Simultaneous Input and State Set-Valued Observers with Applications to Attack-Resilient Estimation

Sze Zheng Yong

Abstract—In this paper, we present a fixed-order set-valued observer for linear discrete-time bounded-error systems that simultaneously finds bounded sets of compatible states and unknown inputs that are optimal in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification. We also analyze the necessary and sufficient conditions for the stability of the observer and its connection to a system property known as strong detectability. Next, we show that the proposed set-valued observer can be used for attack-resilient estimation of state and attack signals when cyber-physical systems are subject to false data injection attacks on both actuator and sensor signals. Moreover, we discuss the implication of strong detectability on resilient state estimation and attack identification. Finally, the effectiveness of our set-valued observer is demonstrated in simulation, including on an IEEE 14-bus electric power system.

I. INTRODUCTION

Cyber-physical systems (CPS) are systems in which computational and communication elements collaborate to control physical entities. Such systems include the power grid, autonomous vehicles, medical devices, etc. Most of these systems are safety-critical and if compromised or malfunctioning, can cause serious harm to the controlled physical entities and the people operating or utilizing them. Recent incidents of attacks on CPS, e.g., the Ukrainian power grid, the Maroochy water service and an Iranian nuclear plant [1]–[3] highlight a need for CPS security and for new designs of resilient estimation and control.

In particular, false data injection attack is one of the most serious forms of attacks on CPS, where malicious and strategic attackers intrude and inject false data signals into the sensor measurements and actuator signals with the goal of causing harm, energy theft etc. Given the strategic nature of these false data injection signals, they are not well-modeled by a zero-mean, Gaussian white noise nor by signals with known bounds. Hence, traditional Kalman filtering and unknown input observers do not apply. Nevertheless, reliable estimates of states and unknown inputs are valuable and needed for purposes of resilient control, attack identification, etc. Similar state and input estimation problems can be found across a wide range of disciplines, from the estimation of mean areal precipitation [4] to fault detection and diagnosis [5] to input estimation in physiological systems [6].

Literature review. Much of the research focus on simultaneous input and state estimation has been on obtaining point estimates for deterministic systems with unknown inputs via asymptotic and sliding mode observers (e.g., [7]–[9]) or for stochastic systems with unknown inputs via unbiased minimum-variance estimation (e.g., [10]–[14]). These methods do not directly apply to bounded-error models, i.e., uncertain dynamic systems with set-valued uncertainties, where instead, the sets of states and unknown inputs that are compatible/consistent with sensor observations are desired. Similarly, while $\mathcal{H}_2$, $\mathcal{H}_\infty$, $\ell_1$ filters (e.g., [15]–[17]) can deal with bounded modeling errors, only point estimates of states are obtained in addition to the fact that it is unsuited to handle large unknown inputs.

In contrast, set-membership or set-valued state observers are capable of estimating the set of compatible states and are preferable to stochastic estimation when hard accuracy bounds are important [18], e.g., to guarantee safety. Since its conception, it was apparent that characterizing the set of states that are compatible with measurements is in general computationally intensive. The complexity of optimal observers [19] grows with time, and also for methods based on an $\ell_1$ model matching problem [20] and polyhedral set computation using Fourier-Motzkin elimination [21]. Thus, fixed-order recursive filters were designed with equalized performance (i.e., with invariant estimation errors) for superstable systems in [18], [22]. However, all these set-membership approaches can only compute the set of compatible states and do not apply when the unknown input signals have unknown bounds, as is often required in attack-resilient estimation where the attack signals are malicious and strategic.

In the context of attack-resilient estimation against false data injection attacks, numerous approaches were proposed for deterministic systems (e.g., [23]–[26]), stochastic systems (e.g., [27]–[29]) and bounded-error systems [30]–[32], but they share the common theme of only obtaining point estimates. In particular, error bounds were computed in [30] for only the initial state and in [31] with the assumption of zero initial state and without optimality considerations. More importantly, only sensor attacks are considered and set-valued estimates of state and attack signals are not computed.

Contributions. The goal of this paper is to bridge the gap between set-valued state estimation without unknown inputs and point-valued state and unknown input estimation. We propose a fixed-order set-valued observer for linear discrete-time bounded-error systems that simultaneously finds bounded sets of states and unknown inputs that contain the true state and unknown input, are compatible/consistent with measurement outputs and are optimal in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification. In addition, we provide the necessary and sufficient...
conditions for observer stability and boundedness of the set-valued estimates, which we show is closely related to a system property known as strong detectability.

We further show that the proposed set-valued observer is applicable for achieving attack-resiliency in cyber-physical systems against false data injection attacks on both actuator and sensor signals. Specifically, the set-valued observer can compute the sets of states and attack signals that are compatible with measurements, where the latter enables not only attack detection but also identification. Moreover, we discuss the implication of strong detectability on resilient state estimation and attack identification. Finally, the effectiveness of our set-valued observer is demonstrated using a benchmark system and an IEEE 14-bus electric power system.

**Notation.** $\mathbb{R}^n$ denotes the $n$-dimensional Euclidean space, $\mathbb{C}$ the field of complex numbers and $\mathbb{N}$ nonnegative integers. For a vector $v \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{p \times q}$, $\|v\| \triangleq \sqrt{v^\top v}$ and $\|M\|$ denote their (induced) 2-norm. Moreover, the transpose, inverse, Moore-Penrose pseudoinverse and rank of $M$ are given by $M^\top$, $M^{-1}$, $M^\dagger$ and $\text{rk}(M)$. For a symmetric matrix $S$, $S \succ 0$ ($S \succeq 0$) is positive (semi-) definite.

