Attack correction for noise-free linear systems subject to sensor attacks

Extended Abstract

Tang, Zhanghan[1] Kuijper, Margreta[1] Chong, Michelle[2] Mareels, Iven[1]
Leckie, Christopher[3]

Abstract— **We address the problem of attack detection and attack correction for multi-output discrete-time linear time-invariant systems under sensor attacks. More specifically, we focus on the situation where adversarial attack signals are added to some of the systems output signals. A 'security index' is defined to characterize the vulnerability of a system against such sensor attacks and methods are given to calculate this index for various system representations. Algorithms are presented to detect and correct for sensor attacks on a noise-free linear system.**

Keywords: linear time-invariant systems, security, sensor attacks, attack correction
AMS subject classifications (2010): 93C05, 93B07, 94B99

I. INTRODUCTION

In today's society, the physical infrastructures in support of critical services can be described as cyber-physical systems. The end-to-end service can therefore be affected by, for example, loss of functionality of the physical assets or malicious attack. In this presentation, the particular attack scenario where some sensor signals may be corrupted by additive malicious attack signals. In this case, the attack signals may be injected based on the knowledge of the targeted system and thus it no longer suffices to treat these external signals as mere disturbances or noise. In fact, such attack signals can be used to control the behavior of a system to achieve the purpose of the attacker. For instance, it is conceivable that a modern autonomous vehicle may be hijacked using sensor spoofing [8].

In this presentation, we consider a discrete-time linear time-invariant (LTI) system whose outputs may be compromised by additive attack signals. To assist in the analysis, the notion of the 'security index' of a system is introduced, which is a quantitative representation-free measure of the vulnerability of a system to sensor attacks. We will show how it relates to the detectability and correctability of attack signals.

Unlike much of the work in this area, the main tool used in this presentation is a kernel representation (see for example [2, Ch2.5]) and behavioral approach rather than a state space representation for the following reasons: 1. Many well-established theorems based on kernel representation can be applied when we are discussing a system using a behavioral approach. 2. Every system in kernel representation can be brought into state space form and every observable system in state space can be transformed into kernel representation. 3. The implementation of the system in kernel representation can be done straightforwardly using shift operators.

The focus of our presentation is on the development of conceptual approaches to attack detection and correction. In this presentation, we restrict ourself to a noise free environment. Our methods then serve as a starting point for further research on the noisy case. We consider only sensor attacks to enable the readers to understand the essence of the proposed methods. We also assume that the inputs

[1] Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne. zhanghant@student.unimelb.edu.au (Zhanghan Tang), mkuijper@unimelb.edu.au (Margreta Kuijper), i.mareels@unimelb.edu.au (Iven Mareels).

[2] Department of Automatic Control, KTH Royal Institute of Technology, Sweden. mchong@kth.se (Michelle Chong).

[3] Department of Computing and Information Systems, University of Melbourne, Melbourne. caleckie@unimelb.edu.au (Chris Leckie).

of the system are known and due to the linearity of the system, we can set it to be zero signals for the purpose of our analysis.

II. PROBLEM STATEMENT

We consider an LTI system Σ in its kernel representation as follows

$$\Sigma : \quad R(\sigma)y = 0, \tag{1}$$

where $y : \mathbb{Z}_+ \to \mathbb{R}^N$ is the sensor output signal of the system Σ and $R(\xi)$ is an $N \times N$ real polynomial matrix of full rank, meaning that the system's behavior is autonomous with no free variables (see for example [2, Ch3.1]). σ represents the forward shift operator, i.e., $\sigma y(t) = y(t+1)$.

Definition 2.1: The behavior of the system Σ is defined as the set given by

$$\mathcal{B} = \{y : \mathbb{Z}_+ \to \mathbb{R}^N \mid R(\sigma)y = 0\}. \tag{2}$$

Now consider additive attack signal $\eta : \mathbb{Z}_+ \to \mathbb{R}^N$ and $\eta \in$ attack set $\mathcal{A}$. A corrupted received signal is $r = y + \eta$.

Definition 2.2: The behavior of the corrupted system $\Sigma_{\mathcal{A}}$ is defined as the set of possible received signals given by

$$\mathcal{B}_{\mathcal{A}} = \{r : \mathbb{Z}_+ \to \mathbb{R}^N \mid r = y + \eta, \text{ where } y \in \mathcal{B}, \eta \in \mathcal{A}\}. \tag{3}$$

The detectability and correctability of a system are defined as follows.

Definition 2.3 (Attack detectability): A non-zero attack signal $\eta \in \mathcal{A}$ is detectable if $\eta \notin \mathcal{B}$.

