An Observer-based Switching Algorithm for Safety under Sensor Denial-of-Service Attacks

Santiago Jimenez Leudo, Kunal Garg, Ricardo G. Sanfelice and Alvaro A. Cardenas

***Abstract*— The design of safe-critical control algorithms for systems under Denial-of-Service (DoS) attacks on the system output is studied in this work. We aim to address scenarios where attack-mitigation approaches are not feasible, and the system needs to maintain safety under adversarial attacks. We propose an attack-recovery strategy by designing a switching observer and characterizing bounds in the error of a state estimation scheme by specifying tolerable limits on the time length of attacks. Then, we propose a switching control algorithm that renders forward invariant a set for the observer. Thus, by satisfying the error bounds of the state estimation, we guarantee that the safe set is rendered conditionally invariant with respect to a set of initial conditions. A numerical example illustrates the efficacy of the approach.**

I. Introduction

The security of Cyber-Physical Systems from a control-theoretic perspective is a growing area of research [1]. Various types of attacks on a control system can occur, such as sensor data or system actuators getting compromised [2], [3]. Attackers can disable the transmission of signals between devices, causing a Denial of Service (DoS) attack [4]. Such attacks can lead to violation of safety requirements, such as avoiding obstacles or keeping the system trajectories in a desired region of the state space [5].

Mitigating and responding to attacks is an active area of research. Some efforts have focused on developing robust observers to prevent compromised data from affecting the feedback loops. Secure estimation uses redundant observers to reconstruct the state, but they assume that only a certain number of sensors (in particular, less than half of the sensors) have been compromised [6]. An alternative to reduce this level of redundancy is to reject outliers with the use of robust statistics [7]. This approach requires precise knowledge of the system's dynamical model. The control signal can also be constrained to prevent attackers from causing damages [8]. However, such approaches may negatively affect the system's performance. There is a plethora of work on resilient or robust control design, see, e.g., [6], [9], [10], that focuses on system performance under attacks, however, without any consideration or guarantees on safety.

Safety is perhaps the most important system property we need to maintain while undergoing an attack. Control barrier function (CBF)-based approaches can help design control algorithms for forward invariance of a safe set [11]. The authors in [12] introduce the notion of fault-tolerant CBF for handling attacks on stochastic systems. In [13], the authors study safe control design under DoS attacks.

In this paper, we focus on the problem of safely recovering from output attacks, i.e., keeping the system trajectories in a safe set even under DoS attacks. The proposed formulation is applicable to several use cases with objectives including obstacle avoidance and collision-free navigation for autonomous vehicles, reach-avoid control problems, surveillance, and convoy of multi-agent systems, among others. We propose a control scheme based on the information available, namely, the uncompromised outputs, that assures safety for systems with outputs experiencing DoS attacks. We consider scenarios in which every attack has finite duration, succeeded by an interval of time without attacks. We are interested in finding the set of initial conditions and the control action such that the state trajectory remains in the safe set at all times. During attacks, the controller relies only on the uncompromised outputs, from which we generate an estimate of the state, whereas the entire output is used when attacks are not present.

In this paper, we design a switching observer scheme that uses the complete output information when there is no attack and uncompromised output information during an attack in the sensors. We provide sufficient conditions involving key properties of the system, such as the maximum tolerable length of the DoS attack and the minimum required length of the interval without an attack for recovery, guaranteeing that the state estimation error remains uniformly bounded. Furthermore, we design CBF-based observer-based feedback laws to render a properly defined set forward invariant for the observer so that with bounded estimation error, the system is safe. This is obtained provided conditional invariance of a set of interest with respect to a set of initial states. Due to space constraints, proofs and other details are not included and will be published elsewhere.

