An Unknown Input Multi-Observer Approach for Estimation and Control under Adversarial Attacks

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

Abstract—We address the problem of state estimation, attack isolation, and control of discrete-time linear time-invariant systems under (potentially unbounded) actuator and sensor false data injection attacks. Using a bank of unknown input observers, each observer leading to an exponentially stable estimation error (in the attack-free case), we propose an observer-based estimator that provides exponential estimates of the system state in spite of actuator and sensor attacks. Exploiting sensor and actuator redundancy, the estimation scheme is guaranteed to work if a sufficiently small subset of sensors and actuators are under attack. Using the proposed estimator, we provide tools for reconstructing and isolating actuator and sensor attacks; and a control scheme capable of stabilizing the closed-loop dynamics by switching off isolated actuators. Simulation results are presented to illustrate the performance of our tools.

Index Terms—Unknown input observers, cyber-physical systems, sensor and actuator attacks, linear systems, control.

I. INTRODUCTION

Networked Control Systems (NCSs) have received considerable attention in recent years due to their numerous advantages (e.g., reduced weight, volume and installation costs, and better maintainability) when compared with traditional control systems where sensors and actuators communicate through point-to-point (wired) links. Networked Control Systems are being used in many engineering applications, e.g., energy, transportation, military, health care, and manufacturing. With the growth of NCSs, new security challenges have become an important issue as wireless communication networks increasingly serve as new access points for adversaries trying to disrupt the process. Cyber-physical attacks on NCSs have caused substantial damage to a number of engineering systems. A well-known example is the StuxNet virus that targeted Siemens' supervisory control and data acquisition systems. Another example is the false data injection attacks on power systems [1]. A more recent incident happend in 2014, where the computers of a German steel mill were hacked and a destruction of a blast furnace was caused. These and many other recent incidents show that tools to identify and deal with attacks on NCSs are needed.

In [2]-[12], various security and privacy problems for linear control systems have been addressed and solved. In general, analysis (synthesis) tools are proposed to quantify (minimize) the performance degradation induced by different classes of attacks, e.g., false-data-injection, replay, zero dynamics, and

denial-of-service. There are also some results addressing the nonlinear case. The problem of state estimation for nonlinear power systems under sensor attacks is solved in [13] by using compressed sensing technique. In [14], the authors address the problem of sensor attack detection and state estimation for uniformly observable continuous-time nonlinear systems. In [15], Satisfiability Modulo Theory (SMT) solvers are used for state estimation for nonlinear differentially flat systems with corrupted sensors. In our previous work [16], [17], the problem of state estimation and attack isolation for a class of nonlinear systems with *positive-slope nonlinearities* is considered. Similar to the ideas given in [18], we provided an observer-based estimation/isolation strategy, using a bank of circle-criterion observers, which provides a robust estimate of the system state in spite of sensor attacks and effectively pinpoints attacked sensors. Most of the existing work assume actuators to be healthy and only consider sensor attacks. There are only a few results dealing with attacked actuators. For instance, in [19], the authors study the effect of actuator attacks on the performance of linear quadratic regulators. In [20] and [21], the problem of state estimation under sensor and actuator attacks is addressed using compressed sensing ideas and SMTbased techniques, respectively. An adaptive control scheme that guarantees uniform ultimate boundedness of the closedloop dynamics despite of sensor and actuator attacks is given in [22].

The core of our estimation scheme is inspired by the work in [11], where the problem of state estimation for *continuous-time* LTI systems is addressed. The authors propose a multi-observer estimator, using a bank of Luenberger observers, that provides a robust estimate of the system state in spite of sensor attacks. In this manuscript, using banks of Unknown Input Observers (UIOs), we address the problem of robust state estimation, attack isolation, and control for discrete-time LTI systems (with matrices (A, B, C) under (potentially unbounded) actuator and sensor attacks. Unknown input observers are dynamical systems capable of estimating the state of the plant without using any input signals. If such an observer exists for the matrices (A, B, \tilde{C}_i) , where \tilde{C}_i denotes a submatrix of C with fewer rows and the same number of columns, then, using a bank of observers, we can perform state estimation and attack isolation when a sufficiently small subset of sensors is attacked (even if all inputs 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, B, C_i) , i.e., the *i*-th observer is driven by the output signals associated with C_i only. If the outputs corresponding to C_i are attack-

This work was supported by the Australian Research Council under the Discovery Project DP170104099.

The authors are with the Department of Electrical and Electronics Engineering, the University of Melbourne, Australia. tianciy@student.unimelb.edu.au

free, 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. If a UIO does not exist for (A, B, \hat{C}_i) , but it does for $(A, \hat{B}_i, \hat{C}_i)$, where B_i is a submatrix of B with fewer columns and the same number of rows, i.e., the *i*-th observer does *not* use the input signals associated with B_i , but it does use the remaining input signals and the output signals corresponding to C_i , then using a bank of these UIOs, 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 and sensors are under attack. If the inputs corresponding to B_i include all the attacked ones and the outputs corresponding to C_i are attack-free, this UIO produces exponentially stable estimation error. For every pair of UIOs in the bank, we compute the largest difference between their estimates and select the pair leading to the smallest difference. We prove that these observers provide exponential estimate of the system state. Once we have an estimate of the state, we provide tools for reconstructing attack signals using model matching techniques. Attacked actuators and sensors are isolated by simply checking the sparsity of the estimated attack signals. Finally, after obtaining state estimates and isolation has been performed, we provide a control scheme for stabilizing the closed-loop dynamics. In the case with sensor attacks only (no actuators attacks), we show that a separation principle between estimation and control holds and the system can be stabilized by closing the loop with the multi-observer estimator and a static output feedback controller. When both sensors and actuator are attacks, we propose an effective technique to stabilize the system by switching off the isolated actuators, and closing the loop with a multi-observer based output timevarying feedback controller. Because attack signals might be zero for some time instants, actuators isolated as attackfree 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) [23] of the closed-loop system with respect to the exponentially stable estimation error to conclude on stability of the closedloop dynamics. Compared to the adaptive controller proposed in [22], where a particular class of attacks is considered and ultimate boundedness of the closed-loop system is guaranteed only, our controller is able drive the system state asymptotically to the origin under arbitrary and potentially unbounded attack signals.

The paper is organized as follows. In Section 2, we present some preliminary results needed for the subsequent sections. In Section 3, we introduce the proposed UIO-based estimation schemes. In Section 4, a method for isolating actuator attacks is described. The proposed control scheme is given in Section V. Finally, in Section 6, 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 $\mathbb{R}^{n \times m}$ the set of $n \times m$ matrices for any $m, n \in \mathbb{N}$. For any vector $v \in \mathbb{R}^n$, we denote v^J the stacking of all $v_i, i \in J, J \subset \{1, \ldots, n\}$, $|v| = \sqrt{v^{\top}v}$, and $\operatorname{supp}(v) = \{i \in \{1, \dots, n\} | v_i \neq 0\}$. For matrices $C \in \mathbb{R}^{p \times n}$, $C^{\top} = (c_1^{\top}, \dots, c_p^{\top})$, we denote C^J the stacking of all rows $c_i \in \mathbb{R}^{1 \times n}$, $i \in J$, $J \subset \{1, \ldots, n\}$. Set J is called a superset of set S if $S \subseteq J$. We denote the cardinality of a set S as 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 $n \times n$, respectively. We simply write 0 and I when their dimensions are evident. A continuous function $\alpha : [0, a) \to [0, \infty)$ is said to belong to class K, if it is strictly increasing and $\alpha(0) = 0$, [24]. Similarity, a continuous function $\beta : [0, a) \times [0, \infty) \rightarrow$ $[0,\infty)$ is said to belong to class KL if, for fixed s, the mapping $\beta(r,s)$ belongs to class K with respect to r and, for fixed r, the mapping $\beta(r,s)$ is decreasing with respect to s and $\beta(r,s) \to 0$ as $s \to \infty$, [24].