**II. PROBLEM STATEMENT**

**System Assumptions.** Consider the linear time-invariant discrete-time bounded-error system

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

$$y_k = Cx_k + Du_k + Hd_k + v_k,$$

where $x_k \in \mathbb{R}^n$ is the state vector at time $k \in \mathbb{N}$, $u_k \in \mathbb{R}^m$ is a known input vector, $d_k \in \mathbb{R}^p$ is an unknown input vector, and $y_k \in \mathbb{R}^l$ is the measurement vector. The process noise $w_k \in \mathbb{R}^n$ and the measurement noise $v_k \in \mathbb{R}^l$ are assumed to be bounded, with $\|w_k\| \leq \eta_w$ and $\|v_k\| \leq \eta_v$ (thus, they are $\ell_\infty$ sequences). We also assume an estimate $\hat{x}_0$ of the initial state $x_0$ is available, where $\|\hat{x}_0 - x_0\| \leq \delta_0$. The matrices $A$, $B$, $C$, $D$, $G$, $H$ and $W$ are known and of appropriate dimensions, where $G$ and $H$ are matrices that encode the locations through which the unknown input or attack signal can affect the system dynamics and measurements. Note that no assumption is made on $H$ to be either the zero matrix (no direct feedthrough), or to have full column rank when there is direct feedthrough. Without loss of generality, we assume that $\text{rk}[G^\top, H^\top] = p$, $n \geq l \geq 1$, $l \geq p \geq 0$ and $m \geq 0$.

**Unknown Input (or Attack) Signal Assumptions.** The unknown inputs $d_k$ are not constrained to be a signal of any type (random or stochastic) nor to follow any model, thus no prior ‘useful’ knowledge of the dynamics of $d_k$ is available (independent of $\{d_k\}$ $\forall k \notin \ell$, $\{w_k\}$ and $\{v_k\}$ $\forall l$). We also do not assume that $d_k$ is bounded or has known bounds and thus, $d_k$ is suitable for representing adversarial attack signals.

The simultaneous input and state set-valued observer design problem is twofold and can be stated as follows:

1) Given a linear discrete-time bounded-error system with unknown inputs (1), design an optimal and stable filter that simultaneously finds bounded sets of compatible states and unknown inputs in the minimum $H_\infty$-norm sense, i.e., with minimum average power amplification.  

2) Develop an attack-resilient set-valued observer for system (1) that computes a bounded set of state estimates that contains the true state and identifies the set of compatible attack signals irrespective of the magnitude of false data injection attacks on its actuators and sensors. In addition, recommend preventative attack mitigation strategies based on detectability conditions.

**III. PRELIMINARY MATERIAL**

**A. System Transformation**

We first carry out a transformation of the system to decouple the output equation into two components, one with a full rank direct feedthrough matrix and the other without direct feedthrough. Note, however, that this similarity transformation is different from the one in [14], which is no longer applicable as it was based on the noise error covariance.

Let $p_H \triangleq \text{rk}(H)$. Using singular value decomposition, we rewrite the direct feedthrough matrix $H$ as

$$H = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} 0 & V_1^\top \\ 0 & V_2^\top \end{bmatrix},$$

where $\Sigma \in \mathbb{R}^{p_H \times p_H}$ is a diagonal matrix of full rank, $U_1 \in \mathbb{R}^{l \times p_H}$, $U_2 \in \mathbb{R}^{l \times (l - p_H)}$, $V_1 \in \mathbb{R}^{p_H \times p_H}$ and $V_2 \in \mathbb{R}^{p_H \times (p - p_H)}$, while $U \triangleq [U_1 \ U_2]$ and $V \triangleq [V_1 \ V_2]$ are unitary matrices. When there is no direct feedthrough, $\Sigma$, $U_1$ and $V_1$ are empty matrices$^a$, and $U_2$ and $V_2$ are arbitrary unitary matrices.

Then, we define two orthogonal components of the unknown input given by

$$d_{1,k} = V_1^\top d_k, \quad d_{2,k} = V_2^\top d_k.$$ 

Since $V$ is unitary, $d_k = V_1d_{1,k} + V_2d_{2,k}$ and the system (1) can be rewritten as

$$x_{k+1} = Ax_k + Bu_k + GV_1d_{1,k} + GV_2d_{2,k} + Ww_k,$$

$$y_k = Cx_k + Du_k + HV_1d_{1,k} + HV_2d_{2,k} + v_k,$$

where $G_1 \triangleq GV_1$, $G_2 \triangleq GV_2$ and $H_1 \triangleq HV_1 = U_1\Sigma$. Next, we decouple the output $y_k$ using a nonsingular transformation $T = \begin{bmatrix} T_1 \ T_2 \end{bmatrix}^\top \triangleq U^\top = \begin{bmatrix} U_1 & U_2 \end{bmatrix}^\top$ to obtain $z_{1,k} \in \mathbb{R}^{p_H}$ and $z_{2,k} \in \mathbb{R}^{l - p_H}$ given by

$$z_{1,k} = T_1y_k = U_1^\top y_k = C_1x_k + D_1u_k + \Sigma d_{1,k} + v_{1,k},$$

$$z_{2,k} = T_2y_k = U_2^\top y_k = C_2x_k + D_2u_k + v_{2,k},$$

where $C_1 \triangleq U_1^\top C$, $C_2 \triangleq U_2^\top C$, $D_1 \triangleq U_1^\top D$, $D_2 \triangleq U_2^\top D$, $v_{1,k} \triangleq U_1^\top v_k$ and $v_{2,k} \triangleq U_2^\top v_k$. This transform is also chosen such that $\| \begin{bmatrix} v_{1,k}^\top \ v_{2,k}^\top \end{bmatrix}^\top \| = \|U^\top v_k\| = \|v_k\|$.

**B. Input and State Detectability (a.k.a. Strong Detectability)**

Similar to the stability of the deterministic and stochastic input and state observers/filters, we will show in Section IV-B that the stability of the set-valued observer is directly related to the notion of strong detectability. Without loss of generality, we assume that $B = 0$ and $D = 0$ in this section, since $u_k$ is known.

