Author/s:
Yang, T;Murguia, C;Kuijper, M;Nesic, D

Title:
An unknown input multi-observer approach for estimation, attack isolation, and control of lti systems under actuator attacks

Date:
2019-06-01

Citation:
Yang, T., Murguia, C., Kuijper, M. & Nesic, D. (2019). An unknown input multi-observer approach for estimation, attack isolation, and control of lti systems under actuator attacks. Proceedings of the 2019 18th European Control Conference (ECC), pp.4350-4355. IEEE. https://doi.org/10.23919/ECC.2019.8796178.

Persistent Link:
https://hdl.handle.net/11343/249537

An Unknown Input Multi-Observer Approach for Estimation, Attack Isolation, and Control of LTI Systems under Actuator Attacks

Tianci Yang, Carlos Murguia, Margreta Kuijper, and Dragan Nešić

***Abstract*— We address the problem of state estimation, attack isolation, and control for discrete-time Linear Time Invariant (LTI) systems under (potentially unbounded) actuator false data injection attacks. Using a bank of Unknown Input Observers (UIOs), each observer leading to an exponentially stable estimation error in the attack-free case, we propose an estimator that provides exponential estimates of the system state and the attack signals when a sufficiently small number of actuators are attacked. We use these estimates to control the system and isolate actuator attacks. Simulations results are presented to illustrate the performance of the results.**

I. Introduction

Networked Control Systems (NCSs) have emerged as a technology that combines control, communication, and computation and offers the necessary flexibility to meet new demands in distributed and large scale systems. Recently, security of NCSs has become an important issue as wireless communication networks might serve as new access points for attackers to adversely affect the operation of the system dynamics. Cyber-physical attacks on NCSs have caused substantial damage to a number of physical processes. One of the most well-known examples is the attack on Maroochy Shire Councils sewage control system in Queensland, Australia that happened in January 2000. The attacker hacked into the controllers that activate and deactivate valves and caused flooding of the grounds of a hotel, a park, and a river with a million liters of sewage. Another incident is the more recent StuxNet virus that targeted Siemens supervisory control and data acquisition systems which are used in many industrial processes. It follows that strategic mechanisms to identify and deal with attacks on NCSs are strongly needed.

In [1]-[23], a range of topics related to security of linear control systems have been discussed. In general, they provide analysis tools for quantifying the performance degradation induced by different classes of attacks; and propose reaction strategies to identify and counter their effect on the system dynamics. There are also some results addressing the nonlinear case. In [24], exploiting sensor redundancy, the authors address the problem of sensor attack detection and state estimation for uniformly observable continuous-time nonlinear systems. Similarly, in [25], the authors provide an algorithm for isolating sensor attacks for a class of discrete-time nonlinear systems with bounded measurement noise.

This work was partially supported by the Australian Research Council under the Discovery Project DP170104099. Tianci Yang, Carlos Murguia, Margreta Kuijper, and Dragan Nešić are with the Department of Electrical and Electronic Engineering, at the University of Melbourne, Australia. Emails: tianciy@student.unimelb.edu.au, carlos.murguia@unimelb.edu.au, mkuijper@unimelb.edu.au, & dnesic@unimelb.edu.au.

In this manuscript, we use Unknown Input Observers (UIOs) to address the problem of state estimation, attack isolation, and control for discrete-time Linear Time Invariant (LTI) systems under (potentially unbounded) actuator attacks. Unknown input observers are dynamical systems capable of estimating the state of the plant *without* using input signals. If such an observer exists and some of the inputs are subject to attacks, we can reconstruct the system state without using inputs; and use these estimates to reconstruct the attack signals by using model matching techniques. The existence of UIOs depend on the system dynamics, i.e., the matrices (A, B, C) comprising the system. If an UIO does not exists for this *complete* (A, B, C) but it does for some *partial* $(A, \tilde{B}_i, C)$, where $\tilde{B}_i$ denotes a submatrix of B with fewer columns and the same number of rows, then, using *a bank of observers*, we can use similar ideas to perform state estimation and attack isolation at the price of only being able to isolate when a sufficiently small subset of actuators are under attack. The main idea behind our multi-observer estimator is the following. Each UIO in the bank is constructed using a triple $(A, \tilde{B}_i, C)$, i.e., the i-th observer does *not* use the input signals associated with $\tilde{B}_i$, but it does use the remaining input signals. If the inputs corresponding to $\tilde{B}_i$ include all the attacked ones, this UIO produces an exponentially stable estimation error. For every pair of UIOs in the bank, we compute the largest difference between their estimates. Then, we select the pair leading to the smallest difference and prove that these observers reconstruct the state of the system exponentially. This idea is guaranteed to work under the assumption that less than half of the actuators are attacked. This multi-observer approach is inspired by the results given in [26] where the problem of state estimation for continuous-time LTI systems under sensor attacks is considered. Once we have an estimate of the state, we reconstruct the attack signals using model matching techniques. Finally, we propose a simple yet effective technique to stabilize the system by switching off the isolated actuators, and closing the loop with a multi-observer based output dynamic feedback controller. Here, we assume that the set of attacked actuators is time-invariant, i.e., if the opponent compromises a set of actuators at some time-instant, only this set will be compromised in forward time. Because attack signals may be zero for some time instants, the actuators isolated as attack-free might arbitrarily switch among all the supersets of the set of attack-free actuators. Therefore, we need a controller able to stabilize the closed-loop dynamics under the arbitrary switching induced by turning off the isolated actuators. To achieve this, we assume

that a *state* feedback controller that stabilizes the switching closed-loop system exists, and use this controller together with the multi-observer estimator to stabilize the system. We use Input-to-State Stability (ISS) [27] of the closed-loop system with respect to the exponentially stable estimation error to conclude on stability of the closed-loop dynamics.