Notation. The symbols $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{N}_{>0}$ denote the sets of real numbers, nonnegative reals, and positive natural numbers, respectively. Let $|x|$ be the Euclidean norm of the vector x. Let $\overline{\mathcal{A}}$ denote the closure of the set $\mathcal{A}$. Let $|A|$ be

S. J. Leudo and R. G. Sanfelice are with the Department of Electrical and Computer Engineering, and A. A. Cardenas is with the Department of Computer Science and Engineering, University of California, Santa Cruz, CA 95064. K. Garg is with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139. Email: {sjimen28, ricardo, alacarde}@ucsc.edu, {kgarg}@mit.edu

Research partially supported by the NSF Grants no. ECS-1710621, CNS-2039054, and CNS-2111688, by the AFOSR Grants no. FA9550-19-1-0169, FA9550-20-1-0238, and FA9550-23-1-0145, by the AFRL Grant nos. FA8651-22-1-0017 and FA8651-23-1-0004, by ARO Grant no. W911NF-20-1-0253, and by Fulbright Colombia - MinTIC. The views and conclusions of this document are those of the authors and should not be interpreted as representing the official policies of the ARO or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes, notwithstanding any copyright notation herein.

the induced matrix 2−norm of A, rank(A) denote its rank, and $\lambda_m(A), \lambda_M(A)$ denote the eigenvalues with minimum and maximum real part, respectively. Let $\mathbb{B} \subset \mathbb{R}^n$ denote the closed unit ball centered at the origin and $p+r\mathbb{B}$ the ball of radius $r \geq 0$ centered at $p \in \mathbb{R}^n$. We denote by $\tilde{\mathcal{O}}(C, A)$ the observability matrix of the pair (C, A) and by $\tilde{\mathcal{C}}(A, B)$ the controllability matrix of the pair (A, B).

II. Preliminaries

Consider the nonlinear system

$$\mathcal{F}: \qquad \dot{z} = F(t, z), \qquad y = H(t, z) \tag{1}$$

where $z \in \mathbb{R}^n$ is the system state, $y \in \mathbb{R}^p$ is the system output, $F : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^n$ is the (potentially nonsmooth) flow map and $H : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^p$ is the output map.

A solution to the system $\mathcal{F}$ is defined as follows.

Definition 1 (Solution to $\mathcal{F}$)**.** *A locally absolutely continuous function $t \mapsto z(t)$ defines a solution to the system $\mathcal{F}$ in (1) if $\frac{d}{dt} z(t) = F(t, z(t))$ for almost all $t \in \mathbb{R}_{\geq 0}$.*

We say that a solution z to $\mathcal{F}$ is maximal if it cannot be extended and we say it is complete when dom $z = [0, \infty)$.

Definition 2 (Safety)**.** *The system (1) is said to be safe with respect to (X_0, X_u), with $X_0 \subset \mathbb{R}^n \setminus X_u$, if for each $z_0 \in X_0$, each solution $t \mapsto z(t)$ to (1) with $z(0) = z_0$ satisfies $z(t) \in \mathbb{R}^n \setminus X_u$ for all $t \in$ dom z.*

Definition 3 (Conditional invariance)**.** *A closed set $S \subset \mathbb{R}^n$ is said to be conditionally invariant for system (1) with respect to $M \subset S$ if, for each $z_0 \in M$, any solution $t \mapsto z(t)$ to (1) from z_0 satisfies $z(t) \in S$ for all $t \in$ dom z.*

It is immediate that the system (1) is safe with respect to (X_0, X_u) if and only if the set $S := \mathbb{R}^n \setminus X_u$ is conditionally invariant for (1) with respect to X_0. For more details see [14].

III. Problem formulation

A. System Model

Consider the linear time-invariant control system

$$\mathcal{S}: \qquad \dot{x} = Ax + Bu, \qquad y = Cx \tag{2}$$

where $x \in \mathbb{R}^n$ is the system state, $y \in \mathbb{R}^p$ is the system output, $u \in \mathcal{U}$ is the control input, and $\mathcal{U} \subset \mathbb{R}^m$. Here, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, and $C \in \mathbb{R}^{p \times n}$.

B. Attack Model

In this work, we consider attacks on the system output y. In particular, we consider an attack where a subset of the components of the system output is compromised. Under such an attack model, the *measured* system output $\bar{y}$ takes the form

$$\bar{y} = (y_s, y_a) \tag{3}$$

where $y_s = \tilde{C}x$, and, for each solution $t \mapsto x(t)$ to (2),

$$y_a(t) = \begin{cases} \bar{C}x(t) & \text{if } t \notin \mathcal{T}_a, \\ Y(t, x(t)) & \text{if } t \in \mathcal{T}_a \end{cases} \tag{4}$$