$^a$ We adopt the convention that the inverse of an empty matrix is also an empty matrix and assume that operations with empty matrices are possible.
Definition 1 (Strong detectability). The linear system (1) is strongly detectable if
\[ y_k = 0 \quad \forall k \geq 0 \quad \text{implies} \quad x_k \to 0 \quad \text{as} \quad k \to \infty \]
for all initial states and input sequences \( \{d_k\}_{k \in \mathbb{Z}} \).

Definition 2 (Invariant Zeros). The invariant zeros \( z \) of the Rosenbrock system matrix \( R_S(z) := [zI - A - G \ C \ H] \) of system (1) are the finite values of \( z \) for which \( R_S(z) \) drops rank, i.e., \( \text{rk}(R_S(z)) < \text{nrank}(R_S) \), where \( \text{nrank}(R_S) \) is the normal rank (maximal rank over \( z \in \mathbb{C} \)) of \( R_S(z) \).

Theorem 1 (Strong detectability). A linear time-invariant discrete-time system is strongly detectable if and only if either of the following holds for \( z \in \mathbb{C}, \ |z| \geq 1 \):

(i) \( \text{rk} R_S(z) = \text{rk} [zI - A - G \ C \ H] = n + p \),

(ii) \( \text{rk} R_S(z) = \text{rk} [zI - A - G \ C \ H] = n + p - p_H \),

(iii) \( \text{rk} \tilde{R}_S(z) = \text{rk} [zI - \tilde{A} - G \ C \ H] = n + p - p_H \),

(iv) \( \text{rk} \tilde{R}_S(z) = \text{rk} [zI - \tilde{A} - G \ C \ H] = n + p - p_H \).

where \( \tilde{A} \equiv A - G \Sigma^{-1} C \) and \( \tilde{A} \equiv (I - G \tilde{M} \tilde{C} \tilde{A})A \) for any \( \tilde{M} \in \mathbb{R}^{(p - p_H) \times (p - p_H)} \). The above conditions are equivalent to the system being minimum-phase (i.e., the invariant zeros of \( R_S(z) \) in Condition (i) are stable).

Moreover, strong detectability implies that the pairs \( (A, C) \) and \( (\tilde{A}, \tilde{C}) \) are detectable, and if \( l = p \), then strong detectability implies that the pairs \( (A, G) \) and \( (\tilde{A}, \tilde{G}) \) are stabilizable.

Proof. The equivalence of Conditions (i) and (ii) with strong detectability in Definition 1 can be found in [14]. Thus, it is sufficient to show the equivalence of Conditions (ii), (iii) and (iv) using the following conditions for all \( z \in \mathbb{C}, \ |z| \neq 1 \):

\[ \text{rk} [zI - \tilde{A} - G \ C \ H] = \text{rk} [zI - A - G \ C \ H] = n + p - p_H \],
\[ \text{rk} [zI - \tilde{A} - G \ C \ H] = \text{rk} [zI - A - G \ C \ H] = n + p - p_H \].

Finally, comparing Conditions (i)–(iv) to the PBH rank test for detectability (and stabilizability), we see that strong detectability implies that \( (A, C) \), \( (\tilde{A}, \tilde{C}) \) and \( (A, G) \) and \( (\tilde{A}, \tilde{G}) \) are detectable, and that \( (A, G) \), \( (\tilde{A}, \tilde{G}) \) are stabilizable if \( l = p \).

IV. FIXED-ORDER SIMULTANEOUS INPUT AND STATE SET-VALUED OBSERVERS

A. Set-Valued Observer Design

We consider a recursive three-step set-valued observer design (similar to [12], [14]), composed of an unknown input estimation step that uses the current measurement and the set of compatible states to estimate the set of compatible unknown inputs, a time update step that propagates the compatible set of states based on the system dynamics, and a measurement update step that updates the set of compatible states using the current measurement. In brief, our goal is to design a recursive three-step set-valued observer of the form:

**Unknown Input Estimation:**
\[ \hat{d}_{k-1} = F_d(x_{k-1}, u_k), \]

**Time Update:**
\[ \hat{X}_k^* = F_s^*(x_{k-1}, \hat{d}_{k-1}, u_k), \]

**Measurement Update:**
\[ \hat{X}_k = F_x(x_{k-1}^*, u_k, y_k), \]

where \( F_d, F_s^* \) and \( F_x \) are the to-be-designed set mappings while \( \hat{d}_{k-1}, \hat{X}_k^* \) and \( \hat{X}_k \) are the sets of compatible unknown inputs at time \( k - 1 \), propagated and updated states at time \( k \), respectively. Note that we have a (one-step) delayed estimate of \( \hat{D}_{k-1} \) because it is the only estimate we can obtain in light of (5) since \( \hat{d}_{k-1} \) does not affect \( z_{1,k} \) and \( z_{2,k} \), and hence, cannot be estimated from \( y_k \). The reader is referred to a previous work [13] for a detailed discussion on when a delay is absent or when further delays are expected.

Since the complexity of optimal observers grows with time [19]–[21], we will only consider fixed-order recursive filters as in [18], [22], where set-valued estimates are of the form:

\[ \hat{d}_{k-1} = \{d \in \mathbb{R}^p : |d_k - \hat{d}_{k-1}| \leq \delta_k^d \}, \]
\[ \hat{X}_k^* = \{x \in \mathbb{R}^n : |x_k - x_{k-1}^*| \leq \delta_k^x \}, \]
\[ \hat{X}_k = \{x \in \mathbb{R}^n : |x_k - \hat{x}_{k-1}^*| \leq \delta_k^x \}, \]

where \( \delta_k^d \) and \( \delta_k^x \) denote the estimation errors to balls of norm \( \delta \). In this case, the observer design problem is reduced to finding the centroids \( \hat{d}_{k-1}^*, \hat{x}_{k-1}^* \) and \( \hat{x}_{k-1} \) as well as the radii \( \delta_k^d \) and \( \delta_k^x \) of the sets \( \hat{D}_{k-1}, \hat{X}_{k-1}^* \) and \( \hat{X}_k \), respectively.