The paper is organized as follows. In Section II, we present some preliminary results needed for the subsequent sections. In Section III, we introduce the proposed UIO-based estimation schemes. In Section IV, a method for isolating actuator attacks is described. The proposed control scheme is given in Section V. Finally, in Section VI, we give concluding remarks.

II. Preliminaries

A. Notation

We denote the set of real numbers by $\mathbb{R}$, the set of natural numbers by $\mathbb{N}$, the set of integers by $\mathbb{Z}$, and the set of $n \times m$ real matrices by $\mathbb{R}^{n\times m}$, $m, n \in \mathbb{N}$. For any vector $v \in \mathbb{R}^{n_v}$, v_J denotes the stacking of all v_i, $i \in J$ and $J \subset \{1, \dots, n_v\}$, $|v| = \sqrt{v^\top v}$ and $\text{supp}(v) = \{i \in \{1, \dots, n_v\} \, | v_i \neq 0\}$. For a sequence of vectors $\{v(k)\}_{k=0}^{\infty}$, we denote by $v_{[0,k]}$ the sequence of vectors $v(i)$, $i = 0, \dots, k$, $||v||_\infty := \sup_{k\geq 0} |v(k)|$ and $||v||_T := \sup_{0\leq k\leq T} |v(k)|$. We say that a sequence $\{v(k)\} \in l_\infty$, if $||v||_\infty < \infty$. We denote the cardinality of a set S as $\text{card}(S)$. The binomial coefficient is denoted as $\binom{a}{b}$, where a, b are nonnegative integers. We denote a variable m uniformly distributed in the interval (z_1, z_2) as $m \sim \mathcal{U}(z_1, z_2)$ and normally distributed with mean μ and variance σ^2 as $m \sim \mathcal{N}(\mu, \sigma^2)$. The notation $\mathbf{0}_n$ and I_n denote the zero matrix and the identity matrix of dimension $\mathbb{R}^{n\times n}$, respectively. We simply write $\mathbf{0}$ and I when their dimensions are evident.

III. Unknown input observer-based estimator

Consider the discrete-time LTI system:

$$\begin{cases} x^+ = Ax + B(u + a), \\ y = Cx, \end{cases} \tag{1}$$

with state $x \in \mathbb{R}^n$, output $y \in \mathbb{R}^{n_y}$, known input $u \in \mathbb{R}^p$, and vector of actuator attacks $a \in \mathbb{R}^p$, $a = (a_1, \dots, a_p)^T$. That is, $a_i(k) = 0$ for all $k \geq 0$ if the i-th actuator is attack-free; otherwise, $a_i(k) \neq 0$ for some (but not necessarily all) time instants $k \geq 0$, and can be arbitrarily large. Matrices A, B, C are of appropriate dimensions, (A, B) is stabilizable, (A, C) is detectable, and B has full column rank. Let the set of *unknown* attacked actuators be denoted by $W \subset \{1, \dots, p\}$, i.e., $a_i(k_i) \neq 0$ for some $k_i \geq 0$ and all $i \in W$.

Assumption 1 *The set of attacked actuators is time invariant, i.e., $W \subset \{1, \dots, p\}$ is a constant set.*

A. Complete Unknown Input Observers

We first treat $(u + a)$ as an unknown input to system (1) and consider an UIO with the following structure:

$$\begin{cases} z^+ = Nz + Ly, \\ \hat{x} = z + Ey, \end{cases} \tag{2}$$

where $z \in \mathbb{R}^n$ is the state of the observer, $\hat{x} \in \mathbb{R}^n$ denotes the estimate of the system state, and (N, L, E) are observer matrices of appropriate dimensions to be designed. It is easy to verify that if (N, L, E) satisfy the following equations:

$$\begin{cases} N(I - EC) + LC + (EC - I)A = \mathbf{0}, \\ (EC - I)B = \mathbf{0}, \end{cases} \tag{3}$$

then, the estimation error $e := \hat{x} - x$ satisfies the difference equation:

$$e^+ = Ne. \tag{4}$$

Hence, if N is Schur, system (2) is an UIO for (1). In [28], it is proved that such an observer exists if and only if the following two conditions are satisfied:

($\mathbf{c_1}$) $\text{rank}(CB) = \text{rank}(B) = p$.

($\mathbf{c_2}$) Matrix E in (2) yields the pair $(A-ECA, C)$ detectable.

Assume that conditions (c_1) and (c_2) are satisfied; then, observer (2) can be constructed by solving (3) for a Schur matrix N. Hence, for such an observer, there exist $c > 0$ and $\lambda \in (0, 1)$ satisfying:

$$|\hat{x}(k) - x(k)| \leq c\lambda^k |\hat{x}(0) - x(0)|, \tag{5}$$

for $k \geq 0$, i.e., observer (2) reconstructs the system state without using any input for arbitrarily large attack signals a.

Example 1: Consider the following system:

$$\begin{cases} x^+ = \begin{bmatrix} 0.2 & 0.5 \\ 0.2 & 0.7 \end{bmatrix} x + \begin{bmatrix} 1 \\ 2 \end{bmatrix} (u + a), \\ y = \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 3 & 2 \end{bmatrix} x. \end{cases} \tag{6}$$

An UIO exists for system (6). We let $W = \{1\}$, $a \in \mathcal{U}(-1, 1)$, $u \in \mathcal{U}(-1, 1)$, and the initial conditions $x_1(0), x_2(0) \in \mathcal{N}(0, 1)$. We solve (3) for some Schur matrix N and construct an unknown input observer for (6). The performance of the estimator is shown in Figure 1 for $\hat{x}(0) = [0, 0]^\top$.