The quantity $\tilde{C}x$ denotes the *secured* output components that *cannot* be attacked with $\tilde{C} \in \mathbb{R}^{\tilde{p} \times n}$ and $0 \leq \tilde{p} < p$, $\bar{C}x$ denotes the *vulnerable* output components that *can* be attacked with $\bar{C} \in \mathbb{R}^{(p-\tilde{p}) \times n}$ such that $C = \begin{bmatrix} \tilde{C} \\ \bar{C} \end{bmatrix}$, and $Y : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^{p-\tilde{p}}$ denotes the attacked output signal. We denote with $\mathcal{T}_a \subset \mathbb{R}_{\geq 0}$ the set of times when an attack is present on the system output, which is assumed to be known provided a DoS attack detection mechanism. The attack model (3) captures Denial-of-Service (DoS) attacks on the system output. Let $[t_1^i, t_2^i)$ with $t_2^i > t_1^i \geq 0$ denote the interval of time over which the $i-$th DoS attack occurs, with $i \in \mathbb{N}_{>0}$. Define $\mathcal{T}_a := \bigcup_i [t_1^i, t_2^i)$, $\mathcal{T}_1 = \bigcup_i \{t_1^i\}$, and $\mathcal{T}_2 = \bigcup_i \{t_2^i\}$ as the intervals of attack, and the sets of the starting and ending time instants of attacks, respectively. To provide sufficient conditions to guarantee safety, we characterize the attacks by defining $T_a := \max_{i \in \{1,2,\dots\}} (t_2^i - t_1^i)$ and $T_{na} := \min_{i \in \{2,3,\dots\}} (t_1^i - t_2^{i-1})$ as the maximum length of the DoS attack and the minimum length of the interval without an attack, respectively. Notice that $t_2^0 := 0$, and when $t_1^1 > 0$, we have $t_1^1 \geq T_{na}$.

C. Problem Statement

Given a nonempty, closed set $S \subset \mathbb{R}^n$, referred to as the *safe* set, the problem to solve is the design of an algorithm such that the set S is conditionally invariant for (2) with respect to the set X_0. Formally, the control design problem studied in this paper is stated as follows.

Problem $(\star)$**.** *Given system (2), a closed set $S \subset \mathbb{R}^n$, and the attack model in (3),*

1) *Find a set of initial states $X_0 \subset S$, and*
2) *Design a control law κ assigning the input u of (2) using measurements of $\bar{y}$*

such that, for each $x_0 \in X_0$, the solution to the resulting closed-loop system, namely $t \mapsto x(t)$, with $x(0) = x_0$, satisfies $x(t) \in S$ for all $t \geq 0$.

D. Proposed Solution

To solve Problem $(\star)$, we propose the design of an observer-based feedback law that induces conditional invariance of S with respect to X_0. Most CBF-based methods for forward invariance rely on measurement of the entire state [15]. We propose to employ a state estimator that reconstructs the system state using the measured output $\bar{y}$. The observer is given as

$$\dot{\hat{x}} = A\hat{x} + Bu + g(\bar{y}, \hat{y}), \qquad \hat{y} = C\hat{x}, \tag{5}$$

where $\hat{x} \in \mathbb{R}^n$ is the estimate of x and $g : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^n$ is the innovation term to be designed such that $g(\bar{y}, \hat{y}) = 0$ at $\bar{y} = \hat{y}$. When the system output is under an attack according to the attack model (3), the actual output information is not available to the state observer. Thus, the observer needs to take into account the attacks on the system output. To this end, we design an observer that uses the *complete output* vector when there is no attack and only the *non-attacked*

output components when the system output is under attack. More specifically, the proposed observer under the attack model (3) is given as

$$\dot{\hat{x}} = \begin{cases} A\hat{x} + Bu + g_1(Cx, C\hat{x}) & \text{if} \quad t \notin \mathcal{T}_a, \\ A\hat{x} + Bu + g_2(\tilde{C}x, C\hat{x}) & \text{if} \quad t \in \mathcal{T}_a \end{cases} \tag{6}$$

where $g_1, g_2 : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^n$ are to be designed. Given a set $\mathcal{T}_a \subset \mathbb{R}$, the feedback law κ assigning u is defined as

$$\kappa(t, \hat{x}, y) = \begin{cases} \kappa_1(\hat{x}, y) & \text{if} \quad t \notin \mathcal{T}_a, \\ \kappa_2(\hat{x}, y) & \text{if} \quad t \in \mathcal{T}_a, \end{cases} \tag{7}$$

where $\kappa_1, \kappa_2 : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m$ are functions to be designed under *nominal* operation (i.e., when the system is not under an attack) and under attack, respectively. Notice that the closed-loop system resulting from the composition of (2) and (6) with κ as in (7) can be expressed as in (1) with $z = (x, \hat{x})$.

