum_{i=k-N}^{k} \|v_i\|^2_{V^{-1}} \\ &\text{subject to} \ \ \text{System (3)} \end{aligned}$$

where P_{k-N}^{-1} is a weighting matrix. See Boulkroune et al. (2010) for a method for determining $\bar{z}_{k-N}$. To solve this problem, we assume that

$$\text{rank}\left(\begin{bmatrix} E \\ H \end{bmatrix}\right) = \text{rank}\left(\begin{bmatrix} I_n & 0_{n\times q} \\ C & C_d \end{bmatrix}\right) = n + q,$$

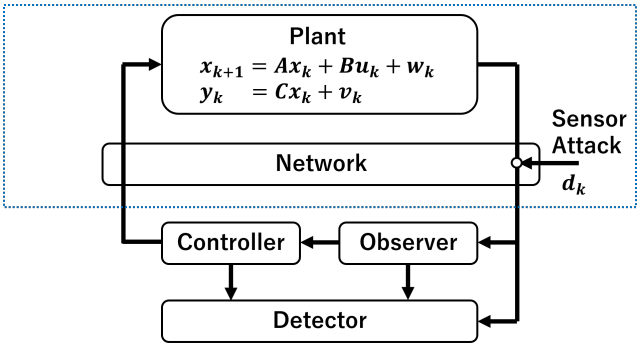

Fig. 1. Control system studied in this paper.

that is,

$$\text{rank}(C_d) = q., \ n + p \geq q. \tag{4}$$

This conditions do not guarantee the precision of estimation, however it is necessary to solve the following optimization problem (i.e., the feasibility can be guaranteed by this condition). Under this assumption, we can analytically obtain the optimal solution using the least squares method. See Boulkroune et al. (2010) for further details.

3. PROPOSED ATTACK DETECTION METHOD

In this section, we propose an attack detection method based on both moving horizon estimation and control performance. First, we explain the problem setting. Since we use moving horizon estimation, we do not have to restrict noise v, w to be Gaussian. Next, we propose a detection method.

3.1 Problem Setting

In this subsection, we give the problem setting to discuss the sensor attack detection problem. Consider the control system shown in Fig. 1. The plant in Fig.1 is a control target system and is given by a discrete-time linear system. We regard sensor attacks as unknown inputs to the output equation. Then, we consider the following system as a plant with sensor attacks (i.e., the inside of the blue dotted line in Fig.1):

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k \\ y_k &= Cx_k + C_d d_k + v_k. \end{aligned} \tag{5}$$

where d_k is the signal from an attacker. The matrix C_d must satisfy the assumption (4). Using C_d, we can characterize sensors that may be attacked. If we set $C_d = I_p$ ($p = q$), then there is a possibility that all sensors are attacked. In the practical situation, we may focus on only some sensors. In such case, we do not need to give C_d as the identity matrix.

For the control system shown in Fig. 1, we use the LQR controller. Then, we assume that the pair (A, B) is stabilizable. We also assume that $Q > 0$, $R > 0$, and the paper $(Q^{1/2}, A)$ is detectable. For the plant, we apply the control input $u_k = K\hat{x}_k$, where K is the optimal state feedback gain K obtained based on LQR.

3.2 Detection Procedure

Focusing on the fact that the controller is designed by LQR, we utilize the control performance (1) as an index. Here, we introduce the following detection index:

$$\tilde{J}_k^* = \frac{1}{l} \sum_{i=k-l+1}^{k} \hat{x}_{i|i}^\top P \hat{x}_{i|i},$$

where l is the length of the moving average. By $\tilde{J}_k^*$, the change of the control performance based on the LQR is evaluated.

In the proposed method, we choose one from the above three methods. Then, it is said that the control system (5) is attacked if the following condition holds:

$$\tilde{J}_k^* > \theta_J,$$

where θ_J is a given threshold. By choosing one from three methods, we can realize sensor attack detection.

Based on the above discussion, we present a detection procedure.

Detection Procedure:

Step 0: Set a threshold and other parameters. Set also the detection start time k_{start}. If $k = k_{\text{start}}$ holds, then go to Step 1.

Step 1: Calculate $\hat{x}_k$, $\hat{d}_{k|k}$ by moving horizon estimation.

Step 2: Calculate a detection index.

Step 3: If the obtained index is greater than the threshold, then determine the system is attacked. Otherwise, update $k := k+1$, and return to Step 1.

In the above procedure, the computational load is only matrix manipulation in moving horizon estimation. Hence, computation at each discrete time can be executed fast.