B. Partial Unknown Input Observers

In [26], the problem of state estimation for continuous-time LTI system under sensor attacks is solved using a bank of Luenberger observers. Inspired by these results, we use a bank of partial UIOs to estimate the state of the system when actuator attacks occur. Here, we are implicitly assuming that either condition (c_1) or (c_2) (or both) cannot be satisfied by the triple (A, B, C). Let B be partitioned as $B = [b_1, \dots, b_i, \dots, b_p]$ where $b_i \in \mathbb{R}^{n\times 1}$ is the i-th column of B. Then, the attacked system (1) can be written as

$$\begin{cases} x^+ = Ax + Bu + b_W a_W, \\ y = Cx, \end{cases} \tag{7}$$

where the attack input a_W can be regarded as an unknown input and the columns of b_W are b_i for $i \in W$. Denote by b_J the matrix whose columns are b_i for $i \in J$. Let q be the largest integer such that for all $J \subset \{1, \dots, p\}$ with

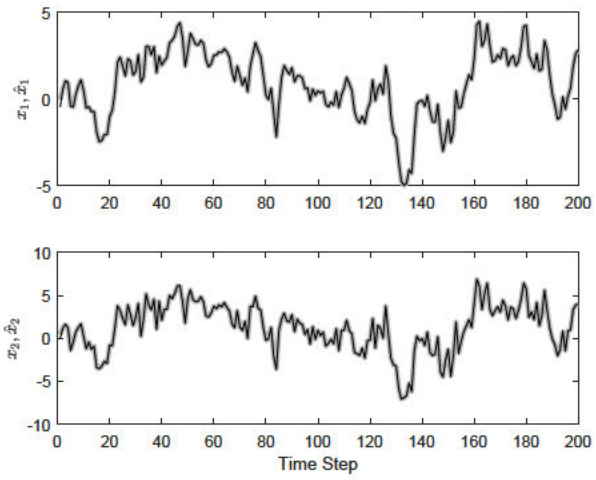

Fig. 1. Estimated states $\hat{x}$ converges to the true states x when $a \sim \mathcal{U}(-1, 1)$. Legend: $\hat{x}$ (grey), true states (black)

$\text{card}(J) \leq 2q$, the following is satisfied:

(c_3) $\text{rank}(Cb_J) = \text{rank}(b_J) = \text{card}(J)$.

(c_4) There exist (N_J, L_J, E_J, T_J) satisfying the equations:

$$\begin{cases} N_J(I - E_JC) + L_JC + (E_JC - I)A = 0, \\ (T_J + E_JC - I)B = 0, \\ (E_JC - I)b_J = 0, \end{cases} \tag{8}$$

with detectable pair $(C, A - E_JCA)$, and Schur N_J.

Then, if conditions (c_3) and (c_4) are satisfied, an UIO with the following structure:

$$\begin{cases} z_J^+ = N_Jz_J + T_JBu + L_Jy, \\ \hat{x}_J = z_J + E_Jy, \end{cases} \tag{9}$$

exists for each b_J, $J \subset \{1, \ldots, p\}$ with $\text{card}(J) \leq 2q < p$, where $z_J \in \mathbb{R}^n$ is the observer state, $\hat{x}_J \in \mathbb{R}^n$ denotes the state estimate, and (N_J, L_J, T_J, E_J) are the observer matrices satisfying (8), see [28] for further details. That is, system (9) is an UIO observer for the system:

$$\begin{cases} x^+ = Ax + Bu + b_Ja_J, \\ y = Cx, \end{cases} \tag{10}$$

with unknown input b_Ja_J and known input Bu. It follows that the estimation error $e_J = \hat{x}_J - x$ satisfies the difference equation:

$$e_J^+ = N_Je_J, \tag{11}$$

with N_J schur.

Assumption 2 *There are at most q attacked actuators, i.e.,*

$$\text{card}(W) \leq q < \frac{p}{2}, \tag{12}$$

where $q > 0$ is the largest integer satisfying (c_3) and (c_4).

Lemma 1 *Under Assumption 2, among all possible sets of q actuators, at least one set includes all the attacked actuators.*

Lemma 2 *Under Assumption 2, for each set of q actuators, among all its supersets of $2q$ actuators, at least one superset includes all the attacked actuators.*

Proof: Lemma 1 and Lemma 2 follow trivially from the fact that $0 < q < p/2$. ■

Note that the existence of an UIO for each b_J with $\text{card}(J) \leq 2q$ implies that if $W \subseteq J$, the estimation error $e_J := \hat{x}_J - x$ satisfies

$$|e_J| \leq c_J\lambda_J^k|e_J(0)|, \tag{13}$$

for some $c_J > 0$, $\lambda_J \in (0, 1)$, all $e_J(0) \in \mathbb{R}^n$ and all $k \geq 0$.