Definition 2.4 (Attack correctability): A non-zero attack signal $\eta \in \mathcal{A}$ is correctable if for all $\eta' \neq \eta$, the following is satisfied

$$\eta' \in \mathcal{A} \Rightarrow \eta - \eta' \notin \mathcal{B}. \tag{4}$$

In this presentation we will first recall results from [1], [3] on the feasibility of attack detection/correction and then propose new methods that are guaranteed to achieve attack correction under certain assumptions about the attack set $\mathcal{A}$. Full proofs on these new results are omitted here but can be found in our submitted work [9].

III. SECURITY INDEX

We recall conditions to achieve attack detectability and correctability via the 'security index' $\delta(\Sigma)$ of the system Σ as defined in [1], [3].

Definition 3.1: The security index of the system Σ is defined as

$$\delta(\Sigma) := \min_{0 \neq y \in \mathcal{B}} \|y\|, \tag{5}$$

here, we use $\|y\|$ to denote the weight (i.e., number of non-zero components) of signal y.

Note that $\delta(\Sigma)$ is an integer between 1 and N and note the analogy of Definition 3.1 with the concept of 'minimum distance' in coding theory (see for example [10, Ch7.2-2]).

Theorem 3.1: **(Attack detection capability of the system)** Let $\mathcal{A} = \{\eta : \mathbb{Z}_+ \to \mathbb{R}^N \mid \|\eta\| < \delta(\Sigma) \text{ and } \eta \neq 0\}$. All attack signals $\eta \in \mathcal{A}$ are detectable.

Theorem 3.2: **(Attack correction capability of the system)** Let $\mathcal{A} = \{\eta : \mathbb{Z}_+ \to \mathbb{R}^N \mid \|\eta\| < \delta(\Sigma)/2 \text{ and } \eta \neq 0\}$. All attack signals $\eta \in \mathcal{A}$ are correctable.

Remark 3.3: We call a system Σ with N outputs *maximally secure* if its security index $\delta(\Sigma) = N$.

It follows from the above theorems that the value of $\delta(\Sigma)$ can be viewed as the minimum number of sensors that have to be attacked in order to implement an undetectable attack. Similarly, the value of $\delta(\Sigma)/2$ corresponds to the minimum number of sensors that have to be attacked in order to implement an uncorrectable attack. Thus the system's security index relates to the feasibility of attack detection/correction.

To calculate the security index of a system Σ, we need the following notation, where $\mathcal{J}$ is assumed to be a subset of $\{1, \dots, N\}$.

Notation 3.1: Define $R_{\mathcal{J}}(\xi)$ as an $N \times \|\mathcal{J}\|$ matrix that consists of the i-th columns of $R(\xi)$ where $i \in \mathcal{J}$.

Next we recall the following theorem which provides a way to calculate the security index of a system.

Theorem 3.4 (Security index calculation): Consider a system Σ whose behavior $\mathcal{B}$ is non-zero and given by (1), where $R(\xi)$ has full rank. Then

$$\delta(\Sigma) = L + 1, \tag{6}$$

where L is the largest integer such that for any subset $\mathcal{J} \subseteq \{1, \dots, N\}$ of cardinality L, the $N \times L$ matrix $R_{\mathcal{J}}(\xi)$ is left unimodular.

IV. Attack detection and correction algorithms

For attack detection, we recall the following algorithm from [1], [3].

Algorithm 1 Attack detection

```
procedure $(R(\xi), r, \eta)$
    ▷ Given $R(\xi)$ and $r$, detect whether $\eta$ is the zero signal.
    Calculate $s = R(\sigma)r$.
    if $s = 0$ then decide no attack, i.e., $\eta = 0$.
    else decide attack occurred, i.e., $\eta \neq 0$.
    end if
end procedure
```

Before we discuss attack correction algorithms, the following theorem on equivalent kernel representations for $\mathcal{B}$ is presented.

Theorem 4.1: (e.g. [4, Theorem 3.9]) Consider two systems Σ and Σ' whose behaviors $\mathcal{B}$ and $\mathcal{B}'$ are given by $\mathcal{B} = \{y | R(\sigma)y = 0\}$ and $\mathcal{B}' = \{y' | R'(\sigma)y' = 0\}$, respectively. Assume that $R(\xi)$ and $R'(\xi)$ are square matrices of the same size, then $\mathcal{B} = \mathcal{B}'$ if and only if $R(\xi)$ and $R^{'}(\xi)$ are left unimodularly equivalent.

Remark 4.2: Theorem 4.1 indicates that changing a kernel representation of $\mathcal{B}$ via left multiplication by a unimodular matrix does not change the behavior $\mathcal{B}$.

The discussion of the attack correction algorithm will be divided into two cases: first for systems that are maximally secure and second for general systems that are not necessarily maximally secure. For both cases, we first single out a canonical form representation which is then used to provide the algorithm for attack correction. Proofs can be found in our submitted work [9]. In the next theorem $\deg\, a(\xi)$ denotes the degree of the polynomial $a(\xi)$.