We make the following assumption on $\mathcal{S}$ in (2).

Assumption 1. *The pair (A, B) is controllable and the pair (C, A) is detectable.*

Based on the structure of the observer in (2) and the observer-based feedback law in (7), the approach followed in this paper for safety under attacks for system (2) is as follows.

Approach: Given a closed set $S \subset \mathbb{R}^n$, the system (2), and the attack model (3), our approach is to compute sets $X_0, \hat{X}_0, \hat{S}_0 \subset S$ and design functions g_1, g_2 for the observer in (6) and functions κ_1, κ_2 for the observer-based feedback law κ as in (7) such that each solution pair $t \mapsto (x(t), \hat{x}(t))$ to the closed-loop system resulting from the composition of (2) and (6) with κ satisfies the following properties:

1) For each $t_0 \in \mathcal{T}_1$ such that $x(t_0) \in X_0$ and $\hat{x}(t_0) \in \hat{X}_0$, the x component of the resulting closed-loop solution satisfies $x(t) \in S$ for all $t \in [t_0, t_0 + T_a)$;
2) For each $t_0 \in \mathcal{T}_2$ such that $x(t_0) \in S$ and $\hat{x}(t_0) \in \hat{S}_0$, and for $\hat{t}_0 = \max\{t_0, \inf_{t \geq t_0} \mathcal{T}_1\}$, the x component of the resulting closed-loop solution satisfies $x(\hat{t}_0) \in X_0$ and $x(t) \in S$ for all $t \in [t_0, \hat{t}_0)$.

Remark 1. *The sets $\hat{X}_0$ and $\hat{S}_0$ denote the sets of estimates before and after an attack, respectively. We will design these sets in the next section. Item 1 in our solution approach encodes conditional invariance of the set S for system (2) with respect to X_0, under an attack with maximum duration. Upon the requirement of the state to be in S at the end of every attack, item 2 encodes safety of system (2) with respect to $(X_0, \mathbb{R}^n \backslash S)$ during the time-intervals with no attacks, and the state to be in X_0 at the beginning of the next attack.*

IV. Switching Observer Design

Under an attack on the system output of the form (3), it might not be possible to reconstruct the state of (2) for a full-state feedback control design. Specifically, under the considered attack model, the rank of the observability matrix $\tilde{\mathcal{O}}$ for the pair $(\tilde{C}, A)$, namely, $\text{rank}(\tilde{\mathcal{O}}) = \tilde{n}$, potentially smaller than n. Thus, there might be $n - \tilde{n} > 0$ eigenvalues in the closed right-half plane for the dynamics of the estimation error resulting for any observer design under attack. Keeping this in mind, the switching observer in (6) is defined as

$$\dot{\hat{x}} = \begin{cases} A\hat{x} + Bu + L(Cx - C\hat{x}) & \text{if} \quad t \notin \mathcal{T}_a, \\ A\hat{x} + Bu + \tilde{L}(\tilde{C}x - \tilde{C}\hat{x}) & \text{if} \quad t \in \mathcal{T}_a, \end{cases} \tag{8}$$

where $L \in \mathbb{R}^{n \times p}$ and $\tilde{L} \in \mathbb{R}^{n \times \tilde{p}}$ is such that $\tilde{n}$ (with $\tilde{n} \leq n$) eigenvalues of the matrix $A - \tilde{L}\tilde{C}$ lie in the open left-half plane. On the other hand, since (C, A) is detectable under Assumption 1, we can design L such that all the eigenvalues of $(A - LC)$ are in the open left-half plane. Now, define $e = x - \hat{x}$ as the estimation error to obtain the error dynamics given as

$$\dot{e} = \begin{cases} (A - LC)e & \text{if} \quad t \notin \mathcal{T}_a, \\ (A - \tilde{L}\tilde{C})e & \text{if} \quad t \in \mathcal{T}_a \end{cases} \tag{9}$$

with $e(0) = x(0) - \hat{x}(0)$. Next, we analyze the error bounds when there is no attack, i.e., at each $t \notin \mathcal{T}_a$.