Let Assumption 2 be satisfied. We construct an UIO for each $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = q$ and each set S with $\text{card}(S) = 2q$. Then, by Lemma 1, there exists at least one set $\bar{J} \subset \{1, \ldots, p\}$ with $\text{card}(\bar{J}) = q$ such that $W \subseteq \bar{J}$ and the estimate produced by the UIO for $\bar{J}$ is a correct state estimate. Thus, the estimates given by any $S \supset \bar{J}$ with $\text{card}(S) = 2q$ will be consistent with that given by $\bar{J}$. This motivates the following estimation strategy:

For each set $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = q$ and all $k \geq 0$, we define $\pi_J(k)$ as the largest deviation between $\hat{x}_J(k)$ and $\hat{x}_S(k)$ that is given by any set $S \supset J$ with $\text{card}(S) = 2q$:

$$\pi_J(k) := \max_{S \supset J : \text{card}(S) = 2q} |\hat{x}_J(k) - \hat{x}_S(k)|, \tag{14}$$

for all $k \geq 0$, and define the sequence $\sigma(k)$ as

$$\sigma(k) := \underset{J \subset \{1, \ldots, p\} : \text{card}(J) = q}{\arg\min} \pi_J(k). \tag{15}$$

The estimate given by the set $\sigma(k)$ is a correct estimate, i.e.,

$$\hat{x}(k) := \hat{x}_{\sigma(k)}(k), \tag{16}$$

where $\hat{x}_{\sigma(k)}(k)$ denotes the estimate given by the set $\sigma(k)$, provides an exponential estimate of the system state. For simplicity and without generality, for all J and S, $z_J(0)$ and $z_S(0)$ are chosen such that $\hat{x}_J(0) = \hat{x}_S(0) = \hat{x}(0)$. The following result summarizes the ideas presented above.

Theorem 1 *Consider system* (1). *Let conditions* (c_3) *and* (c_4)*, and Assumption 1 and Assumption 2 be satisfied, and consider the multi-observer* (14)-(16). *Define the estimation error $e(k) := \hat{x}_{\sigma(k)}(k) - x(k)$; then, there exist constants $\bar{c} > 0$ and $\bar{\lambda} \in (0, 1)$ satisfying:*

$$|e(k)| \leq \bar{c}\bar{\lambda}^k|e(0)|, \tag{17}$$

for all $e(0) \in \mathbb{R}^n$, $k \geq 0$.

Proof: By Lemma 1, there exists at least one set $\bar{J}$ with $\text{card}(\bar{J}) = q$ such that $\bar{J} \supset W$. By (c_3) and (c_4), for $J = \bar{J} \supset W$ with $\text{card}(\bar{J}) = q$, there exist $c_{\bar{J}} > 0$ and $\lambda_{\bar{J}} \in (0, 1)$, such that

$$|e_{\bar{J}}(k)| \leq c_{\bar{J}}\lambda_{\bar{J}}^k|e(0)|, \tag{18}$$

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$. Moreover, for any set $S \supset \bar{J}$ with $\text{card}(S) = 2q$, we have $S \supset W$ $\forall k \geq 0$; hence, by (c_3) and (c_4), there exist $c_S > 0$ and $\lambda_S \in (0, 1)$ such that

$$|e_S(k)| \leq c_S\lambda_S^k|e(0)|, \tag{19}$$

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$. Consider $\pi_{\bar{J}}$ in (14). Combining the above results, we have that

$$\begin{aligned}\pi_{\bar{J}}(k) =& \max_{S \supset \bar{J}} |\hat{x}_{\bar{J}}(k) - \hat{x}_S(k)| \\ =& \max_{S \supset \bar{J}} |\hat{x}_{\bar{J}}(k) - x(k) + x(k) - \hat{x}_S(k)| \\ \leq& |e_{\bar{J}}(k)| + \max_{S \supset \bar{J}} |e_S(k)|,\end{aligned} \tag{20}$$

for all $k \geq 0$. From (18) and (19), we obtain

$$\pi_{\bar{J}}(k) \leq 2c'_{\bar{J}} \lambda'^k_{\bar{J}} |e(0)|, \tag{21}$$

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$, where

$$c'_{\bar{J}} := \max_{S \supset \bar{J}} \{c_{\bar{J}}, c_S\}, \lambda'_{\bar{J}} := \max_{S \supset \bar{J}} \{\lambda_{\bar{J}}, \lambda_S\}.$$

Note that $S \supset \bar{J}$ with $\text{card}(S) = 2q$, thus, from (15), we have $\pi_{\sigma(k)}(k) \leq \pi_{\bar{J}}(k)$. From Lemma 2, we know that there exists at least one set $\bar{S} \supset \sigma(k)$ with $\text{card}(\bar{S}) = 2q$ such that $\bar{S} \supset W \ \forall k \geq 0$, and, by ($c_3$) and ($c_4$), there exist $c_{\bar{S}} > 0$ and $\lambda_{\bar{S}} \in (0, 1)$ such that

$$|e_{\bar{S}}(k)| \leq c_{\bar{S}} \lambda^k_{\bar{S}} |e(0)|, \tag{22}$$

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$. From (14), by construction

$$\begin{aligned}\pi_{\sigma(k)}(k) =& \max_{S \supset \sigma(k): \text{card}(S) = 2q} |\hat{x}_{\sigma(k)}(k) - \hat{x}_S(k)| \\ \geq& |\hat{x}_{\sigma(k)}(k) - \hat{x}_{\bar{S}}(k)|,\end{aligned}$$

using the above lower bound on $\pi_{\sigma(k)}(k)$ and the triangle inequality, we have that

$$\begin{aligned}|e_{\sigma(k)}(k)| =& |\hat{x}_{\sigma(k)}(k) - x(k)| \\ =& |\hat{x}_{\sigma(k)}(k) - \hat{x}_{\bar{S}}(k) + \hat{x}_{\bar{S}}(k) - x(k)| \\ \leq& |\hat{x}_{\sigma(k)}(k) - \hat{x}_{\bar{S}}(k)| + |e_{\bar{S}}(k)| \\ \leq& \pi_{\sigma(k)}(k) + |e_{\bar{S}}(k)| \\ \leq& \pi_{\bar{J}}(k) + |e_{\bar{S}}(k)|,\end{aligned} \tag{23}$$