III. ESTIMATION

In [18], 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 UIOs to estimate the state of the system when sensor and actuator attacks both occur. Consider a discrete-time linear system under sensor and actuator attacks:

$$\begin{cases} x^+ = Ax + B(u + a_u) \\ y = Cx + a_y \end{cases}$$
(1)

with state $x \in \mathbb{R}^n$, output $y \in \mathbb{R}^{n_y}$, known input $u \in \mathbb{R}^{n_u}$, vector of actuator attacks $a_u \in \mathbb{R}^{n_u}$, $a_u = (a_{u1}, \ldots, a_{un_u})^{\top}$, i.e., $a_{ui}(k) = 0$ for all $k \ge 0$ if the *i*-th actuator is attackfree; otherwise, $a_{ui}(k_i) \ne 0$ for some $k_i \ge 0$ and can be arbitrarily large, and vector of sensor attacks $a_y \in \mathbb{R}^{n_y}$, $a_y = (a_{y1}, \ldots, a_{yn_y})^{\top}$, i.e., $a_{yi}(k) = 0$ for all $k \ge 0$ if the *i*th sensor is attack-free; otherwise, $a_{yi}(k_i) \ne 0$ for some $k_i \ge 0$ and can be arbitrarily large. Matrices A, B, C are of appropriate dimensions, and we assume that (A, B) is stabilizable, (A, C) is detectable, and B has full column rank. Let $W_u \subset \{1, \ldots, n_u\}$ denotes the *unknown* set of attacked actuators, and $W_y \subset \{1, \ldots, n_y\}$ denotes the *unknown* set of attacked sensors.

Assumption 1 The sets of attacked actuators and sensors do not change over time, i.e., $W_u \subset \{1, \ldots, n_u\}, W_y \subset \{1, \ldots, n_y\}$ are constant (time-invariant) and $\operatorname{supp}(a_u(k)) \subseteq W_u$, $\operatorname{supp}(a_y(k)) \subseteq W_y$, for all $k \ge 0$.

A. Complete Unknown Input Observers

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

$$\begin{cases} z_{J_s}^+ = N_{J_s} z_{J_s} + L_{J_s} y^{J_s}, \\ \hat{x}_{J_s} = z_{J_s} + E_{J_s} y^{J_s}, \end{cases}$$
(2)

where $z_{J_s} \in \mathbb{R}^n$ is the state of the observer, $\hat{x}_{J_s} \in \mathbb{R}^n$ denotes the estimate of the system state, $(N_{J_s}, L_{J_s}, E_{J_s})$ are observer matrices of appropriate dimensions to be designed. It is easy to verify that if $(N_{J_s}, L_{J_s}, E_{J_s})$ satisfy the following equations:

$$\begin{cases} N_{J_s}(I - E_{J_s}C^{J_s}) + L_{J_s}C^{J_s} + (E_{J_s}C^{J_s} - I)A = 0, \\ (E_{J_s}C^{J_s} - I)B = 0; \end{cases}$$
(3)

then, the estimation error $e_{J_s} = \hat{x}_{J_s} - x$ satisfies:

$$e_{J_s}^+ = N_{J_s} e_{J_s}.$$
 (4)

If N_{J_s} is Schur, system (2) is called a UIO for (1). In [25], it is proved that such observer exists if and only if the following two conditions are satisfied:

(c₁) rank
$$(C^{J_s}B)$$
 = rank $(B) = n_u$

(c₂) The pair $(C^{J_s}, A - E_{J_s}C^{J_s}A)$ is detectable.

Let q be the largest integer such that for all $J_s \subset \{1, \ldots, n_y\}$ with $\operatorname{card}(J_s) \ge n_y - 2q > 0$, conditions (c_1) and (c_2) are satisfied; then, observer (2) can be constructed for any C^{J_s} with $\operatorname{card}(J_s) \ge n_y - 2q$ by solving (3) for a Schur matrix N_{J_s} . Hence, for such an observer, if $a_y^{J_s}(k) = 0$ for all $k \ge 0$, there exist $c_{J_s} > 0$, $\lambda_{J_s} \in (0, 1)$ satisfying:

$$|e_{J_s}(k)| \le c_{J_s} \lambda_{J_s}^k |e_{J_s}(0)|, \tag{5}$$

for all $k \ge 0$ [25], where $e_{J_s} = \hat{x}_{J_s} - x$.

Assumption 2 There are at most q sensors attacked by an adversary, i.e.,

$$\operatorname{card}(W_y) \le q < \frac{n_y}{2},$$
 (6)

where q is the largest positive integer satisfying conditions (c_1) and (c_2) .

Lemma 1 Under Assumption 2, among each set of $n_y - q$ sensors, at least $n_y - 2q > 0$ of them are attack-free.

Let Assumption 2 be satisfied. Inspired by the ideas in [11], we use a UIO for each subset $J_s \subset \{1, \ldots, n_y\}$ of sensors with $\operatorname{card}(J_s) = n_y - q$ and for each subset $S_s \subset \{1, \ldots, n_y\}$ of sensors with $\operatorname{card}(S_s) = n_y - 2q$. Under Assumption 2, there exists at least one set $\overline{J}_s \subset \{1, \ldots, n_y\}$ with $\operatorname{card}(\overline{J}_s) = n_y - q$ such that $a_y^{J_s}(k) = 0$ for all $k \ge 0$. Then, the estimate given by the UIO for \overline{J}_s is a correct estimate, and the estimate given by the UIO for any $S_s \subset \overline{J}_s$ with $\operatorname{card}(S_s) = n_y - 2q$ is consistent with that given by \overline{J}_s . This motivates the following estimation strategy.

For each set J_s with $\operatorname{card}(J_s) = n_y - q$, we define $\pi_{J_s}(k)$ as the largest deviation between \hat{x}_{J_s} and \hat{x}_{S_s} that is given by any $S_s \subset J_s$ with $\operatorname{card}(S_s) = n_y - 2q$, i.e.,

$$\pi_{J_s}(k) := \max_{\substack{S_s \subset J_s: \text{card}(S_s) = n_y - 2q}} |\hat{x}_{J_s}(k) - \hat{x}_{S_s}(k)|, \quad (7)$$

for all $k \ge 0$, and the sequence $\sigma_s(k)$ as

$$\sigma_s(k) := \arg\min_{J_s \subset \{1,\dots,n_y\}: \operatorname{card}(J_s) = n_y - q} \pi_{J_s}(k).$$
(8)

Then, as proved below, the estimate indexed by $\sigma_s(k)$:

$$\hat{x}(k) := \hat{x}_{\sigma_s(k)}(k), \tag{9}$$

is an exponential attack-free estimate of the system state. For simplicity and without generality, for all J_s and S_s , $z_{J_s}(0)$ and $z_{S_s}(0)$ are chosen such that $\hat{x}_{J_s}(0) = \hat{x}_{S_s}(0) = \hat{x}(0)$. The following result summarizes the ideas presented above.