A. Analysis under No Attacks

Consider the starting instant of an interval during which there is no attack on the system output, namely $t_2^i \in \mathcal{T}_2 \cup \{0\}$, with $i \in \mathbb{N}$. The following result is the initial step to guarantee conditional invariance of S with respect to X_0 for the system (2) when there are no attacks.

Lemma 1. *Given system (2), suppose Assumption 1 holds. For given $T_{na}, \bar{e}_0 > 0$, an associated observer (8), and corresponding error dynamics (9), if at the $i-$th interval of no attacks with $i \in \mathbb{N}$, $|e(t_2^i)| \leq \bar{e}_0$ with $t_2^i \in \mathcal{T}_2$, then the state estimation error satisfies $|e(t)| \leq \gamma_1(t - t_2^i)\bar{e}_0$ for all $t \in [t_2^i, t_1^{i+1}]$, where*

$$\gamma_1(t) \coloneqq c_1 \exp\left(-\bar{\lambda}_1 t\right) \tag{10}$$

with $\bar{\lambda}_1 = \frac{\lambda_m(Q)}{2\lambda_M(P)}$, $c_1 = \sqrt{\frac{\lambda_M(P)}{\lambda_m(P)}}$, and L such that for some symmetric positive definite matrices P and Q, $-Q = (A - LC)^\top P + P(A - LC)$ holds.

Notice that the above analysis (with a nominal Luenberger observer) can be used to show that starting from $e(t_2^i)$ with $t_2^i \in \mathcal{T}_2 \cup \{0\}, i \in \mathbb{N}$, the error exponentially converges to $\delta\mathbb{B}$ in time T_{na}, where $\delta = \gamma_1(T_{na})|e(t_2^i)|$, and stays in that ball until the next attack starts at t_1^{i+1}.

Remark 2. *The Luenberger observer used when there are no attacks is just one choice of a state estimator. It is also possible to use a finite-time stable state estimator [16], or any other observer that has faster convergence guarantees.*

B. Analysis under Attacks

During the attack on the output, we use a different observer gain designed for the pair $(\tilde{C}, A)$. Since it might not be possible to place all the eigenvalues of $A - \tilde{L}\tilde{C}$ in the open left-half plane, the matrix $\tilde{L}$ in (8) can be designed to minimize the maximum eigenvalue of $A - \tilde{L}\tilde{C}$, which minimizes the rate of growth of the error during attacks.

Based on $\tilde{L}$, we compute the maximum growth rate possible in the estimation error e during intervals of attacks in the system output, assuming a worst-case attack.

Under the attack model (3), a subset of the state space may still be detectable for the pair $(\tilde{C}, A)$. Thus, under the observer (8) for $t \in \mathcal{T}_a$, it is possible that some of the eigenvalues of the matrix $A - \tilde{L}\tilde{C}$ are in the open left-half plane. To bound the error growth during the attack, we consider the general case in which we can decompose the matrix $A - \tilde{L}\tilde{C}$ into submatrices $\hat{A}_{11}$ and $\hat{A}_{22}$, such that the eigenvalues of $\hat{A}_{11}$ are in the open left-half plane. To this end, let $\Phi \in \mathbb{R}^{n\times n}$ be an invertible matrix consisting of the generalized eigenvectors of the matrix $A - \tilde{L}\tilde{C}$ such that

$$\Phi^{-1}(A - \tilde{L}\tilde{C})\Phi = \begin{bmatrix} \hat{A}_{11} & 0_{\tilde{n}\times(n-\tilde{n})} \\ 0_{(n-\tilde{n})\times\tilde{n}} & \hat{A}_{22} \end{bmatrix} \tag{11}$$