We further limit our attention to observers for the centroids \( \hat{d}_{k-1}^*, \hat{x}_{k-1}^* \) and \( \hat{x}_{k-1} \) that belong to the class of three-step recursive filters given in [12], [14], defined as follows for each time \( k \) (with \( \hat{x}_{0} = \hat{x}_0 \)):

**Unknown Input Estimation:**
\[ \hat{d}_{1,k} = M_1(z_{1,k} - C\hat{x}_{k-1} - D_1u_k), \]
\[ \hat{d}_{2,k} = M_2(z_{2,k} - C\hat{x}_{k-1} - D_2u_k), \]
\[ \hat{d}_{k-1} = V_1\hat{d}_{1,k-1} + V_2\hat{d}_{2,k-1}, \]

**Time Update:**
\[ \hat{x}_{k-1} = A\hat{x}_{k-1} + Bu_k + G_1\hat{d}_{1,k-1}, \]
\[ \hat{x}_{k-1}^* = \hat{x}_{k-1}^* + G_2\hat{d}_{2,k-1}, \]

**Measurement Update:**
\[ \hat{x}_{k-1}^* = \hat{x}_{k-1}^* + L(y_k - C\hat{x}_{k-1} - D_1u_k), \]
\[ \hat{x}_{k-1}^* = \hat{x}_{k-1}^* + L(y_k - C\hat{x}_{k-1} - D_2u_k), \]

where \( L \in \mathbb{R}^{n \times l}, \hat{L} \equiv LU_2 \in \mathbb{R}^{n \times (l - p_H)}, \) and \( M_1 \in \mathbb{R}^{p_H \times p_H} \) and \( M_2 \in \mathbb{R}^{(p - p_H) \times (p - p_H)} \) are observer gain matrices that are chosen in the following lemma and theorem to minimize the “volume” of the set of compatible states and unknown inputs, quantified by the radii \( \delta_k^d \), \( \delta_k^x \) and \( \delta_k^z \). Their proofs will be provided in the appendix. Note also that we applied \( L = LU_2U_2^T \), where \( U_2^T \) from the following lemma into (14).

**Lemma 1** (Necessary Conditions for Boundedness of Set-Valued Estimates). The input and state estimation errors, \( \hat{d}_{k-1} = d_{k-1} - \hat{d}_{k-1} \) and \( \hat{x}_{k-1} = x_{k-1} - \hat{x}_{k-1} \), are bounded for
all $k$ (i.e., the set-valued estimates are bounded with radii $\delta^{d}_{k-1}, \delta^{r}_{k}, \delta^{e}_{k} < \infty$), only if $M_{1}\Sigma = I$, $M_{2}C_{2}G_{2} = I$ and $LU_{1} = 0$. Consequently, $\text{rk}(C_{2}G_{2}) = p - p_{H}$, $M_{1} = \Sigma^{-1}$, $M_{2} = (C_{2}G_{2})^{\dagger}$ and $L = LU_{2}U_{2}^{\dagger} = LU_{2}$.

Theorem 2. Suppose Lemma 1 holds, and let $T_{x,w,v}$ denote the transfer function matrix that maps the noise signals $\tilde{w}_{k} \triangleq \begin{bmatrix} w_{k}^{T} & v_{k}^{T} \end{bmatrix}$ to the updated state estimation error $\tilde{x}_{k|k} \triangleq x_{k} - \tilde{x}_{k|k}$. Moreover, assume that the following hold:

(A.1) $(A_{x}, C_{x})$ is detectable,

(A.2) $D_{x}C_{x} > 0$ and

(A.3) $\text{rk} \begin{bmatrix} A_{x} - c_{j}^{2}I & B_{x} \end{bmatrix}C_{x} = n + l - p_{H}, \forall \omega \in [0, 2\pi]$, with $\hat{A} \triangleq A - G_{1}M_{1}C_{1}$, $\Phi \triangleq I - G_{2}M_{2}C_{2}$, $\bar{A} \triangleq \Phi A$, $A_{x} \triangleq \bar{A}$, $B_{x} \triangleq [\Phi W - \Phi G_{1}M_{1} \sqrt{2} G_{2}M_{2} | 0]$, $C_{x} \triangleq C_{2}A$ and $D_{x} \triangleq \begin{bmatrix} C_{2}\Phi W & -C_{2}\Phi G_{1}M_{1} - \sqrt{2} G_{2}M_{2} \end{bmatrix}$. Then, there exists an $H_{\infty}$-observer that satisfies $\|T_{x,w,v}\|_{\infty} \leq \gamma$, i.e., the maximum average signal power amplification is upper-bounded by $\gamma^{2}$:

$$\lim_{k \to \infty} \frac{\sum_{i=0}^{k-1} \sum_{i=0}^{\infty} (\tilde{w}_{k}^{T} \tilde{w}_{i})^{2}}{\sum_{i=0}^{\infty} (\tilde{w}_{k}^{T} \tilde{w}_{i})} = ||T_{x,w,v}||^{2}_{\infty} \leq \gamma^{2},$$ (15)

if and only if $P = P^{T} > 0$ satisfies the following discrete-time algebraic Riccati equation (DARE):b

$$P = - (A_{x}P + C_{j}^{T} + B_{x}B_{x}^{T} + R_{1})^{-1} (A_{P}C_{j} + C_{j}C_{x} + B_{x}B_{x}^{T} + A_{x}PA_{x})^{T},$$ (16)