Theorem 1 Consider system (1), observer (2), and the complete multi-observer estimator (7)-(9). Define the estimation error $e(k) := \hat{x}_{\sigma_s(k)}(k) - x(k)$, and let conditions (c_1) - (c_2) and Assumptions 1-2 be satisfied; then, there exist constants $\bar{c} > 0$, $\bar{\lambda} \in (0, 1)$ satisfying:

$$|e(k)| \le \bar{c}\bar{\lambda}^k |e(0)|, \tag{10}$$

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

Proof: Under Assumption 2, there exists at least one set J_s with $\operatorname{card}(\bar{J}_s) = n_y - q$ such that $a_y^{\bar{J}_s}(k) = 0$ for all $k \ge 0$. Then, there exist $c_{\bar{J}_s} > 0$ and $\lambda_{\bar{J}_s} \in (0, 1)$ such that

$$|e_{\bar{J}_s}(k)| \le c_{\bar{J}_s} \lambda_{\bar{J}_s}^k |e(0)|, \tag{11}$$

for all $e(0) \in \mathbb{R}^n$ and $k \ge 0$. Moreover, for any set $S_s \subset \overline{J}_s$ with $\operatorname{card}(S_s) = n_y - 2q$, we have $a_y^{S_s}(k) = 0 \ \forall k \ge 0$; hence, there exist $c_{S_s} > 0$ and $\lambda_{S_s} \in (0, 1)$ such that

$$|e_{S_s}(k)| \le c_{S_s} \lambda_{S_s}^k |e(0)|,$$
 (12)

for all $e(0) \in \mathbb{R}^n$ and $k \ge 0$. Consider π_{J_s} in (7). Combining the above inequalities, we have

$$\pi_{\bar{J}_{s}}(k) = \max_{S_{s} \subset \bar{J}_{s}} |\hat{x}_{\bar{J}_{s}}(k) - \hat{x}_{S_{s}}(k)|$$

$$= \max_{S_{s} \subset \bar{J}_{s}} |\hat{x}_{\bar{J}_{s}}(k) - x(k) + x(k) - \hat{x}_{S_{s}}(k)|$$

$$\leq |e_{\bar{J}_{s}}(k)| + \max_{S_{s} \subset \bar{J}_{s}} |e_{S_{s}}(k)|$$

$$\leq 2c'_{\bar{J}_{s}} \lambda'_{\bar{J}_{s}}^{k} |e(0)|,$$
(13)

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

$$c'_{\bar{J}_s} := \max_{S_s \subset \bar{J}_s} \left\{ c_{\bar{J}_s}, c_{S_s} \right\},$$
$$\lambda'_{\bar{J}_s} := \max_{S_s \subset \bar{J}_s} \left\{ \lambda_{\bar{J}_s}, \lambda_{S_s} \right\}.$$

Note that $S_s \subset \overline{J}_s$, $\operatorname{card}(S_s) = n_y - 2q_2$. Then, from (8), we have $\pi_{\sigma_s(k)}(k) \leq \pi_{\overline{J}_s}(k)$. From Lemma 1, we know that there exist at least one set $\overline{S}_s \subset \sigma_s(k)$ with $\operatorname{card}(\overline{S}_s) = n_y - 2q$, such that $a_y^{\overline{S}_s}(k) = 0$ for all $k \geq 0$, and there exist $c_{\overline{S}_s} > 0$ and $\lambda_{\overline{S}_s} \in (0, 1)$ such that

$$|e_{\bar{S}_{s}}(k)| \le c_{\bar{S}_{s}}\lambda_{\bar{S}_{s}}^{k}|e(0)|, \tag{14}$$

for all $e(0) \in \mathbb{R}^n$ and $k \ge 0$. From (7), we have

$$\pi_{\sigma_s(k)}(k) = \max_{S_s \subset \sigma_s(k)} |\hat{x}_{\sigma_s(k)}(k) - \hat{x}_{S_s}(k)|$$

$$\geq |\hat{x}_{\sigma_s(k)}(k) - \hat{x}_{\bar{S}_s}(k)|.$$



Fig. 1. Estimated states \hat{x} converges to the true states x when $a_u, a_{y3} \sim \mathcal{U}(-10, 10)$. Legend: \hat{x} (blue), true states (black)

Using this lower bound on $\pi_{\sigma_s(k)}(k)$ and the triangle inequality we have that

$$e_{\sigma_{s}(k)}(k)| = |\hat{x}_{\sigma_{s}(k)}(k) - x(k)|$$

$$= |\hat{x}_{\sigma_{s}(k)}(k) - \hat{x}_{\bar{S}_{s}}(k) + \hat{x}_{\bar{S}_{s}}(k) - x(k)|$$

$$\leq |\hat{x}_{\sigma_{s}(k)}(k) - \hat{x}_{\bar{S}_{s}}(k)| + |e_{\bar{S}_{s}}(k)| \qquad (15)$$

$$\leq \pi_{\sigma_{s}(k)}(k) + |e_{\bar{S}_{s}}(k)|$$

$$\leq \pi_{\bar{J}_{s}}(k) + |e_{\bar{S}_{s}}(k)|,$$

for all $k \ge 0$. Hence, from (13) and (14), we have

$$|e_{\sigma_s(k)}(k)| \le \bar{c}\bar{\lambda}^k |e(0)|,\tag{16}$$

for all $e(0) \in \mathbb{R}^n$ and $k \ge 0$, where $\bar{c} = 3 \max\{c_{\bar{S}_s}, c'_{\bar{J}_s}\}$ and $\bar{\lambda} = \max\{\lambda_{\bar{S}_s}, \lambda'_{\bar{J}_s}\}$. Inequality (16) is of the form (10) and the result follows.

Example 1: Consider the following system subject to actuator and sensor attacks:

$$\begin{cases} x^{+} = \begin{bmatrix} 0.2 & 0.5 \\ 0.2 & 0.7 \end{bmatrix} x + \begin{bmatrix} 1 \\ 2 \end{bmatrix} (u + a_{u}), \\ y = \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 3 & 2 \\ 2 & 1 \end{bmatrix} x + a_{y}. \end{cases}$$
(17)

It can be verified that a UIO of the form (2) exists for each C^{J_s} with $J_s \subset \{1, 2, 3, 4\}$ and $\operatorname{card}(J_s) \geq 2$; then, 4-2q=2, i.e., q=1 and at most one sensor is attacked. We attack the actuator and let $W_y = \{3\}$, i.e., the third sensor is attacked. We let $u \sim \mathcal{U}(-1, 1)$, $a_u, a_{y3} \sim \mathcal{U}(-10, 10)$. We design a UIO for each J_s with $\operatorname{card}(J_s) = 3$, and for each S_s with $\operatorname{card}(S_s) = 2$. Therefore, totally $\binom{4}{3} + \binom{4}{2} = 10$ UIOs are designed and they are all initialized at $\hat{x}(0) = [0, 0]^{\top}$. For $k \in [0, 19]$, the estimator (2), (26)-(28) is used to construct $\hat{x}(k)$. The performance of the estimator is shown in Figure 1.