where $\hat{A}_{11}$ and $\hat{A}_{22}$ are Jordan blocks such that $\lambda_M(\hat{A}_{11}) < 0$ and $0_{p\times q} \in \mathbb{R}^{p\times q}$ is a matrix consisting of zeros[1]. Also, let $\Phi^{-1} = \left[\hat{\Phi}_1^\top, \hat{\Phi}_2^\top\right]$, and define the change of coordinates $z = \Phi^{-1}e$. Then, $e = \Phi z$, and in the new coordinates, the error dynamics are expressed as

$$\dot{z} = \Phi^{-1}\dot{e} = \begin{bmatrix} \hat{A}_{11} & 0_{\tilde{n}\times(n-\tilde{n})} \\ 0_{(n-\tilde{n})\times\tilde{n}} & \hat{A}_{22} \end{bmatrix} z.$$

Define $z = (z_{11}, z_{22})$, where $z_{11} \in \mathbb{R}^{\tilde{n}}$ and $z_{22} \in \mathbb{R}^{n-\tilde{n}}$ so that we have $\dot{z} = (\dot{z}_{11}, \dot{z}_{22}) = (\hat{A}_{11}z_{11}, \hat{A}_{22}z_{22})$. We can now state the following result providing a bound on the state estimation error under attacks.

Lemma 2. *Given system (2), suppose Assumption 1 holds. For given $T_a, \bar{e}_0 > 0$, an associated observer (8), and corresponding error dynamics (9), if at the $i-$th interval of attack with $i \in \mathbb{N}_{>0}$ and maximum length T_a, $|e(t_1^i)| \le \bar{e}_0$ with $t_1^i \in \mathcal{T}_1$, then the state estimation error satisfies $|e(t)| \le \gamma_2(T_a)\bar{e}_0$ for all $t \in [t_1^i, t_2^i]$, where*

$$\gamma_2(T_a) \coloneqq \max_{t\in[0,T_a]} \hat{c}_1 \exp\left(-\hat{\lambda}_1 t\right) + \hat{c}_2 \exp\left(\hat{\lambda}_2 t\right) \tag{12}$$

with

$$\hat{c}_1 = |\Phi||\hat{\Phi}_1|\sqrt{\frac{\lambda_M(\hat{P})}{\lambda_m(\hat{P})}},\ \hat{c}_2 = |\Phi||\hat{\Phi}_2|,\ \hat{\lambda}_1 = \frac{\lambda_m(\hat{Q})}{2\lambda_M(\hat{P})},\ \hat{\lambda}_2 = |\hat{A}_{22}|,$$

and $\tilde{L}$ such that for some symmetric positive definite matrices $\hat{P}$ and $\hat{Q}$, $-\hat{Q} = \hat{A}_{11}^\top\hat{P} + \hat{P}\hat{A}_{11}$ holds.

C. Global Bound on Estimation Error

Before we state the first main result of the paper, we make the following assumption on the initial state estimation error.

Assumption 2. *The closed set $S \subset \mathbb{R}^n$ is such that there exists $\bar{E} > 0$ such that, for the initial state $x(0) \in S$ and initial estimate $\hat{x}(0) \in S$, the error satisfies $|e(0)| = |x(0) - \hat{x}(0)| \le \bar{E}$.*

A pre-defined initial error bound helps us guarantee the existence of a switching observer of the form (6) such that safety is guaranteed.

[1]Note that it is always possible to find the Jordan form of the matrix $A-\tilde{L}\tilde{C}$, even when it is not diagonalizable.

Now, we provide a result on bounds on the state estimation error under the proposed switching observer algorithm.

Theorem 1. *Given system (2), suppose Assumptions 1 and 2 hold for $\bar{E} > 0$. For given $T_{na}, T_a > 0$, an associated observer (8), and corresponding error dynamics (9), let $c_1, \bar{\lambda}_1, \hat{c}_1, \hat{c}_2, \hat{\lambda}_1, \hat{\lambda}_2 > 0$ be defined as per Lemma 1 and Lemma 2. If T_{na} and T_a are such that $\gamma_1(T_{na})\gamma_2(T_a) \le 1$ with γ_1 as in (10) and γ_2 as in (12), then $|e(t)| \le \gamma_1(0)\gamma_2(T_a)\bar{E}$ for all $t \ge 0$. In addition,*