for all $k \geq 0$. Hence, from (21) and (22), we have

$$|e_{\sigma(k)}(k)| \leq \bar{c} \bar{\lambda}^k |e(0)|, \tag{24}$$

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$, where $\bar{c} = 3 \max\left\{c_{\bar{S}}, c'_{\bar{J}}\right\}$ and $\bar{\lambda} = \max\left\{\lambda_{\bar{S}}, \lambda'_{\bar{J}}\right\}$. Inequality (24) is of the form (17), and the result follows. ■

Example 2 Consider the following system:

$$\begin{cases} x^+ = \begin{bmatrix} 0.5 & 0 & 0.1 \\ 0.2 & 0.7 & 0 \\ 1 & 0 & 0.3 \end{bmatrix} x + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} (u + a), \\ y = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \end{bmatrix} x. \end{cases} \tag{25}$$

It can be verified that a complete UIO for (25) does not exist. However, partial UIOs exist for each b_J with $\text{card}(J) \leq 2$; then, $2q = 2$, i.e., $q = 1$. We let $W = \{2\}$, i.e., the second actuator is attacked. We let $u_1, u_2, u_3 \in \mathcal{U}(-1, 1)$, $a_2 \in \mathcal{U}(-1, 1)$, and $x_1(0), x_2(0), x_3(0) \in \mathcal{N}(0, 1)$. We construct an UIO for each set $J \subset \{1, 2, 3\}$ with $\text{card}(J) = 1$ and each $S \subset \{1, 2, 3\}$ with $\text{card}(S) = 2$. Totally $\binom{3}{1} + \binom{3}{2} = 6$ UIOs are designed and they are all initialized at $\hat{x}(0) = [0, 0, 0]^\top$. The performance of the estimator is shown in Figure 2.

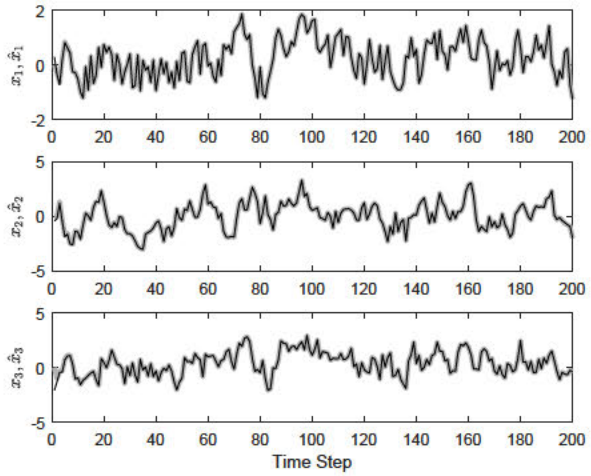

Fig. 2. Estimated states $\hat{x}$ converges to the true states x when $a_2 \sim \mathcal{U}(-1, 1)$. Legend: $\hat{x}$ (grey), true states (black)

IV. Isolation of Actuator Attacks

Once we have an estimate $\hat{x}(k)$ of $x(k)$, either using the complete observer in Section III-A or the partial multi-observer estimator in Section III-B, we can use these estimates, the system model (1), and the known inputs to exponentially reconstruct the attack signals. First, consider the complete observer in Section III-A. By construction, the estimation error $e = \hat{x} - x$ satisfies the difference equation (4) for some Schur matrix N. Note that $e = \hat{x} - x \Rightarrow x = \hat{x} - e \Rightarrow x^+ = \hat{x}^+ - e^+$. Then, the system dynamics (1) can be written in terms of e and $\hat{x}$ as follows:

$$\begin{cases} \hat{x}^+ = e^+ + A(\hat{x} - e) + B(u + a), \\ \Downarrow \\ a = B^{-1}_{left}(\hat{x}^+ - A\hat{x}) - u + B^{-1}_{left}(e^+ - Ae), \end{cases} \tag{26}$$

because B has full column rank, where B^{-1}_{left} denotes the Moore-Penrose pseudoinverse of B. Therefore, because e (and thus e^+ as well) vanishes exponentially, the following attack estimate:

$$\hat{a}(k) \quad = B^{-1}_{left}(\hat{x}(k) - A\hat{x}(k-1)) - u(k-1), \tag{27}$$

exponentially reconstructs the attack signals $a(k-1)$, i.e.,

$$\lim_{k \to \infty} (\hat{a}(k) - a(k-1)) = 0. \tag{28}$$

Then, for sufficiently large k, we assume $\text{supp}(a(k)) = \text{supp}(a(k-1))$, thus, the sparsity pattern of $\hat{a}(k)$ can be used to isolate actuator attacks at time k, i.e.,

$$\hat{W}(k) = \text{supp}(\hat{a}(k)), \tag{29}$$

where $\hat{W}(k)$ denotes the set of isolated actuators at time k. Note that we can only estimate a from $\hat{x}^+$ and e^+, which implies that we always have, at least, one-step delay.