B. Partial Unknown Input Observers

Here, we are implicitly assuming that either condition (c_1) or (c_2) (or both) cannot be satisfied for any C^{J_s} with

 $\operatorname{card}(J_s) = n_y - 2q$ with $q \ge 1$. Let *B* be partitioned as $B = [b_1, \ldots, b_i, \ldots, b_{n_u}]$ 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

$$\begin{cases} x^+ = Ax + Bu + b_{W_u} a^{W_u}, \\ y = Cx + a_y, \end{cases}$$
(18)

where the attack input a^{W_u} can be regarded as an unknown input and the columns of b_{W_u} are b_i , $i \in W_u$. Denote by b_{J_u} the matrix whose columns are b_i for $i \in J_u$. Let q_1 and q_2 be the largest integers such that for all $J_u \subset \{1, \ldots, n_u\}$ with $\operatorname{card}(J_u) \leq 2q_1 < n_u$ and $J_s \subset \{1, \ldots, n_y\}$ with $\operatorname{card}(J_s) \geq$ $n_y - 2q_2 > 0$, the following is satisfied:

(c₃) rank
$$(C^{J_s}b_{J_u}) = \operatorname{rank}(b_{J_u}) = \operatorname{card}(J_u).$$

(c₄) There exists $(N_{J_{us}}, L_{J_{us}}, E_{J_{us}}, T_{J_{us}})$ satisfying:

$$\begin{cases} N_{J_{us}}(I - E_{J_{us}}C^{J_s}) + L_{J_{us}}C^{J_s} + (E_{J_{us}}C^{J_s} - I)A = 0, \\ (T_{J_{us}} + E_{J_{us}}C^{J_s} - I)B = 0, \\ (E_{J_{us}}C^{J_s} - I)b_{J_u} = 0, \\ (19) \end{cases}$$

with detectable pair $(C^{J_{us}}, A - E_{J_{us}}C^{J_{us}}A)$ and Schur $N_{J_{us}}$. If conditions (c_3) and (c_4) are satisfied, a UIO with the following structure exists for each b_{J_u} with $J_u \subset \{1, \ldots, n_u\}$, $card(J_u) \leq 2q_1 < n_u$ and each C^{J_s} with $J_s \subset \{1, \ldots, n_y\}$, $card(J_s) \geq n_y - 2q_2 > 0$:

$$\begin{cases} z_{J_{us}}^{+} = N_{J_{us}} z_{J_{us}} + T_{J_{us}} B u + L_{J_{us}} y^{J_s}, \\ \hat{x}_{J_{us}} = z_{J_{us}} + E_{J_{us}} y^{J_s}, \end{cases}$$
(20)

where $z_{J_{us}} \in \mathbb{R}^n$ is the observer state, $\hat{x}_{J_{us}}$ denotes the state estimate, and $(N_{J_{us}}, L_{J_{us}}, T_{J_{us}}, E_{J_{us}})$ are the observer matrices satisfying (19), see [25] for further details. That is, system (20) is a UIO for the system:

$$\begin{cases} x^{+} = Ax + Bu + b_{J_{u}}a_{u}^{J_{u}}, \\ y^{J_{s}} = C^{J_{s}}x + a_{y}^{J_{s}}, \end{cases}$$
(21)

with unknown input $b_{J_u}a^{J_u}$ and known input Bu. It follows that the estimation error $e_{J_{us}} = \hat{x}_{J_{us}} - x$ satisfies:

$$e_{J_{us}}^+ = N_{J_{us}} e_{J_{us}}, (22)$$

for some Schur matrix $N_{J_{us}}$. We refer to UIOs of the form (21) as *partial* UIOs for the pair (J_u, J_s) .

Assumption 3 There are at most q_1 actuators and at most q_2 sensors attacked by an adversary, i.e.,

$$\operatorname{card}(W_u) \le q_1 < \frac{n_u}{2} \tag{23}$$

$$\operatorname{card}(W_y) \le q_2 < \frac{n_y}{2},\tag{24}$$

where q_1 and q_2 are the largest positive integers satisfying (c_3) and (c_4) .

Remark 1 Note that if conditions (c_3) and (c_4) are satisfied for b_{J_u} with $card(J_u) = 2q_1 = n_u$, then conditions (c_1) and (c_2) are satisfied, and (20) is a complete UIO for (1) for $T_{J_{us}} = 0$. Since we are considering partial UIOs, we assume $2q_1 < n_u$ to exclude this case. **Lemma 2** Under Assumption 3, for each set of q_1 actuators, among all its supersets with $2q_1$ actuators, at least one set is a superset of W_u .

Lemma 3 Under Assumption 3, among each set of $n_y - q_2$ sensors, at least $n_y - 2q_2 > 0$ sensors are attack-free.

Proof: Lemmas 2 and 3 follow trivially from Assumption 3.

Note that the existence of a UIO for each pair (J_u, J_s) with $\operatorname{card}(J_u) \leq 2q_1$ and $\operatorname{card}(J_s) \geq n_y - 2q_2$ means that if $W_u \subseteq J_u$ and $a_y^{J_s}(k) = 0$ for all $k \geq 0$, the estimation error $e_{J_{us}} = \hat{x}_{J_{us}} - x$ satisfies

$$|e_{J_{us}}| \le c_{J_{us}} \lambda_{J_{us}}^k |e_{J_{us}}(0)|, \tag{25}$$

for some $c_{J_{us}} > 0$ and $\lambda_{J_{us}} \in (0, 1)$, all $e_{J_{us}}(0) \in \mathbb{R}^n$, and $k \geq 0$. Let Assumption 3 be satisfied. We use a UIO for each pair (J_u, J_s) with $\operatorname{card}(J_u) = q_1$ and $\operatorname{card}(J_s) =$ $n_y - q_2$. Then, we use a UIO for each pair (S_u, S_s) with $S_u \subset \{1, \ldots, n_u\}$, $\operatorname{card}(S_u) = 2q_1$ and $S_s \subset \{1, \ldots, n_y\}$, $\operatorname{card}(S_s) = n_y - 2q_2$. Under Assumption 3, there exists at least one set \overline{J}_u with $\operatorname{card}(\overline{J}_u) = q_1$ such that $W_u \subseteq \overline{J}_u$ and at least one set \overline{J}_s with $\operatorname{card}(\overline{J}_s) = n_y - q_2$ such that $a_y^{\overline{J}_s}(k) = 0$ for all $k \geq 0$. Then, the estimate given by the UIO for $(\overline{J}_u, \overline{J}_s)$ is a correct estimate, and the estimates given by the UIOs for any (S_u, S_s) (denoted as $\hat{x}_{S_{us}}$), where $S_u \supset \overline{J}_u$, $\operatorname{card}(S_u) = 2q_1$ and $S_s \subset \overline{J}_s$, $\operatorname{card}(J_s) = n_y - 2q_2$, are consistent with $\hat{x}_{J_{us}}$. This motivates the following estimation strategy.