4. NUMERICAL EXAMPLE

4.1 Setting

In this section, we present a numerical example of the proposed attack detection methods. First, the matrices in the plant are given by

$$A = \begin{bmatrix} 1.03 & -0.4 & 0.2 \\ 0 & 1.07 & -0.3 \\ 0 & 0 & 0.9 \end{bmatrix}, \; B = \begin{bmatrix} 1 \\ 0 \\ 0.2 \end{bmatrix},$$
$$C = I_3, \; C_d = [1 \; 0 \; 1]^\top,$$

respectively. The initial state is given by $[5 \; -5 \; 5]^\top$.

The weights in LQR are given by $Q = 10000 I_3$, $R = 100$, respectively. Solving the discrete-time algebraic Riccati equation, the solution and the gain of the optimal state feedback controller can be obtained by

$$P = \begin{bmatrix} 3.4 \times 10^4 & 4.3 \times 10^4 & -1.2 \times 10^5 \\ 4.3 \times 10^4 & 2.0 \times 10^5 & -2.4 \times 10^5 \\ -1.2 \times 10^5 & -2.4 \times 10^5 & 5.7 \times 10^5 \end{bmatrix},$$
$$K = [1.0 \; -0.85 \; 0.21],$$

respectively. The weights in the unknown input observer are given by

$$W = \text{diag}(1000, 1000, 1000), \; V = \text{diag}(100, 100, 100),$$

respectively. The prediction horizon N is given by $N = 50$. For the detection algorithm, the parameters in the proposed method are given by $l = 5$ and $\theta_J = 1.0 \times 10^7$. Furthermore, as

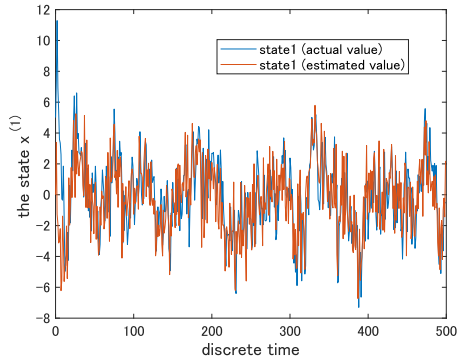

Fig. 2. Time response of the state x_1.

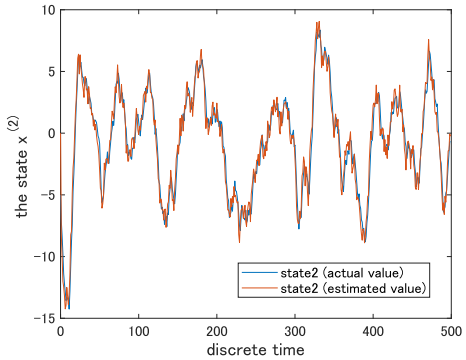

Fig. 3. Time response of the state x_2.

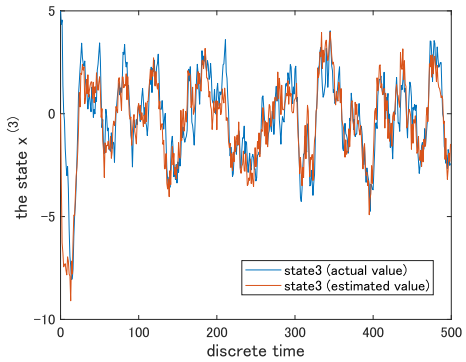

Fig. 4. Time response of the state x_3.

the conventional detection indices, we introduce two indices, i.e., the norm of the unknown input and the norm of the state. Then, we consider two detection conditions as follows: $\sum_{i=k-l+1}^{k} \|\hat{d}_{i|i}\|/l > \theta_d$ and $\sum_{i=k-l+1}^{k} \|\hat{x}_{i|i}\|/l > \theta_x$, where we set $\theta_d = 1.8$ and $\theta_x = 10$.

In the numerical simulation, the noise v, w are respectively random numbers, and uniformly distributed among the interval $(-1.0, 1.0)$. We apply the attack detection algorithm since $k = 300$, and we suppose that the attack starts at $k = 300$ as follows:

$$d_k = \begin{cases} 0.5 & \text{If } 300 \leq \text{k} < 400 \\ 0 & \text{otherwise.} \end{cases}$$

In this case, we suppose that the observation outputs $y^{(1)}$, $y^{(3)}$ are attacked simultaneously.

4.2 Result

We present the computation result. Figures 2, 3, and 4 show the estimated values of the state $x^{(1)}$, $x^{(2)}$, and $x^{(3)}$. From these

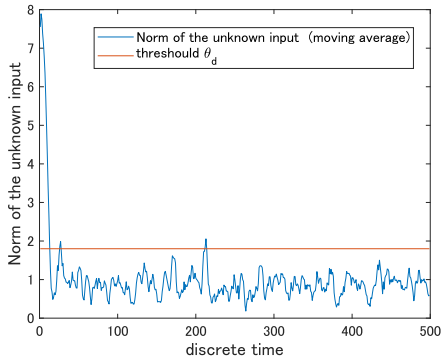

Fig. 5. Time response of the norm of the unknown input.