Theorem 4.3: **(Kronecker-Hermite canonical kernel representation of $R(\xi)$—maximally secure case)** [5, Theorem 2.40], [7, Theorem 7.5], [6] Let $R(\xi)$ be an $N \times N$ polynomial matrix whose determinant is non-zero. Assume that all $N \times (N-1)$ submatrices of $R(\xi)$ are left unimodular. Then there exists a unimodular matrix $U(\xi)$ such that

$$U(\xi)R(\xi) = \left[\begin{array}{c|c} & -c_1(\xi) \\ & -c_2(\xi) \\ \mathbb{I}_{N-1} & \vdots \\ & -c_{N-1}(\xi) \\ \hline 0 \ \dots \ 0 & a(\xi) \end{array}\right], \tag{7}$$

where $c_j(\xi)$ is coprime with $a(\xi)$ and $\deg\, c_j(\xi) < \deg\, a(\xi)$ for all $j \in \{1, ..., N-1\}$.

Notation 4.1: Because of the coprimeness of $c_j(\xi)$ and $a(\xi)$ there exist polynomials $p_j(\xi)$ and $q_j(\xi)$ satisfying

$$\begin{bmatrix} p_j(\xi) & q_j(\xi) \end{bmatrix} \begin{bmatrix} c_j(\xi) \\ a(\xi) \end{bmatrix} = 1, \forall j \in \{1, ..., N-1\}. \tag{8}$$

Notation 4.2: Maj$\{v_1, v_2, \dots, v_L\}$, (majority vote) denotes the most frequently occurring signal in the set of v_j's.

Given Theorem 4.3, without loss of generality, we assume the polynomial matrix $R(\xi)$ for a maximally secure system is in the form of (7). Algorithm 2 presents the attack correction method for systems that are maximally secure. We can prove that $\hat{y} = y$ if $\|\eta\| < \delta(\Sigma)/2$, see[9].

Algorithm 2 Attack correction for a maximally secure system given by (7)

```
procedure $(a(\xi), c_1(\xi), \dots, c_{N-1}(\xi), r, \hat{y})$
    ▷ Given $a(\xi)$, $c_j(\xi)$'s and $r$, compute $\hat{y}$.
    Calculate
        $$\hat{y}_N = \text{Maj}\{p_1(\sigma)r_1, \dots, p_{N-1}(\sigma)r_{N-1}, r_N\}, \tag{9}$$
    where $p_j(\xi)$ is defined as in (8).
    $\hat{y}_j = c_j(\sigma)\hat{y}_N$ for $j = 1, 2, \dots, N-1$.
    return $\hat{y} = \begin{bmatrix} \hat{y}_1 & \hat{y}_2 & \dots & \hat{y}_N \end{bmatrix}$.
end procedure
```

Theorem 4.4: **(Kronecker-Hermite canonical kernel representation of $R(\xi)$—general case)** Let $R(\xi)$ be an $N \times N$ polynomial matrix whose determinant is nonzero. Let L be the largest integer such that, for any subset $\mathcal{J} \subseteq \{1, \dots, N\}$ of cardinality L, the $N \times L$ matrix $R_{\mathcal{J}}(\xi)$ is left unimodular. Then there exists a unimodular matrix $U(\xi)$ such that

$$U(\xi)R(\xi) = \left[\begin{array}{c|c} \mathbb{I}_L & -M_1(\xi) \\ \hline 0 \ \dots \ 0 & D(\xi) \end{array}\right], \tag{10}$$

where $D(\xi)$ is an upper triangular matrix and the degree of the diagonal entities of $D(\xi)$, denoted as $\deg d_{ii}(\xi)$ for $i \in \{1, ..., N-L\}$, is strictly the highest within the corresponding column of (10).

$$\begin{bmatrix} \mathbb{I}_{\delta(\Sigma)-1} & 0 \\ 0 & \mathbb{I}_{N-\delta(\Sigma)+1} \\ 0 & 0 \end{bmatrix} y = \begin{bmatrix} M_1(\sigma) \\ \mathbb{I}_{N-\delta(\Sigma)+1} \\ D(\sigma) \end{bmatrix} \ell, \quad (11)$$

where the signal ℓ is an auxiliary signal that can be interpreted as a 'state signal' that drives the system's behavior. In the representation (11) the signal ℓ simply coincides with the last $m = N - \delta(\Sigma)+1$ components of y. More generally, we have

$$\begin{bmatrix} \mathbb{I}_N \\ 0 \end{bmatrix} y = \begin{bmatrix} M(\sigma) \\ D(\sigma) \end{bmatrix} \ell, \quad (12)$$

where $\ell : \mathbb{Z}_+ \to \mathbb{R}^m$.