For each pair (J_u, J_s) with $\operatorname{card}(J_u) = q_1$ and $\operatorname{card}(J_s) = n_y - q_2$, define $\pi_{J_{us}}(k)$ as the largest deviation between $\hat{x}_{J_{us}}(k)$ and $\hat{x}_{S_{us}}(k)$ that is given by any pair (S_u, S_s) , where $S_u \supset J_u$ with $\operatorname{card}(S_u) = 2q_1$ and $S_s \subset J_s$ with $\operatorname{card}(S_s) = n_y - 2q_2$. That is,

$$\pi_{J_{us}}(k) := \max_{S_u \supset J_u, S_s \subset J_s} |\hat{x}_{J_{us}}(k) - \hat{x}_{S_{us}}(k)|, \qquad (26)$$

for all $k \ge 0$. Define the sequences $\sigma_u(k)$ and $\sigma_s(k)$ as

$$(\sigma_u(k), \sigma_s(k)) := \underset{J_u, J_s}{\operatorname{arg\,min}} \pi_{J_{us}}(k). \tag{27}$$

Then, as proven below, the estimate indexed by $(\sigma_u(k), \sigma_s(k))$:

$$\hat{x}(k) = \hat{x}_{\sigma_{us}(k)}(k), \tag{28}$$

is an exponential attack-free estimate of the system state. For simplicity and without generality, for all J and S, $z_{J_{us}}(0)$ and $z_{S_{us}}(0)$ are chosen such that $\hat{x}_{J_{us}}(0) = \hat{x}_{S_{us}}(0) = \hat{x}(0)$. The following result summarizes the ideas presented above.

Theorem 2 Consider system (1), observer (20), and the partial multi-observer estimator (26)-(28). Define the estimation error $e(k) := \hat{x}_{\sigma_{us}(k)}(k) - x(k)$ and let (c_3) - (c_4) and Assumptions 1,3 be satisfied; then, there exist positive constants $\bar{c} > 0$ and $\bar{\lambda} \in (0, 1)$ satisfying:

$$|e(k)| \le \bar{c}\bar{\lambda}^k |e(0)|, \tag{29}$$

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

Proof: Under Assumption 3, there exists at least one set J_u with $card(\bar{J}) = q_1$ such that $\bar{J}_u \supset W_u$, and at least one set \bar{J}_s

with $\operatorname{card}(\bar{J}_s) = n_y - q_2$ such that $a_y^{J_s}(k) = 0$ for all $k \ge 0$; then, there exist $c_{\bar{J}_{us}} > 0$ and $\lambda_{\bar{J}_{us}} \in (0, 1)$ satisfying

$$|e_{\bar{J}_{us}}(k)| \le c_{\bar{J}_{us}} \lambda_{\bar{J}_{us}}^k |e(0)|, \tag{30}$$

for all $e(0) \in \mathbb{R}^n$ and $k \ge 0$. Moreover, for any set $S_u \supset \overline{J}_u$ with $\operatorname{card}(S_u) = 2q_1$ and $S_s \subset \overline{J}_s$ with $\operatorname{card}(S_s) = n_y - 2q_2$, we have $S_u \supset W_u$ and $a_y^{S_s}(k) = 0$ for all $k \ge 0$; hence, there exist $c_{S_{us}} > 0$ and $\lambda_{S_{us}} \in (0, 1)$ such that

$$|e_{S_{us}}(k)| \le c_{S_{us}}\lambda_{S_{us}}^k|e(0)|,$$
 (31)

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$. Consider $\pi_{\bar{J}_{us}}$ in (26). Combining the above results, we have that

$$\begin{aligned} \pi_{\bar{J}_{us}}(k) &= \max_{S_u \supset \bar{J}_u, S_s \subset \bar{J}_s} |\hat{x}_{\bar{J}_{us}}(k) - \hat{x}_{S_{us}}(k)| \\ &= \max_{S_u \supset \bar{J}_u, S_s \subset \bar{J}_s} |\hat{x}_{\bar{J}_{us}}(k) - x(k) + x(k) - \hat{x}_{S_{us}}(k)| \\ &\leq |e_{\bar{J}_{us}}(k)| + \max_{S_u \supset \bar{J}_u, S_s \subset \bar{J}_s} |e_{S_{us}}(k)|, \end{aligned}$$

for all $k \ge 0$. From (30) and (31), we obtain

$$\pi_{\bar{J}_{us}}(k) \le 2c'_{\bar{J}_{us}}\lambda'^{k}_{\bar{J}_{us}}|e(0)|, \tag{32}$$

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

$$c'_{\bar{J}_{us}} := \max_{S_u \supset \bar{J}_u, S_s \subset \bar{J}_s} \left\{ c_{\bar{J}_{us}}, c_{S_{us}} \right\},$$
$$\lambda'_{\bar{J}_{us}} := \max_{S_u \supset \bar{J}_u, S_s \subset \bar{J}_s} \left\{ \lambda_{\bar{J}_{us}}, \lambda_{S_{us}} \right\}.$$

Note that $S_u \supset \overline{J}_u$, $\operatorname{card}(J_u) = 2q_1$, and $S_s \subset \overline{J}_s$, $\operatorname{card}(S_s) = n_y - 2q_2$. Then, from (27), we have $\pi_{\sigma_{us}(k)}(k) \le \pi_{\overline{J}_{us}}(k)$. By Lemmas 2 and 3, we know that there exists at least one set $\overline{S}_u \supset \sigma_u(k)$ with $\operatorname{card}(\overline{S}_u) = 2q_1$ and at least one set $\overline{S}_s \subset \sigma_s(k)$ with $\operatorname{card}(\overline{S}_s) = n_y - 2q_1$ such that $\overline{S}_u \supset W_u$ and $a_{\overline{S}_s}^{\overline{J}}(k) = 0$ for all $k \ge 0$. Hence, there exist $c_{\overline{S}_{us}} > 0$ and $\lambda_{\overline{S}_{us}} \in (0, 1)$ satisfying

$$|e_{\bar{S}_{us}}(k)| \le c_{\bar{S}_{us}}\lambda_{\bar{S}_{us}}^{k}|e(0)|,$$
(33)

for all $e(0) \in \mathbb{R}^n$ and $k \ge 0$. From (26), by construction

$$\pi_{\sigma_{us}(k)}(k) = \max_{\substack{S_u \supset \sigma_u(k), S_s \subset \sigma_s(k)}} |\hat{x}_{\sigma_{us}(k)}(k) - \hat{x}_{S_{us}}(k)|$$
$$\geq |\hat{x}_{\sigma_{us}(k)}(k) - \hat{x}_{\bar{S}_{us}}(k)|.$$

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

$$|e_{\sigma_{us}(k)}(k)| = |\hat{x}_{\sigma_{us}(k)}(k) - x(k)| = |\hat{x}_{\sigma_{us}(k)}(k) - \hat{x}_{\bar{S}_{us}}(k) + \hat{x}_{\bar{S}_{us}}(k) - x(k)| \leq |\hat{x}_{\sigma_{us}(k)}(k) - \hat{x}_{\bar{S}_{us}}(k)| + |e_{\bar{S}_{us}}(k)| \leq \pi_{\sigma_{us}(k)}(k) + |e_{\bar{S}_{us}}(k)| \leq \pi_{\bar{J}_{us}}(k) + |e_{\bar{S}_{us}}(k)|,$$
(34)

for all $k \ge 0$. Hence, from (32) and (33), we have

$$e_{\sigma_{us}(k)}(k)| \le \bar{c}\bar{\lambda}^k |e(0)|, \tag{35}$$

for all $e(0) \in \mathbb{R}^n$ and $k \ge 0$, where $\bar{c} = 3 \max\{c_{\bar{S}_{us}}, c'_{\bar{J}_{us}}\}$, $\bar{\lambda} = \max\{\lambda_{\bar{S}_{us}}, \lambda'_{\bar{J}_{us}}\}$. Inequality (35) is of the form (29), and the result follows.