Next, consider the partial multi-observer estimator given in Section III-B. In this case, the attack vector a can also be written as (26) but the estimation error dynamics is now given by some *nonlinear* difference equation characterized

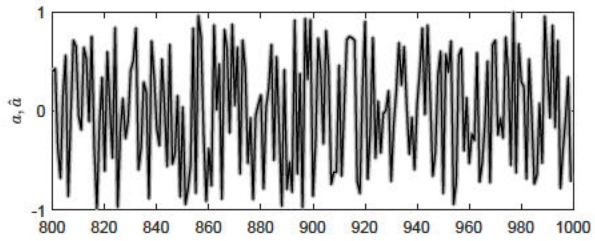

Fig. 3. Estimated attack $\hat{a}$ converges to a when $a \sim \mathcal{U}(-1,1)$. Legend: $\hat{a}$ (grey), a (black)

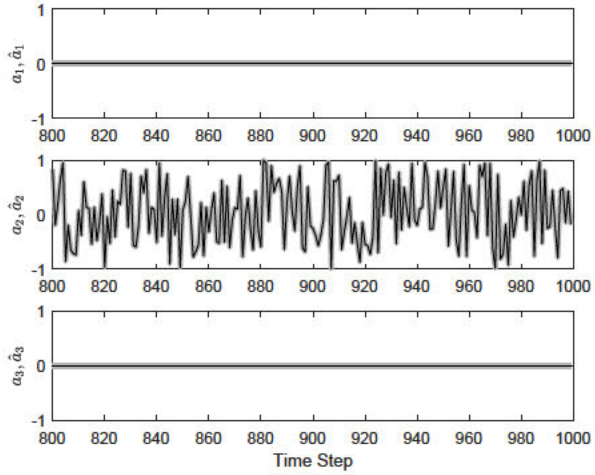

Fig. 4. Estimated states $\hat{a}$ converges to a when $a_2 \sim \mathcal{U}(-1,1)$. Legend: $\hat{a}$ (grey), a (black)

by the estimator structure in (14)-(16). Let the estimation error dynamics be given by

$$e^+ = f(e, x, a), \tag{30}$$

for some nonlinear function $f : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$. That is, the estimation error is given by some nonlinear function of the state and the attack signals. However, in Theorem 1, we have proved that e converges to the origin exponentially. Hence, the terms depending on e and e^+ in the expression for a in (26) vanishes exponentially and therefore the attack estimate in (27) exponentially reconstructs the attack signals. Again, the sparsity pattern of $\hat{a}(k)$ can be used to isolate actuator attacks using (29).

Example 3 Consider system (6) and the complete UIO in Example 1. Let $W = \{1\}$, $a \in \mathcal{U}(-1,1)$, $u \in \mathcal{U}(-1,1)$, and $x_1(0), x_2(0) \in \mathcal{N}(0,1)$. For $k \in [800, 1000]$, we obtain the attack estimates $\hat{a}(k)$ using (27). The performance is shown in Figure 3.

Example 4 Consider system (25) and the multi-observer estimator in Example 2. Let $q^* = 1$, $W = \{2\}$, $u_1, u_2, u_3 \in \mathcal{U}(-1,1)$, $a_2 \in \mathcal{U}(-1,1)$, and $x_1(0), x_2(0), x_3(0) \in \mathcal{N}(0, 1^2)$. For $k \in [800, 1000]$, we obtain $\hat{x}(k)$ using (14)-(16) and the vector $\hat{a}(k)$ from (27). The performance is shown in Figure 4. Note that $\hat{W}(k) = \{2\}$, i.e., the second actuator is correctly isolated.

V. Control

In this section, we propose a simple yet effective technique to stabilize the system by switching off the isolated actuators, i.e., by removing the columns of B which correspond to the isolated actuators, and closing the loop with a multi-observer (observer) based output dynamic feedback controller. Indeed, we need the system to be stabilizable after switching off the isolated actuators. We first consider the case when a complete UIO exists (Section III-A), i.e., $\hat{x}$ is generated by (2). We estimate $\hat{a}(k)$ using (27) and obtain $\hat{W}(k)$ from (29). Again, let B be partitioned as $B = [b_1, \ldots, b_i, \ldots, b_p]$. Define $\bar{J}(k) := \{1, \ldots, p\} \setminus \hat{W}(k)$ and $b_{\bar{J}(k)}$ as the matrix whose columns are b_i for $i \in \bar{J}(k)$, i.e., $\bar{J}(k) \subset \{1, \ldots, p\}$ is the set of isolated attack-free actuators and the columns of $b_{\bar{J}(k)}$ are the corresponding columns of B. Therefore, after switching off the set $\hat{W}(k)$ of actuators, the system has the following form:

$$x^+ = Ax + b_{\bar{J}(k)}\bar{u} \tag{31}$$

where $\bar{u} \in \mathbb{R}^{\mathrm{card}(\bar{J}(k))}$ is the set of isolated attack-free inputs. Let $q^\star$ be the largest integer such that (A, b_J) is stabilizable for each set $J \subset \{1, \ldots, p\}$ with $\mathrm{card}(J) \geq p - q^\star$ where b_J denotes a matrix whose columns are b_i for $i \in J$. We assume that at most $q^\star$ actuators are attacked. It follows that $p - q^\star \leq \mathrm{card}(\bar{J}(k)) \leq p$. We assume the following.

Assumption 3 *For any subset J with cardinality $\mathrm{card}(J) = p - q^\star$, there exists a linear switching state feedback controller $\bar{u} = K_{\bar{J}(k)}x$ such that the closed-loop dynamics:*

$$x^+ = (A + b_{\bar{J}(k)}K_{\bar{J}(k)})x, \tag{32}$$

is GAS for $b_{\bar{J}(k)}$ arbitrarily switching among all $b_{J'}$ with $J \subset J' \subset \{1, \ldots, p\}$ and $p - q^\star \leq \mathrm{card}(J') \leq p$.