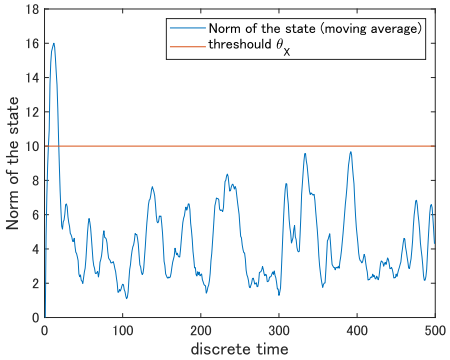

Fig. 6. Time response of the norm of the state.

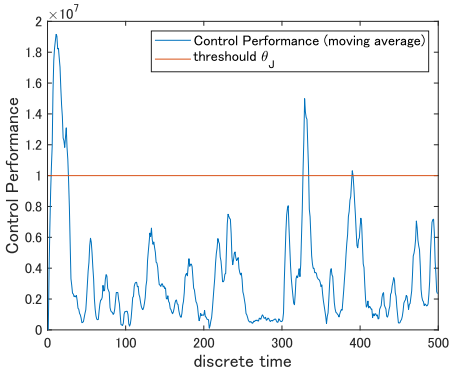

Fig. 7. Time response of the control performance.

figures, we see that the noise particularly affects $x^{(1)}$ and $x^{(3)}$ adversely.

Figure 5 shows the norm of the estimated unknown input. In this example, the attack is detected at $k = 213$. In other words, attack detection is failed. Figure 6 shows the norm of the estimated state. It is difficult to distinguish an attack from a noise, and the attack is not detected. This is failure. Figure 7 shows the control performance. The attack is detected at $k = 328$, and attack detection is successful.

Figure 8 shows both $\|y - C\hat{x}\|$ (the estimation error of the output equation) and its moving average. The state estimation error is one of the conventional indices for attack detection (see, e.g., Shinohara and Namerikawa (2017)). However, in this case, the detection method using the estimation error does not work, because estimation error does not show clear difference between the safe situation and attacked situation. In contrast, by utilizing the control performance, attack detection can be performed correctly.

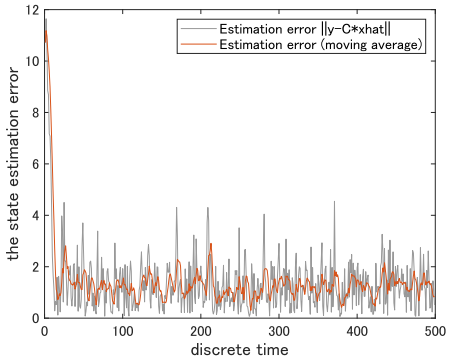

Fig. 8. Time response of the estimation error.

5. CONCLUSION

In this paper, we proposed a new method of sensor attack detection based on moving horizon estimation and control performance. The detection condition proposed in this paper uses the estimated state and the control performance by the LQR, and is simple. The effectiveness of the proposed method was presented by a numerical example.

In future work, it is important to develop a method for detecting both an attack to actuators and an attack to sensors.

REFERENCES

B. Boulkroune, M. Darouach, and M. Zasadzinski, Moving horizon state estimation for linear discrete-time singular systems, *IET Control Theory & Applications*, vol. 4, no. 3, pp. 339–350, 2010.

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

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

J. P. Hespanha, *Linear Systems Theory*, Second Edition, Princeton University Press, 2018.

K. Isono, K. Kobayashi, R. Adachi, and Y. Yamashita, Attack detection in control systems based on unknown input observers and control performance, *Proc. of the 34th Int'l Technical Conf. on Circuits/Systems, Computers and Communications*, pp. 290–293, 2019.

F. Pasqualetti, F. Dörfler, and F. Bullo, Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems, *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 110–127, 2015.

L. Negash, S.-H. Kim, and H.-L. Choi, Distributed observers for cyberattack detection and isolation in formation-flying unmanned aerial vehicles, *Journal of Aerospace Information Systems*, vol. 14, no. 10, pp. 551–565, 2017.

T. Shinohara and T. Namerikawa, On the vulnerabilities due to manipulative zero-stealthy attacks in Cyber-physical systems, *SICE Journal of Control, Measurement, and System Integration*, vol. 10, no. 6, pp. 563–570, 2017.

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