Fig. 2. Estimated states \hat{x} converges to the true states x when $a_{u3}, a_{y2} \sim \mathcal{U}(-10, 10)$. Legend: \hat{x} (blue), true states (black)

Example 2: Consider a linear system subject to actuator and sensor attacks:

$$\begin{cases} x^{+} = \begin{bmatrix} 0.5 & 0 & 0.1 \\ 0.2 & 0.7 & 0 \\ 1 & 0 & 0.3 \end{bmatrix} x + \begin{bmatrix} 0.5 & 0 & 0.5 \\ 1 & 1 & 0.1 \\ 0 & 0 & 0.5 \end{bmatrix} (u + a_{u}), \\ y = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} x + a_{y}.$$
(36)

It can be verified that complete UIOs do not exist for any C^{J_s} with $\operatorname{card}(J_s) \leq 2$. However, a partial UIO exists for each pair (J_u, J_s) with $\operatorname{card}(J_u) \leq 2$ and $\operatorname{card}(J_s) \geq 2$; then, $2q_1 = 2$ and $4 - 2q_2 = 2$, i.e., $q_1 = q_2 = 1$. We let $W_u = \{3\}, W_y = \{2\}$, i.e., the third actuator and the second sensor are attacked, $u \sim \mathcal{U}(-1, 1)$, and $a_{u3}, a_{y2} \sim \mathcal{U}(-10, 10)$. We construct a partial UIO for each pair (J_u, J_s) with $\operatorname{card}(J_u) = 1, \operatorname{card}(J_s) = 3$ and each set (S_u, S_s) with $\operatorname{card}(S_u) = 2, \operatorname{card}(S_s) = 2$. Therefore, totally $\binom{3}{1} \times \binom{4}{3} + \binom{3}{2} \times \binom{4}{2} = 30$ partial UIOs are designed. We initiate the observers at $\hat{x}(0) = [0, 0, 0]^{\top}$. Estimator (20), (26)-(28) is used to construct $\hat{x}(k)$. The performance of the estimator is shown in Figure 2.

IV. ATTACK ISOLATION AND RECONSTRUCTION

Once we have an estimate $\hat{x}(k)$ of x(k), either using the complete multi-observer estimator 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. 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:

$$\begin{cases} \hat{x}^{+} = e^{+} + A(\hat{x} - e) + B(u + a_{u}), \\ & \downarrow \\ a_{u} = B_{left}^{-1}(\hat{x}^{+} - A\hat{x}) - u - B_{left}^{-1}(e^{+} - Ae), \end{cases}$$
(37)

$$\begin{cases} y = Cx + a_y = C\hat{x} - Ce + a_y, \\ & \downarrow \\ a_y = y - C\hat{x} + Ce. \end{cases}$$
(38)

First, consider the complete multi-observer in Section III-A. Let the estimation error dynamics characterized by (7)-(9) be given by

$$e^+ = f_1(e, x, a_y, a_u),$$
 (39)

where $f_1 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_u} \to \mathbb{R}^n$ denotes some nonlinear function. 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_u and a_y in (37) and (38) vanishes exponentially and therefore, the following attack estimate:

$$\hat{a}_u(k) = B_{left}^{-1}(\hat{x}(k) - A\hat{x}(k-1)) - u(k-1), \qquad (40)$$

and

$$\hat{a}_y(k) = y(k) - C\hat{x}(k),$$
 (41)

exponentially reconstruct the attack signals $a_u(k-1)$ and $a_u(k)$, i.e.,

$$\lim_{k \to \infty} (\hat{a}_u(k) - a_u(k-1)) = 0,$$
(42)

and

$$\lim_{k \to \infty} (\hat{a}_y(k) - a_y(k)) = 0.$$
(43)

Then, for sufficiently large k, the sparsity pattern of $\hat{a}_u(k)$ and $\hat{a}_u(k)$ can be used to isolate attacks, i.e.,

$$\hat{W}_u(k) = \operatorname{supp}(\hat{a}_u(k)), \tag{44}$$

and

$$\hat{W}_y(k) = \operatorname{supp}(\hat{a}_y(k)), \tag{45}$$

where $\hat{W}_u(k)$ denotes the set of isolated attacked actuators, and $\hat{W}_y(k)$ denotes the set of isolated attacked sensors. Note that we can only estimate a_u from \hat{x}^+ and e^+ , which implies that we always have, at least, one-step delay for actuator attacks isolation.

Next, consider the partial multi-observer estimator given in Section III-B. In this case, the attack vector a_u and a_y can also be written as (37) and (38), and the estimation error dynamics is given by some nonlinear difference equation characterized by the estimator structure in (26)-(28). Let the estimation error dynamics be given by

$$e^+ = f_2(e, x, a_y, a_u),$$
 (46)

for some nonlinear function $f_2 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_u} \to \mathbb{R}^n$. In Theorem 2, we have proved that *e* converges to the origin exponentially. Hence, the attack estimate in (40) and (41) exponentially reconstructs the attack signals. Again, the sparsity pattern of $\hat{a}_u(k)$ and $\hat{a}_y(k)$ can be used to isolate actuator and sensor attacks using (44) and (45).



Fig. 3. Estimated actuator attacks \hat{a}_u^+ converges to a_u when $a_u, a_{y3} \sim \mathcal{U}(-10, 10)$. Legend: \hat{a}_u^+ (blue), a_u (black).



Fig. 4. Estimated sensor attacks \hat{a}_y converges to a_y when $a_u, a_{y3} \sim \mathcal{U}(-10, 10)$. Legend: \hat{a}_y (blue), a_y (black).

Example 3: Consider system (17) and the complete multiobserver estimator in Example 1. Let $W_u = \{1\}$, $W_y = \{3\}$, $u \sim \mathcal{U}(-1, 1)$, $a_u, a_{y2} \sim \mathcal{U}(-10, 10)$, and $(x_1(0), x_2(0)) \sim \mathcal{N}(0, 1^2)$. We obtain $\hat{a}_u(k)$ and $\hat{a}_y(k)$ from (40) and (41). The reconstructed attack signals are depicted in Figures 3-4. By checking the sparsity of these signals, actuator and sensor 3 are isolated as attacked.

Example 4: Here we consider system (36) and the partial multi-observer estimator in Example 2. Let $W_u = \{3\}$, $W_y = \{2\}$, $(u_1, u_2, u_3) \sim \mathcal{U}(-1, 1)$, $a_{u_3}, a_{y_2} \sim \mathcal{U}(-10, 10)$, and $(x_1(0), x_2(0), x_3(0)) \sim \mathcal{N}(0, 1^2)$. We obtain $\hat{a}_u(k)$ and $\hat{a}_y(k)$ from (40) and (41). The reconstructed attacks are shown in Figures 5-6. In this case, using sparsity of the estimated attacks, actuator 3 and sensor 2 are correctly isolated.