with $C_{j} \triangleq \begin{bmatrix} C_{x} & C_{x} \end{bmatrix}$, $D_{j} \triangleq \begin{bmatrix} D_{x} \ 0 \end{bmatrix}$ and $R_{1} \triangleq \begin{bmatrix} D_{x}D_{x}^{T} & 0 \\ 0 & -I \end{bmatrix}$, such that $U_{x}^{T} \triangleq I - \gamma^{-2}P > 0$ and $\hat{A} \triangleq A_{x} - (A_{x}P_{C_{j}} + B_{x}B_{x}^{T} + R_{1})C_{j}$ is asymptotically stable, i.e., $|\lambda_{i}(\hat{A})| < 1$ for all eigenvalues $\lambda_{i}$ of $A_{x}$. When such a $P$ matrix exists, the filter gain $L$ is given by

$$L = (B_{x}D_{x}^{T} + A_{x}V_{x}C_{j}C_{x}^{T} + D_{x}D_{x}^{T})^{-1}$$ (17)

with $V_{x} = P + \gamma^{-2}P_{U}^{T}P$. Moreover, $A_{x} \triangleq \bar{A}(I - LC_{2}) \bar{A}$ and $A_{x}^{*} \triangleq \bar{A}(I - LC_{2})$ are asymptotically stable.

Thus, we can search over $\gamma$ (e.g., via bisection) to find the smallest $\gamma$ and the corresponding optimal observer gain $\bar{L}$ in the minimum $H_{\infty}$-norm sense. Further, by upper-bounding the estimation errors, we find the radii of the sets of compatible inputs and states to be (cf. proof in the appendix):

$$\begin{align*}
\delta^{d}_{k-1} &= \delta^{e}_{k} ||V_{x}A_{x}^{e-1}|| + \eta_{w}(||V_{x}M_{2}C_{2}|| \\
&+ \sum_{i=0}^{k-2} ||V_{x}A_{x}^{i}(B_{x,v} + A_{x}B_{x,v})|| + \sum_{i=0}^{k-3} ||V_{x}A_{x}^{i}(B_{x,v} + A_{x}B_{x,v})|| \\
&+ ||V_{x}M_{2}T_{2}|| + ||V_{x}B_{x,v} + (V_{x} - V_{x}M_{2}C_{2}G_{2})M_{1}T_{1}|| \\
&+ ||V_{x}A_{x}^{k-2}B_{x,v}||),
\end{align*}$$

$$\begin{align*}
\delta^{r}_{k-1} &= \delta^{e}_{k} ||A_{x}^{k-1}|| + \eta_{w}(||A_{x}B_{x,v}|| + ||B_{x,v}|| \\
&+ \sum_{i=0}^{k-2} ||A_{x}B_{x,v}|| + ||B_{x,v}|| \\
&+ \sum_{i=0}^{k-3} ||A_{x}B_{x,v}|| + ||B_{x,v}||),
\end{align*}$$

b The DARE equation in (16) can be solved with control system software. For example, in MATLAB’s Control System Toolbox, we can use the command $\text{DARE}(A_{x}, C_{j}, \Phi(I + G_{1}M_{1}G_{1}^{T}G_{1}^{T}) \Phi^{T}, (C_{j}C_{2} - \frac{1}{\gamma^{2}}I))^{-1}$.
Algorithm 1 Fixed-Order Input & State Set-Valued Observer

1: Initialize: \( M_1 = \Sigma^{-1} \); \( M_2 = (C_2G_2)^T \); \( \hat{A} = A - G_1M_1C_1 \);
\( \Phi = I - G_2M_2G_2 \); \( \Phi_1 = \Phi A \);
\( V_\epsilon = V_1M_1C_1 + V_2M_2C_2A \);
Compute \( \hat{L} \) via Theorem 2 and perform bisection to minimize \( \gamma \);
\( A_c = (I - LC_2)\tilde{X} \); \( B_{c,w} = (I - LC_2)\Phi_1 \);
\( B_{c,v} = -(I - LC_2)G_2M_2 + (LC_2)T_2 \);
\( \hat{x}_{0,j} = \hat{x}_{0}(0) \);
\( 0 = \min\{|x - \hat{x}_{0}(0)| \leq \delta, \forall x \in \tilde{X}_0 \} \);
\( d_1 = M_1(z_{1,0} - C_1\hat{x}_{0}(0) - D_1u_0) \);

2: for \( k = 1 \) to \( N \) do
   \( \triangleright \) Estimation of \( d_{2,k-1} \) and \( d_{k-1} \)
   \( \hat{x}_{k,k-1} = A\hat{x}_{k-1,k-1} + B_{c,k-1} + G_1\hat{d}_{1,k-1} \);
   \( \hat{d}_{2,k-1} = M_2(z_{2,k} - C_2\hat{x}_{k,k-1} - D_2u_k) \);
   \( \hat{d}_{k-1} = V_1\hat{d}_{1,k-1} + V_2\hat{d}_{2,k-1} \);
   \( \delta_j^d = \delta_j^d \left[ ||V_1A_{c}^j|| \right] + \eta_j \sum_{i=0}^{j-2} ||V_2A_{c}^i|| ||B_{c,w}|| + \sum_{i=0}^{j-1} ||V_2A_{c}^i|| ||B_{c,v}|| + ||V_1B_{c,v} + (V_2 - V_2M_2C_2G_2)T_1|| \);
   \( \Delta_k = \{ d \in \mathbb{R}^{2} : ||d - \hat{d}_{d-1}|| \leq \delta_k^d \} \);
   \( \triangleright \) Time update
   \( \hat{x}_{k,k} = \hat{x}_{k,k-1} + G_2d_{k-1} \);
   \( \triangleright \) Measurement update
   \( \hat{x}_{k,k} = \hat{x}_{k,k} + L(z_{2,k} - C_2\hat{x}_{k,k} - D_2u_k) \);
   \( \delta_k = \delta_k \left[ ||A_{c}^k|| + \eta_k \sum_{i=0}^{k-3} ||A_{c}^i|| ||B_{c,w}|| + \eta_k ||||B_{c,v}|| \right. \)
   \( + \sum_{i=0}^{k-2} ||A_{c}^i|| ||B_{c,v}|| + \sum_{i=0}^{k-1} ||A_{c}^i|| ||B_{c,v} + A_kB_{c,v}|| \} \);
   \( \hat{X}_k = \{ x \in \mathbb{R}^{n} : ||x - \hat{x}_{k,k}|| \leq \delta_k \} \);
   \( \triangleright \) Estimation of \( d_{1,k} \)
   \( d_{1,k} = M_1(z_{1,k} - C_1\hat{x}_{k,k} - D_1u_k) \);
3: end for