Remark 1 *We do not give a method for designing the linear switching state feedback controller $\bar{u} = K_{\bar{J}(k)}x$. We refer the interested reader to, e.g.,* [29] *and references therein, for design methods of linear switching controllers.*

By switching off the set $\hat{W}(k)$ of actuators at time k, using the controller designed for the set $\bar{J}(k)$, and letting $\bar{u} = K_{\bar{J}(k)}\hat{x}$, the closed-loop system can be written as

$$x^+ = (A + b_{\bar{J}(k)}K_{\bar{J}(k)})x + b_{\bar{J}(k)}K_{\bar{J}(k)}e, \tag{33}$$

with estimation error $e = \hat{x} - x$ satisfying the difference equation (4) for some Schur matrix N. Because $e(k)$ converges to zero exponentially, $e(k)$ in (33) is a vanishing perturbation. Hence, under Assumption 3, it follows that $\lim_{k\to\infty} x(k) = 0$.

Next, assume that a complete UIO does not exist but partial UIOs exist for each b_J with $\mathrm{card}(J) \leq 2q < p$ (Section III-B) and $q \leq q^\star$. We assume that at most q actuators are attacked. We construct $\hat{x}(k)$ from (14)-(16), estimate $\hat{a}(k)$ using (27), and obtain $\hat{W}(k)$ from (29). After switching off the set $\hat{W}(k)$ of actuators, the system has the form (31) with $p - q \leq \mathrm{card}(\bar{J}(k)) \leq p$. We assume the following.

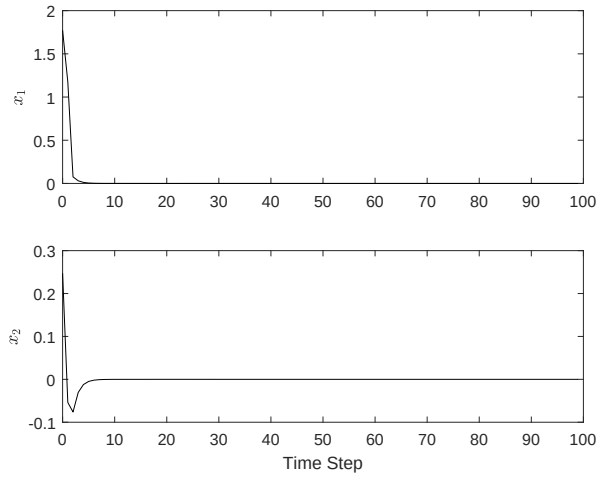

Fig. 5. State trajectories when $a_1 \sim \mathcal{U}(-1, 1)$.

Assumption 4 *For any subset J with cardinality $\mathrm{card}(J) = p-q$, there exists a linear switching state feedback controller $\bar{u} = K_{\bar{J}(k)}x$ such that the closed-loop dynamics:*

$$x^+ = (A + b_{\bar{J}(k)}K_{\bar{J}(k)})x, \tag{34}$$

is GAS for $b_{\bar{J}(k)}$ arbitrarily switching among all $b_{J'}$ with $J \subset J' \subset \{1, \ldots, p\}$ and $p - q \leq \mathrm{card}(J') \leq p$.

Then, by switching off the set $\hat{W}(k)$ of actuators at time k, using the controller designed for the set $\bar{J}(k)$, and letting $\bar{u} = K_{\bar{J}(k)}\hat{x}$, the closed-loop dynamics can be written in the form (33). Then, in this case, $e(k)$ is generated by some nonlinear difference equation of the form (30). Under Assumption 4, the closed-loop dynamics (33) is Input-to-State Stable (ISS) with input $e(k)$ and some linear gain, see [30]. Moreover, in Theorem 1, we have proved that $e(k)$ converges to the origin exponentially uniformly in $x(k)$ and $a(k)$. The latter and ISS of the system dynamics imply that $\lim_{k\to\infty} x(k) = 0$ [31].

Example 5 Consider the following system:

$$\begin{cases} x^+ = \begin{bmatrix} 1.5 & 0 & 0.1 \\ 0.2 & 0.7 & 0 \\ 1 & 0 & 0.3 \end{bmatrix} x + \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} (u + a), \\ y = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \end{bmatrix} x, \end{cases} \tag{35}$$

with $u = K_{\bar{J}(k)}\hat{x}$. Since (A, b_i) is stabilizable for $i \in \{1, 2, 3\}$, we have $q^\star = 2$. It can be verified that there does not exist a complete UIO for this system but partial UIOs exist for each b_J with $\mathrm{card}(J) \leq 2$, then we have $q = 1$. We let $W = \{1\}$, and $a_1 \in \mathcal{U}(-1, 1)$. We construct $\binom{3}{1} + \binom{3}{2} = 6$ UIOs and use the design method given in [29] to build controllers for actuators $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, $\{1, 2, 3\}$. Then, we use the multi-observer approach in Section III-B to estimate the state, reconstruct the attack signals, and control the system. The state of the system is shown in Figure 5.

VI. Conclusion

We have addressed the problem of state estimation, attack isolation, and control for discrete-time LTI systems under (potentially unbounded) actuator false data injection attacks. Using a bank of Unknown Input Observers (UIOs), we have proposed an estimator that reconstructs the system state and the attack signals. We have proved that the designed estimator provides exponentially stable estimation errors for potentially unbounded attack signals. We used these estimates to control the system and isolate actuator attacks. We have provided simulations results to illustrate the performance of the results.