V. CONTROL

In this section, we introduce a method to use the proposed multi-observer estimators to asymptotically stabilize the system dynamics.

A. Sensor attacks only

We first consider the case when only sensors are attacked and actuators are attack-free. Then, the system is given by

$$\begin{cases} x^+ = Ax + Bu, \\ y = Cx + a_y. \end{cases}$$
(47)

Let $u = K\hat{x}$, where \hat{x} is the estimate given by the complete multi-observer estimator in Section III-A or the partial multi-



Fig. 5. Estimated actuator attacks \hat{a}_u^+ converges to a_u when $a_{u3}, a_{y2} \sim \mathcal{U}(-10, 10)$. Legend: \hat{a}_u^+ (blue), a_u (black).



Fig. 6. Estimated sensor attacks \hat{a}_y converges to a_y when $a_{u3}, a_{y2} \sim \mathcal{U}(-10, 10)$. \hat{a}_y (blue), a_y (black).

observer estimator in Section III-B, and K is chosen such that A + BK is Schur. Then, the closed-loop system is given by

$$x^+ = Ax + BK\hat{x},\tag{48}$$

or in terms of the estimation error as

$$x^{+} = Ax + B(K(\hat{x} - x + x)), = (A + BK)x + BKe.$$
(49)

For the complete multi-observer estimator, let the estimation error dynamics be given by

$$e^+ = f_1(e, x, a_y),$$
 (50)

for some nonlinear function $f_1 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n_y} \to \mathbb{R}^n$. For the partial multi-observer estimator, let the estimation error dynamics be given by

$$e^+ = f_2(e, x, a_y),$$
 (51)

for some nonlinear function $f_2 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n_y} \to \mathbb{R}^n$. Since A + BK is Schur, the closed-loop dynamics (49) is Input-to-State Stable (ISS) with respect to input e(k) and some linear



Fig. 7. Controlled states when $a_{y2} \sim \mathcal{U}(-10, 10)$.

gain, see [26]. Moreover, in Theorems 1 and 2, we have proved that (50) and (51) are exponentially stable uniformly in x(k) and $a_y(k)$. The latter and ISS of the system dynamics imply that $\lim_{k\to\infty} x(k) = 0$ [26].

Example 5: Consider the open-loop unstable system

$$\begin{cases} x^{+} = \begin{bmatrix} 1.2 & 0.5 \\ 0.2 & 0.7 \end{bmatrix} x + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} K \hat{x}, \\ y = \begin{bmatrix} 1 & 1 & 3 & 2 \\ 3 & 1 & 2 & 1 \end{bmatrix}^{\top} x + a_{y}. \end{cases}$$
(52)

It can be verified that a UIO of the form (2) exists for each $J_s \subset \{1, 2, 3, 4\}$ with $\operatorname{card}(J_2) \geq 2$; then, 4 - 2q = 2 and q = 1. We let $W_y = \{2\}$ and $a_{y2} \sim \mathcal{U}(-10, 10)$. We construct $\binom{4}{3} + \binom{4}{2} = 10$ UIOs initialized at $\hat{x}(0) = [0, 0]^{\top}$ and let

$$K = \begin{bmatrix} -1.2 & 0.7 \\ -0.2 & -0.7 \end{bmatrix}$$

We use the complete multi-observer in Section III-A to estimate the state. The state of the closed-loop system is shown in Figure 7.

B. Sensor and actuator attacks

Here, we consider sensor and actuator attacks. 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 that correspond to the isolated actuators, and closing the loop with a multi-observer based output dynamic feedback controller, see Figure 8. We introduce a switching signal $\rho(k) \subseteq \{1, \ldots, n_u\}$, containing the isolated attack-free actuators, i.e., $\rho(k) := \{1, \ldots, n_u\} \setminus \hat{W}_u(k)$. This $\rho(k)$ is used to denote actuators that are switched on. That is, $\rho(k) = J$ if the subset $J \subseteq \{1, \ldots, n_u\}$ of actuators are switched on and the remaining actuators are switched off at time k. Again, let B be partitioned as $B = [b_1, \ldots, b_i, \ldots, b_{n_u}]$. After switching off the subset $\{1, \ldots, n_u\} \setminus \rho(k)$ of actuators, system (1) is written as follows

$$\begin{cases} x^{+} = Ax + b_{\rho(k)}(u^{\rho(k)} + a_{u}^{\rho(k)}), \\ y = Cx + a_{y}, \end{cases}$$
(53)



Fig. 8. Estimation, isolation, and control diagram

where $b_{\rho(k)}$ is the matrix whose columns are $b_i, i \in \rho(k)$, vectors $u^{\rho(k)}$ and $a_u^{\rho(k)}$ are the inputs and attacks corresponding to the switched-on actuators, respectively. We first consider the case when the complete multi-observer estimator in Section III-A exists, i.e., \hat{x} is generated by (7)-(9). We estimate $\hat{a}_u(k)$ using (40) and obtain $\hat{W}_u(k)$ from (44). Then, we switch off the set \hat{W}_u of actuators by letting $\rho(k) = \bar{J}(k) = \{1, \ldots, n_u\} \setminus \hat{W}_u(k)$. Since $a_i(k) = 0, i \in \bar{J}(k)$, system (53) has the following form:

$$x^{+} = Ax + b_{\bar{J}(k)} u^{J(k)}$$
(54)

where $u^{\bar{J}(k)} \in \mathbb{R}^{\operatorname{card}(\bar{J}(k))}$ is the set of isolated attack-free inputs. Let $0 < q^* < n_u$ be the largest integer such that (A, b_J) is stabilizable for each set $J \subset \{1, \ldots, n_u\}$ with $\operatorname{card}(J) \ge n_u - q^*$ where b_J denotes a matrix whose columns are b_i for $i \in J$. We assume that at most q^* actuators are attacked. It follows that $n_u - q^* \le \operatorname{card}(\bar{J}(k)) \le n_u$. We assume the following.

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

$$x^{+} = (A + b_{\bar{J}(k)} K_{\bar{J}(k)}) x + b_{\bar{J}(k)} K_{\bar{J}(k)} e, \qquad (55)$$

is ISS with input e for $b_{\bar{J}(k)}$ arbitrarily switching among all $b_{J'}$ with $J \subset J' \subset \{1, \ldots, n_u\}$ and $n_u - q^* \leq \operatorname{card}(J') \leq n_u$.

Remark 2 We do not give a method for designing the linear switching state feedback controller $u^{\overline{J}(k)} = K_{\overline{J}(k)}x$. Standard results for designing switching controllers, for instance results in [27] and references therein, can be used to design controllers satisfying Assumption 4.

By switching off the set $\hat{W}_u(k)$ of actuators at time k, using the controller designed for the set $\bar{J}(k)$, and letting $u^{\bar{J}(k)} = K_{\bar{J}(k)}\hat{x}$, the closed-loop system can be written as (55) with estimation error $e = \hat{x} - x$ generated by some nonlinear difference equation (39). Because in Theorem 1, we have proved that e(k) converges to zero exponentially uniformly in x(k), $a_y(k)$ and $a_u(k)$, the error e(k) in (55) is a vanishing perturbation. Hence, under Assumption 4, it follows that $\lim_{k\to\infty} x(k) = 0$.