- *if there is an attack at time $t = 0$, then $|e(t)| \le \bar{E}$ for all $t \in \mathcal{T}_1 \cup \{0\}$, and*
- *if the first attack is launched after at least T_{na} seconds, then $|e(t)| \le \bar{E}$ for all $t \in \mathcal{T}_2 \cup \{0\}$.*

Remark 3. *Consider a set X_0, and for a given $x_0 \in X_0$ such that $x(0) = x_0$, define the set $\hat{X}_0(x_0) := \{x \in \mathbb{R}^n : x \in x_0 + \bar{E}\mathbb{B}\}$. Notice that thanks to Theorem 1, for each $x_0 \in X_0$, and each $\hat{x}_0 \in \hat{X}_0(x_0)$ we have that each solution pair $t \mapsto (x(t), \hat{x}(t))$ to (2) from $x(0) = x_0, \hat{x}(0) = \hat{x}_0$ satisfies*

1) *Boundedness of error at all times: $|x(t) - \hat{x}(t)| \le \bar{E}$ for all $t \ge 0$;*
2) *Maximum error at the beginning of each attack: $|x(t_1^i) - \hat{x}(t_1^i)| \le \gamma_1(T_{na})$ for each $i \in \mathbb{N}_{>0}$.*

Under an attack, it is possible that the error grows, and when there is no attack, the error decreases. However, using the proposed observer, the norm of the error always remains bounded by $\gamma_1(0)\gamma_2(T_a)\bar{E}$, as long as Assumption 2 on the initial estimation error holds.

V. Observer-Based Feedback Law Design

A. Construction of Sets of Initial Conditions

Consider a closed set $S \subset \mathbb{R}^n$, $T_a, T_{na} > 0$, maps γ_1 and γ_2 as in (10) and (12), and $\bar{E} > 0$ in Assumption 2. Pick $\varepsilon > (1 + \gamma_1(0)\gamma_2(T_a))\bar{E}$ Define the set of initial states as

$$X_0 \coloneqq S \setminus (\partial S + \varepsilon\mathbb{B}). \tag{13}$$

Note that under Assumption 2, X_0 is nonempty. Now, given $x_0 \in X_0$, set $x(0) = x_0$ and define the set-valued map

$$\hat{X}_0(x_0) \coloneqq x_0 + \bar{E}\mathbb{B}. \tag{14}$$

Thus, for each $x_0 \in X_0$ and $\hat{x}_0 \in \hat{X}_0(x_0)$, it holds that $|x_0 - \hat{x}_0| \le \bar{E}$. Additionally, notice that $\hat{x}_0 \in \tilde{X}$, where

$$\tilde{X} \coloneqq X_0 + \bar{E}\mathbb{B} \tag{15}$$

which is an inflation of X_0 by $\bar{E}$. This construction of the sets of initial conditions, namely, X_0 and $\hat{X}_0$, leads to conditional invariance of S, as shown below.

Lemma 3. *Given the system (2), the observer (8), the observer-based feedback law κ (7), a closed set $S \subset \mathbb{R}^n$, X_0 as in (13), and $\hat{X}_0$ as in (14), consider the solution $t \mapsto (x(t), \hat{x}(t))$ to the resulting closed-loop system from the composition of (2) and (8) with κ from $x(0) \in X_0$, $\hat{x}(0) \in \hat{X}_0(x(0))$ and $T_a, T_{na}, \bar{E}$, such that conditions of Theorem 1 are satisfied. If $S \setminus (\partial S + (1 + \gamma_1(0)\gamma_2(T_a))\bar{E}\mathbb{B}) \neq \emptyset$ and $\hat{x}(t) \in \tilde{X}$ for all $t \ge 0$, then $x(t) \in S$ for all $t \ge 0$.*

In words, the set of initial states X_0 and the set of initial estimates $\hat{X}_0$ are defined such that the initial estimation error is upper bounded by $\bar{E}$. Furthermore, we define $\tilde{X}$ in (15) as the set resulting from an inflation of X_0 by $\bar{E}$. Under this construction, for the resulting closed-loop system from the composition of (2) and (8) with κ, forward invariance of $\tilde{X}$ for the observer (8) implies conditional invariance of the set S for the system (2) with respect to X_0. Thus, the control objective is to enforce the estimate $\hat{x}$ in the set $\bar{X}$ at all times to guarantee safety of S.