References

[1] H. Fawzi, P. Tabuada, and S. Diggavi, "Security for control systems under sensor and actuator attacks," in *IEEE 51st Conference on Desision and Control (CDC)*, 2012, pp. 3412–3417.

[2] M. Massoumnia, G. C. Verghese, and A. S. Willsky, "Failure detection and identification in linear time-invariant systems," *Technology*, no. July, 1986.

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

[4] Y. Mo and B. Sinopoli, "Resilient detection in the presence of integrity attacks," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 31–43, 2014.

[5] K. G. Vamvoudakis, J. P. Hespanha, B. Sinopoli, and Y. Mo, "Detection in adversarial environments," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3209–3223, 2015.

[6] M. S. Chong and M. Kuijper, "Characterising the vulnerability of linear control systems under sensor attacks using a system's security index," in *IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 5906–5911.

[7] K. G. Vamvoudakis, J. P. Hespanha, B. Sinopoli, and Y. Mo, "Adversarial detection as a zero-sum game," in *IEEE 51st Conference on Decision and Control (CDC)*, 2012, pp. 7133–7138.

[8] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. a. Seshia, and P. Tabuada, "Secure State Estimation For Cyber Physical Systems Under Sensor Attacks: A Satisfiability Modulo Theory Approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 4917 – 4932, 2017.

[9] S. Z. Yong, M. Zhu, and E. Frazzoli, "Resilient state estimation against switching attacks on stochastic cyber-physical systems," in *IEEE 54th Conference on Decision and Control (CDC)*, 2015, pp. 5162–5169.

[10] J. Park, J. Weimer, and I. Lee, "Sensor attack detection in the presence of transient faults," *6th International Conference on Cyber-Physical Systems*, no. April, pp. 1–10, 2015.

[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, pp. 21–32, 2009.

[12] Z. Tang, K. Margreta, M. Chong, and I. Mareels, "Sensor attack correction for linear systems with known inputs," *7th IFAC Workshop on Distributed Estimation and Control in Networked Systems*, no. May, 2018.

[13] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," *2012 50th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2012*, pp. 1806–1813, 2012.

[14] C. Murguia and J. Ruths, "Characterization of a CUSUM model-based sensor attack detector," in *2016 IEEE 55th Conference on Decision and Control, CDC 2016*, 2016, pp. 1303–1309.

[15] V. S. Dolk, P. Tesi, C. D. Persis, and W. P. M. H. Heemels, "Event-triggered control systems under denial-of-service attacks," *IEEE Transactions on Control of Network Systems*, vol. 4, pp. 93–105, 2017.

[16] N. Hashemil, C. Murguia, and J. Ruths, "A comparison of stealthy sensor attacks on control systems," in *proceedings of the American Control Conference (ACC)*, 2017.

[17] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2715–2729, 2013.
[18] C. Murguia and J. Ruths, "On reachable sets of hidden cps sensor attacks," in *proceedings of the American Control Conference (ACC)*, 2017.
[19] J. Giraldo, A. Cardenas, and N. Quijano, "Integrity attacks on real-time pricing in smart grids: Impact and countermeasures," *IEEE Transactions on Smart Grid*, 2016.
[20] C. Murguia and J. Ruths, "Cusum and chi-squared attack detection of compromised sensors," in *proceedings of the IEEE Multi-Conference on Systems and Control (MSC)*, 2016.
[21] S. H. Kafash, J. Giraldo, C. Murguia, A. A. Cardenas, and J. Ruths, "Constraining attacker capabilities through actuator saturation," in *proceedings of the American Control Conference (ACC)*, 2017.
[22] C. Murguia, N. van de Wouw, and J. Ruths, "Reachable sets of hidden cps sensor attacks: Analysis and synthesis tools," in *proceedings of the IFAC World Congress*, 2016.
[23] C. Murguia, I. Shames, F. Farokhi, and D. Nešić, "On privacy of quantized sensor measurements through additive noise," in *proceedings of the 57th IEEE Conference on Decision and Control (CDC)*, 2018.
[24] J. Kim, C. Lee, H. Shim, Y. Eun, and J. H. Seo, "Detection of sensor attack and resilient state estimation for uniformly observable nonlinear systems," *IEEE 55th Conference on Decision and Control (CDC)*, pp. 1297–1302, 2016.
[25] T. Yang, C. Murguia, M. Kuijper, and D. Nešić, "Attack detection and isolation for discrete-time nonlinear systems," in *proceedings of the Australian and New Zealand Control Conference (ANZCC)*, 2018.
[26] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," *Proceedings of the American Control Conference*, vol. 2015-July, pp. 2439–2444, 2015.
[27] E. D. Sontag, "Input to state stability: Basic concepts and results," *Lecture Notes in Mathematics*, vol. 1932, pp. 163–220, 2008.
[28] S. X. Ding, *Model-based fault diagnosis techniques: Design schemes, algorithms, and tools*, 2008.
[29] J. Daafouz, P. Riedinger, and C. Iung, "Stability Analysis and Control Synthesis for Switched Systems: A switched Lyapunov function approach," *IEEE Trans. on Automat. Contr.*, vol. 47, no. 11, pp. 1883–1887, 2002.
[30] E. D. Sontag and Y. Wang, "Output-to-state stability and detectability of nonlinear systems," *Systems & Control Letters*, vol. 29, no. September, pp. 279–290, 1997.
[31] Z.-P. Jiang and Y. Wang, "Input-to-state stability for discrete-time nonlinear systems," *Automatica*, vol. 37, no. 6, pp. 857–869, 2001.