Next, assume that a complete multi-observer estimator does not exist but a partial multi-observer estimator exists (Section III-B), i.e., \hat{x} is generated from (26)-(28) and $q_1 \leq q^*$. We assume that at most q_1 actuators are attacked. We construct $\hat{x}(k)$ from (26)-(28), estimate $\hat{a}_u(k)$ using (40), and obtain $\hat{W}_u(k)$ from (44). After switching off the set $\hat{W}_u(k)$ of actuators, the system has the form (54) with $n_u - q_1 \leq \operatorname{card}(\bar{J}(k)) \leq n_u$. We assume the following.

Assumption 5 For any subset J with cardinality $\operatorname{card}(J) = n_u - q_1$, there exists a linear switching state feedback controller $u^{\overline{J}(k)} = K_{\overline{J}(k)}x$ such that the closed-loop dynamics (55) is ISS with respect to e for $b_{\overline{J}(k)}$ arbitrarily switching among all $b_{J'}$ with $J \subset J' \subset \{1, \ldots, n_u\}$ and $n_u - q_1 \leq \operatorname{card}(J') \leq n_u$.

Using the controller designed for the set $\bar{J}(k)$, and letting $u^{\bar{J}(k)} = K_{\bar{J}(k)}\hat{x}$, the closed-loop dynamics can be written in the form (55). Then, in this case, e(k) is generated by some nonlinear difference equation of the form (46). Under Assumption 5, the closed-loop dynamics (55) is ISS with input e(k), see [26]. Moreover, in Theorem 2, we have proved that e(k) converges to the origin exponentially uniformly in x(k), $a_u(k)$ and $a_y(k)$. The latter and ISS of the system dynamics imply that $\lim_{k\to\infty} x(k) = 0$ [26].

Example 6: Consider the following system:

$$\begin{cases} x^{+} = \begin{bmatrix} 0.5 & 0 & 0.1 \\ 0.2 & 1.7 & 0 \\ 1 & 0 & 0.3 \end{bmatrix} x + \begin{bmatrix} 0.5 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} (u + a_{u}), \\ y = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} x + a_{y}.$$
(56)

Since (A, b_i) is stabilizable for $i \in \{1, 2, 3\}$, we have $q^* = 2$. It can be verified that there does not exist a complete UIO for any $S_s \subset \{1, 2, 3, 4\}$ with $\operatorname{card}(S_s) = 2$, but partial UIOs exists for each pair (J_u, J_s) with $\operatorname{card}(J_u) \leq 2$ and $\operatorname{card}(J_s) \geq 2$; then, we have $q_1 = q_2 = 1$ and $q_1 < q^*$. We let $W_u = \{3\}$, $W_y = \{2\}$, and $a_{u3}, a_{y2} \sim \mathcal{U}(-10, 10)$. We construct $\binom{3}{1} \times \binom{4}{3} + \binom{3}{2} \times \binom{4}{2} = 30$ UIOs and use the design method given in [27] to build controllers for actuators $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, $\{1, 2, 3\}$. Then, we use the partial 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 9.

VI. CONCLUSION

We have addressed the problem of state estimation, attack isolation, and control for discrete-time linear time-invariant (LTI) systems under (potentially unbounded) actuator and sensor false data injection attacks. Using a bank of Unknown Input Observers (UIOs), we have proposed an estimator that reconstructs the system states and the attack signals. We use these estimates to isolate attacks and control the system. We propose an effective technique to stabilize the system by switching off the isolated actuators. Simulation results are provided to illustrate our results.



Fig. 9. State trjectories when $a_{u3}, a_{y2} \sim \mathcal{U}(-10, 10)$.

REFERENCES

- 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.
- [2] 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.
- [3] 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.
- [4] 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)*, pp. 5906–5911, 2016.
- [5] Y. Shoukry, P. Nuzzo, A. Puggelli, A. 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] C. Murguia and J. Ruths, "Characterization of a CUSUM model-based sensor attack detector," in *IEEE 55th Conference on Decision and Control, CDC*, 2016.
- [10] 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.
- [11] M. S. Chong, M. Wakaiki, and P. Hespanha, "Observability of linear systems under adversarial attacks *," *Proc. American Control Conf.* (ACC), pp. 2439–2444, 2015.
- [12] Z. Tang, M. Kuijper, M. S. Chong, I. Mareels, and C. Leckie, "Linear system security-detection and correction of adversarial sensor attacks in the noise-free case," *Automatica*, vol. 101, pp. 53–59, 2019.
- [13] Q. Hu, D. Fooladivanda, Y. H. Chang, and C. J. Tomlin, "Secure State Estimation and Control for Cyber Security of the Nonlinear Power Systems," *IEEE Transactions on Control of Network Systems*, pp. 1310 – 1321, 2017.
- [14] 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.
- [15] Y. Shoukry, P. Nuzzo, N. Bezzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state reconstruction in differentially flat systems under sensor attacks using satisfiability modulo theory solving,"

54th IEEE Conference on Decision and Control, CDC, pp. 3804–3809, 2015.

- [16] T. Yang, C. Murguia, M. Kuijper, and D. Nešić, "A robust circlecriterion observer-based estimator for discrete-time nonlinear systems in the presence of sensor attacks," *IEEE 57th Conference on Decision* and Control, CDC, 2018.
- [17] T. Yang, C. Murguia, M. Kuijper, and D. Nešić, "Attack detection and isolation for discrete-time nonlinear systems," 2018 Australian & New Zealand Control Conference (ANZCC), 2018.
- [18] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," *American Control Conference*, 2015.
- [19] S. M. Djouadi, A. M. Melin, E. M. Ferragut, J. A. Laska, J. Dong, and A. Drira, "Finite energy and bounded actuator attacks on cyber-physical systems," 2015 European Control Conference, ECC 2015, pp. 3659– 3664, 2015.
- [20] 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), pp. 3412–3417, 2012.
- [21] M. Showkatbakhsh, Y. Shoukry, R. H. Chen, S. Diggavi, and P. Tabuada, "An SMT-based approach to secure state estimation under sensor and actuator attacks," 2017 IEEE 56th Annual Conference on Decision and Control, CDC 2017, 2017.
- [22] M. Yadegar, N. Meskin, and W. M. Haddad, "An Output Feedback Adaptive Control Architecture for Mitigating Actuator Attacks in Cyberphysical Systems," *IEEE Transactions on Control of Network Systems*, vol. 62, no. 11, pp. 6058–6064, 2017.
- [23] E. D. Sontag, "Input to state stability: Basic concepts and results," *Lecture Notes in Mathematics*, vol. 1932, pp. 163–220, 2008.
- [24] H. K. Khalil, Nonlinear systems; 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [25] S. X. Ding, Model-based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools. Springer, 2013.
- [26] Z.-P. Jiang and Y. Wang, "Input-to-state stability for discrete-time nonlinear systems," *Automatica*, vol. 37, no. 6, pp. 857–869, 2001.
- [27] 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.