Theorem 4.5: Consider a system Σ whose behavior $\mathcal{B}$ is nonzero and given by (12). Then

$$\delta(\Sigma) = N + 1 - \tilde{L}, \quad (13)$$

where $\tilde{L}$ is the smallest integer such that for any subset $\mathcal{J} \subseteq \{1, \ldots, N\}$ of cardinality $\tilde{L}$, the $(\tilde{L} + m) \times m$ matrix $\begin{bmatrix} M_{\mathcal{J}}(\xi) \\ D(\xi) \end{bmatrix}$ is left unimodular.

Notation 4.3: Let $\mathcal{J}$ be a subset of $\{1, \ldots, N\}$ of cardinality $N + 1 - \delta(\Sigma)$. Suppose that the matrix $\begin{bmatrix} M_{\mathcal{J}}(\xi) \\ D(\xi) \end{bmatrix}$ is left unimodular. Then there exist polynomial matrices $P^{\mathcal{J}}(\xi)$ and $Q^{\mathcal{J}}(\xi)$ such that

$$\begin{bmatrix} P^{\mathcal{J}}(\xi) & Q^{\mathcal{J}}(\xi) \end{bmatrix} \begin{bmatrix} M_{\mathcal{J}}(\xi) \\ D(\xi) \end{bmatrix} = \mathbb{I}_{N+1-\delta(\Sigma)}. \quad (14)$$

Algorithm 3 presents the attack correction method for general systems (not necessarily maximally secure). We can prove that $\hat{y} = y$ if $\|\eta\| < \delta(\Sigma)/2$, see [9]. Note that polynomial matrices $P^{\mathcal{J}}(\xi)$s in Algorithm 3 and $p_j(\xi)$s in Algorithm 2 can be considered as state observers that provide accurate estimations of the state values of the system Σ.

V. CONCLUSIONS

In this presentation, we propose attack correction methods for zero input discrete-time LTI systems in the noise-free case. The purpose of this presentation is to provide a proof of concept around the application of ideas from coding theory and representations of system behavior to handle attacks on LTI systems. We use the notion of ”system security index” from earlier work which quantifies the vulnerability of the system against sensor attacks. We propose attack correction methods that exploit the outputs' redundancy as well as the known dynamics of the system.

Algorithm 3 Attack correction for general system given by (12)

```
procedure $(M(\xi), D(\xi), \delta(\Sigma), r, \hat{y})$
    ▷ Given $M(\xi), D(\xi)$, $\delta(\Sigma)$ and $r$, compute $\hat{y}$.
    Calculate
        $\hat{\ell} = \text{Maj}\{P^{\mathcal{J}}(\sigma) r_{\mathcal{J}}\},$          (15)
    where the majority vote is taken over all subsets $\mathcal{J}$ of cardinality $N + 1 - \delta(\Sigma)$ and $P^{\mathcal{J}}(\xi)$ is defined as in (14).
    $\hat{y} = M(\sigma)\hat{\ell}$.
    return $\hat{y}$.
end procedure
```

REFERENCES

[1] M. Chong and M. Kuijper, Characterising the vulnerability of linear control system under sensor attack using a system's security index. In Proc. IEEE 55th Conf. on Decision and Control, Las Vegas, 2016, pp. 5906-5911.

[2] J. Polderman and J. Willems, Introduction to mathematical systems theory: a behavioral approach, New York, USA: Springer, 1997, vol. 26.

[3] M. Chong and M. Kuijiper, Vulnerability of linear systems against sensor attacks - a system's security index. In Proc. 22nd International Symposium on Mathematical Theory of Networks and Systems, Minneapolis, USA, 2016.

[4] M. Kuijper. First-Order Representation of Linear Systems. Systems and Control: Foundations and Applications. Boston, USA: Birkhäuser, 1994.

[5] P. A. Fuhrmann and U. Helmke, The Mahematics of Networks of Linear System, Springer, 2015

[6] D. Hinrichsen, U. Helmke and D. Prätzel- Wolters. Generalized hermite matrices and complete invariants of strict system equivalence. SIAM J. Control and Optimization, pp. 21:289 - 306, 1983.

[7] F. Colonius, U, Helmke, D. Prätzel- Wolters and F. Wirth. System and Control: Fundations and applications: Springer, 2001.

[8] J. S. Warner and R. G. Johnstib. A simple demonstration that the global positioning system is vulnerable to spoofing. In Proc. Journal of Security Administration, pp. 19-27, 2002.

[9] Z. Tang, M. Kuijper, M. Chong, I. Mareels and C. Leckie. Linear system security – detection and correction of adversarial attacks in the noise-free case, Nov. 2017. Submitted, available at arXiv:1711.05400 [eess.SP].

[10] J. P. Proakis and M. Salehi. Digital communications 5th edition. McGraw Hill, New York, USA, 2008.