Assumption (A.3) in Theorem 2. Then, an \( \mathcal{H}_\infty \)-observer with cost \( \gamma \) exist if \( U_\infty > 0 \) and \( \hat{A} \) is asymptotically stable (with \( U_\infty \) and \( \hat{A} \) defined in Theorem 2).

V. APPLICATION TO ATTACK-RESILIENT ESTIMATION

In this section, we show that the proposed set-valued observer can be applied for attack-resilient estimation of state and attack signals when a system is subject to false data injection attacks on both the actuator and sensor signals. Note that we are not considering sparse false data injection attacks, which is a subject of ongoing research. But rather, we assume that the attack locations are known (i.e., the \( G \) and \( H \) matrices that encode the attack locations are given) while the attack magnitudes \( d_k \) at any time \( k \) are unknown.

As previously discussed, the false data injection attack magnitudes \( d_k \) on the actuator and sensor signals are adversarial and strategic. Hence, we ought not make any assumptions on the attack model (random or deterministic) because a strategic adversary could otherwise simply select a different attack model than is assumed. This ‘non-assumption’ aligns perfectly with the unknown input signal model in Section II.

Therefore, for any linear time-invariant bounded-error models of a cyber-physical system, the system description in (1) can capture false data injection attacks on the actuator and sensor signals on the system without any limitations. Moreover, the fixed-order simultaneous input and state set-valued observer proposed in Section IV is capable not only of reliably estimating the set of all compatible states, \( \hat{X}_k \), but also of identifying the attack signals via the estimation of the set of all compatible unknown inputs, \( D_{k-1} \).

A. Implications on Attack-Resilient Estimation

Having established that the proposed set-valued observer is applicable to attack detection and estimation, we now discuss the implication of the relationship between observer stability and strong detectability on resilient state estimation and attack identification. First, we introduce the following definitions:

Definition 3 (Resilient Set of Compatible States). We say that the estimated set of compatible states is resilient, if for any initial state \( x_0 \in \mathbb{R}^n \) and signal attack sequence \( \{d_j\}_{j \in \mathbb{N}} \) in \( \mathbb{R}^p \), the true state \( x_k \) is contained in the set estimates \( \hat{X}_k \) and \( \hat{X}_k \), and these sets remain bounded for all \( k \).

Definition 4 (Data Injection Attack Identification). A false data injection attack is identified, if for any initial state \( x_0 \in \mathbb{R}^n \) and signal attack sequence \( \{d_j\}_{j \in \mathbb{N}} \) in \( \mathbb{R}^p \), we have a set-valued observer to compute the resilient set of compatible states and to identify the false data injection attack signals.

Based on these definitions as well as the observer design in Section IV, we have the following conclusions:

Proposition 1 (Resilient Set-Valued State Estimation). A resilient set of compatible states can be obtained if and only if the system (1) is strongly detectable and Lemma 1 holds.

Proposition 2 (Attack Identification). A false data injection attack is identified if and only if the system (1) is strongly detectable and Lemma 1 holds.

Note that strong detectability is a system property that is independent of the observer design. Hence, the necessity of strong detectability can serve as a guide to determine and recommend which actuators or sensors need to be safeguarded to guarantee resilient estimation as a preventative attack mitigation method (cf. Section VI-B for an example).

Moreover, we can derive an upper bound on the maximum number of false data injection attacks that can be asymptotically corrected based on strong detectability.

Theorem 5. The maximum number of correctable actuators and sensors signals attacks, \( p^* \), for system (1) is equal to the number of sensors, \( l \), i.e., \( p^* \leq l \) (upper bound is achievable).\(^a\)

Proof. The theorem follows immediately as an implication of Theorem 3 and [27, Theorem 1], [28, Theorem 4.3].

\(^a\)This definition is distinct from [25, Definition 1] that is defined for exact finite-time point estimation (after \( \tau \) steps) and requires strong observability [25]. Instead, it is related to boundedness in infinite time, similar to infinite-time point estimation that only requires strong detectability [27], [28].

\(^b\)By contrast, the stronger requirement of strong observability in [25] (implies strong detectability [14]) leads to a maximum of \( p^* \leq \left\lceil \frac{l}{2} \right\rceil - 1 \).
VI. SIMULATION EXAMPLES

A. Benchmark System

In this example, we consider a system that has been used as a benchmark for many state and input filters (e.g., [14]):

\[
A = \begin{bmatrix}
0.5 & 2 & 0 & 0 & 0 \\
0 & 0.2 & 1 & 0 & 1 \\
0 & 0 & 0.3 & 0 & 1 \\
0 & 0 & 0 & 0.7 & 1 \\
0 & 0 & 0 & 0 & 0.1
\end{bmatrix};
\]

\[B = 0_{5 \times 1}; C = I_5; D = 0_{5 \times 1}.\]

The unknown inputs used in this example are as given in Fig. 1, while the initial state estimate and noise signals (drawn uniformly) have bounds \(\delta_x = 0.5, \eta_w = 0.02\) and \(\eta_v = 10^{-4}\). The invariant zeros of the system matrix \(R_S(z)\) are \(\{0.3, 0.8\}\). Thus, this system is strongly detectable.

We observe from Fig. 1 that the proposed algorithm is able to find set-valued estimates of the states and unknown inputs. The actual estimation errors are also within the predicted upper bounds (cf. Fig. 2), which converges to a steady-state that is subject to data injection attacks. The system consists of 5 synchronous generators and 14 buses, with secure phasor measurement units (PMUs) being installed on the buses depicted in Fig. 3. It is represented by \(n = 10\) states comprising the rotor angles and frequencies of each generator. The dynamics of the system can be represented by an uncertain LTI model [23], [25] that is discretized with a sampling interval of \(dT = 0.05s\) to obtain the model in (1), where \(p = 35\) sensors are deployed to measure the real power injections at every bus, the real power flows along every branch and the rotor angle of generator 1. The initial state estimate and noise signals (drawn uniformly) have bounds \(\delta_x = 1, \eta_w = 0.03\) and \(\eta_v = 0.03\).

In this example, we assume that all unsecured PMU measurements are attacked (sensor attacks with known locations). Nevertheless, we observe from Fig. 4 that the proposed algorithm is able to find set-valued state estimates that contain the true state, as well as to identify a bounded set that contains the actual attack signals. Individual state and attack signals as in Fig. 1 can also be estimated but are omitted for brevity. Moreover, performing strong detectability tests (necessary condition by Propositions 1 and 2) for various attack locations, i.e., \(G\) and \(H\), we found that false data injection attacks on the sensors are correctable if at least one sensor from among sensors 10, 14 and 15 is protected.

VII. CONCLUSION

We presented an optimal fixed-order set-valued observer that simultaneously computes the set of compatible states and unknown inputs for linear discrete-time bounded-error systems in the minimum \(H_\infty\)-norm sense. Necessary and sufficient conditions for the stability of the observer and its connection to strong detectability were also derived. Then, we showed that the proposed set-valued observer is applicable for attack-resilient estimation of state and attack signals when cyber-physical systems are subject to malicious false data injection attacks on both actuator and sensor signals, as well as discussed the implication of strong detectability on resilient state estimation and attack identification.

REFERENCES


A. Proof of Lemma 1

We observe from (5), (9) and (10) that
\[ \hat{d}_{1,k} = M_1(C_1 x_{k|k} + \Sigma d_{1,k} + v_{1,k}), \]
\[ \hat{d}_{2,k-1} = M_2(C_2(A_{k-1} x_{k-1|k-1} + G_1 \hat{d}_{1,k-1} + w_{k-1}) + v_{2,k} + C_2 G_2 d_{k-1}). \]

(21)

(22)

Then, from (12) and (13), as well as (4) and (14), the propagated and updated state estimate errors are found as
\[ \tilde{x}_{k|k} = A \tilde{x}_{k-1|k-1} + G_1 \hat{d}_{1,k-1} + G_2 \hat{d}_{2,k-1} + w_{k-1}, \]
\[ \tilde{x}_{k|k} = (I - LC) \tilde{x}_{k|k} - LU_1 \Sigma d_{1,k} - Lu_k. \]

(23)

(24)

We show by induction that the estimates \( \hat{d}_{1,k}, \tilde{x}_{k|k} \) and \( \tilde{x}_{k|k} \) are bounded, assuming at the moment that their dynamics are stable (which we will show to hold in Theorem 3). For the base case, since \( \hat{x}_{0|0}, \tilde{x}_{0|0} \) and the noise signals are bounded by assumption, from (21) and (22), we find that \( \hat{d}_{1,0} \) and \( \hat{d}_{2,0} \) are bounded, only if \( M_1 \Sigma = I \) and \( M_2 C_2 G_2 = I \), since we assumed that \( d_{1,0} \) and \( d_{2,0} \) can be unbounded. Hence, \( \hat{x}_{0|0} \) is bounded. In the inductive step, we assume that \( \tilde{x}_{k-1|k-1} \) and \( \tilde{x}_{k-1|k-1} \) are bounded. Then, the input estimates \( d_{1,k-1} \) and \( d_{2,k-1} \) are bounded, only if \( M_1 \Sigma = I \) and \( M_2 C_2 G_2 = I \), since the unknown inputs \( d_{1,k} \) and \( d_{2,k-1} \) can be unbounded although the noise signals are bounded. Similarly, from (24) with a bounded measurement noise, we need the constraint \( LU_1 = 0 \) such that \( \tilde{x}_{k|k} \) remains bounded. Thus, by induction, \( \tilde{x}_{k|k} \) and \( \tilde{x}_{k|k} \) are bounded for all \( k \). Since we require \( M_2 C_2 G_2 = I \) for the existence of bounded input estimates, it follows that \( rk(C_2 G_2) = p - pH \) is a necessary condition. Furthermore, \( L = LU_1^T = LU_{22}^T = LU_2^T \) since \( LU_1 = 0 \).

B. Proof of Theorem 2

First, we reduce the system with unknown inputs to one without unknown inputs. From (14) and (5), we have \( \tilde{x}_{k|k} = \tilde{x}_{k|k} - L(C_2 x_{k|k} + v_{2,k}) \). Then, substituting (21) with \( M_1 = \Sigma^{-1} \) into (23) and the above, and rearranging, we obtain
\[ \tilde{x}_{k|k} = \tilde{x}_{k-1|k-1} + w_{k-1} - L_k(C_2 \tilde{x}_{k-1|k-1} + C_2 w_{k-1}) \]
\[ \tilde{x}_{k|k} = -(I - G_2 M_2 C_2)(G_1 M_1 v_{1,k-1} - w_{k-1}) -
\]
\[ \tilde{x}_{k|k} = -(I - G_2 M_2 C_2)(G_1 M_1 v_{1,k-1} - w_{k-1}) -
\]

with the state matrix \( \tilde{A} \) and noise signal \( w_{k-1} \).
As it turns out, the updated state estimate error dynamics above is the same for an a posteriori $H_{\infty}$ filter [15, Eq. (5.2)] for a linear system without unknown inputs

$$x_{k+1} = \bar{A}x_k + w_k, \quad y_k = Cx_k + v_k.$$ 

Since the objective for both systems is the same, i.e., to obtain the observer gain $\hat{L}$ with an a posteriori $H_{\infty}$ filter, they are equivalent systems from the perspective of estimation. Furthermore, from the analysis in [15, Eq. (5.3)], it can be seen that an equivalent observer gain $\hat{L}$ can be obtained with the a priori $H_{\infty}$ filter for the following system

$$\dot{x}_{k+1} = A\dot{x}_k + B_\infty w_k, \quad y_k = C\dot{x}_k + v_k,$$

with $A_\infty \triangleq \bar{A}$, $B_\infty \triangleq [\Phi W - \Phi G_1 M_1 - \sqrt{2}G_2 M_2 \ 0]$, $C_\infty \triangleq C_2 \bar{A}$, $D_\infty \triangleq [0 \ 0 \ 0 \ \sqrt{2I}] + C_2 B_\infty$ and $w_k \triangleq \left[ w_k^T \ v_{k-1}^T \ \frac{1}{\sqrt{2}} v_{k-2,1}^T \ \frac{1}{\sqrt{2}} v_{k-2}^T \right]^T$, where the factors $\sqrt{2}$ are included such that

$$\lim_{k \to \infty} \sum_{i=0}^{k} \|w_i\|^2 = \lim_{k \to \infty} \sum_{i=0}^{k} \|w_i\|^2 = 0.$$

The design of the observer gain $\hat{L}$ is then a direct application of the a priori $H_{\infty}$ filter (e.g., [34, Theorem 2.2]). Next, we consider the following identity $\forall \omega \in [0, \pi]$, (equivalently, $\forall \omega \in [C, 1], |z| = 1$)

$$\text{rk} \left[ A_\infty - e^{\omega i} I \begin{bmatrix} B_\infty \\ C_\infty \end{bmatrix} D_\infty \right] = \text{rk} \left[ A_\infty - z I \begin{bmatrix} B_\infty \\ C_\infty \end{bmatrix} \right].$$

By Theorem 1, strong detectability implies that Assumption (A.3) drops rank (cf. Theorem 1). Then, there exists $[\nu^T \mu^T]^T \neq 0$ such that

$$(zI - \bar{A})\nu - G_2 \mu = 0, \quad C_2 \nu = 0.$$ 

Premultiplying the former with $(I - G_2 M_2 C_\bar{A})$ and applying the latter as well as the fact that $M_2 G_2 C_\bar{A} = I$, we have

$$(I - G_2 M_2 C_\bar{A}) (zI - \bar{A})\nu + (I - G_2 M_2 C_\bar{A}) G_2 \mu = 0$$

$$(zI - \bar{A})\nu \neq 0, \quad \frac{(zI - \bar{A})\nu}{(zI - \bar{A})\nu}.$$ 

If $\nu = 0$, then $G_2 \mu = 0$ and, in turn, $\mu = 0$, which is a contradiction. Hence, $\nu \neq 0$ and the determinant of $zI - A_\infty$ is zero, i.e., any invariant zero of $\bar{R}(z)$ is a unique eigenvalue of the propagated state estimation error dynamics of $\ddot{x}_{k|k}$, which can be found from (27) to be

$$\ddot{x}_{k|k}^* = \bar{A}(I - \bar{L}C_2)^{-1} x_{k-1|k-1} + B_{e,v} w_{k-1} + B_{e,w} e_{k|k} - \bar{L} \bar{T}.$$ 

and by extension, the state matrix $A_\infty$ in (25), since $A_\infty$ and $A_\infty^*$ have the same eigenvalues [36, Theorem 1.3.22].

E. Proof of Theorem 3

Let Lemma 1 hold. We will prove that strong detectability implies that state estimation errors with $A^*_\infty = \bar{A}(I - \bar{L}C_2)$ and $A_{e} = (I - \bar{L}C_2)\bar{A}$ (they have same eigenvalues [36]) are asymptotically stable and bounded, and vice versa $(\Rightarrow)$: By Theorem 1, strong detectability implies that $A_{e}$ is detectable. Hence, there exists $\bar{L}$ such that $A_{e}$, and by extension $A_e^*$, are asymptotically stable, which in turn implies bounded-input, bounded-state stability of (25), (28).

$(\Leftarrow)$: We will show this by contraposition. By Lemma 2, we know that (1) being not strongly detectable implies that $A_e^*$, and by extension $A_e$, are unstable. This implies that

$$\lim_{k \to \infty} \|A^*_e k\| = \lim_{k \to \infty} \|A^*_e k\| = \infty. $$

thus, from (19) and (20), the estimation errors are unbounded for any $\delta_0 \neq 0$.

F. Proof of Theorem 4

By Theorem 1, strong detectability implies that Assumption (A.1) holds. Moreover, $D_\infty D_{\infty}^T = C_2 G_2 M_2^T G_2^T + \Phi (W W^T + G_1 M_1 M_1^T G_1^T) \Phi^T + 2I \geq 0$ readily satisfies Assumption (A.2). Finally, for Assumption (A.3) to hold, we require that

$$\text{rk} \left[ B_{e,v} - \bar{A} G_1 M_1 - \sqrt{2}G_2 M_2 \right]$$

is stable on the unit circle. Moreover, if $p = l$, from Theorem 1, strong detectability satisfies the rank condition above and thus Assumption (A.3).