B. QP-based Feedback Law Synthesis

We use a control barrier function (CBF)-based approach for guaranteeing forward invariance of a subset $\bar{X}$ of the set $\tilde{X}$ in (15) for (8) (see [15]). In order to use CBF for forward invariance, we need a zero sublevel set representation of the set $\bar{X}$. To this end, consider the function $h : \mathbb{R}^n \to \mathbb{R}$ and define a set

$$\bar{X} := \{\hat{x} \mid h(\hat{x}) \leq 0\} \subset \tilde{X}. \tag{16}$$

Given an observer-based feedback law κ assigning the input $u = \kappa(t, \hat{x}, \bar{y})$ of (8), consider a solution $t \mapsto \hat{x}(t)$ to (8) from $\hat{x}(0) \in \bar{X}$. For the given measurement $\bar{y}$, it is sufficient to ensure that for each $\hat{x}(0) \in \bar{X}$, the estimate satisfies $\hat{x}(t) \in \bar{X} \subset \tilde{X}$, for all $t \geq 0$. The CBF condition for guaranteeing this when there is no attack is:

$$\frac{\partial}{\partial \hat{x}} h(\hat{x}(t)) \left(A\hat{x}(t) + B\kappa_1(\hat{x}(t), \bar{y}(t)) + L(\bar{y}(t) - C\hat{x}(t))\right) \leq \alpha_1(-h(\hat{x}(t))), \tag{17}$$

for all $t \geq 0$, where $t \mapsto \bar{y}(t)$ is the measured output signal, and the CBF condition under attack is

$$\frac{\partial}{\partial \hat{x}} h(\hat{x}(t)) \left(A\hat{x}(t) + B\kappa_2(\hat{x}(t), \bar{y}(t)) + \tilde{L}(\bar{y}_s(t) - \tilde{C}\hat{x}(t))\right) \leq \alpha_1(-h(\hat{x}(t))), \tag{18}$$

for all $t \geq 0$, where $t \mapsto y_s(t)$ is the secured output signal and α_1, α_2 are class-$\mathcal{K}$ functions. We can use a Quadratic Programming (QP) formulation to compute the input u in the respective cases.

Consider the following QP for each $\hat{x} \in \bar{X}$ and $\bar{y}$ such that $x \in S$ for input synthesis when there is no attack:

$$\min_{(v,\eta)} \quad \frac{1}{2}|v - K\hat{x}|^2 + \frac{1}{2}\eta^2 \tag{19a}$$

$$\text{s.t.} \quad \frac{\partial}{\partial \hat{x}} h(\hat{x}) \left(A\hat{x} + Bv + L(\bar{y} - C\hat{x})\right) \leq -\eta h(\hat{x}), \tag{19b}$$

where K is the optimal LQR gain for the pair (A, B). Next, we use a similar QP to compute the input under attack. Consider the following QP for each $\hat{x} \in \bar{X}$ and $y_s = \tilde{C}x$ such that $x \in S$:

$$\min_{(v_s,\zeta)} \quad \frac{1}{2}|v_s - K\hat{x}|^2 + \frac{1}{2}\zeta^2 \tag{20a}$$

$$\text{s.t.} \quad \frac{\partial}{\partial \hat{x}} h(\hat{x}) \left(A\hat{x} + Bv_s + \tilde{L}(y_s - \tilde{C}\hat{x})\right) \leq -\zeta h(\hat{x}). \tag{20b}$$

The objective functions in (19) and (20) set the convex minimization problem to obtain the closest control action to the LQR control that satisfies the constraints. The additional decision variables, namely (η, ζ), respectively, are slack variables. Denote the solutions to (19) and (20) as $t \mapsto u_1^*(\hat{x}(t), \bar{y}(t))$ and $t \mapsto u_2^*(\hat{x}(t), \bar{y}(t))$, respectively. To guarantee continuity of these solutions with respect to $\hat{x}$, we need to impose the strict complementary slackness condition (see [17]). In brief, if the $i-$th constraint of (19) (or (20)), with $i \in \{1, 2\}$, is written as $G_i(\hat{x}, \bar{y}, u_{QP}) \leq 0$ with $u_{QP} = (v, \eta)$ (respectively, $u_{QP} = (v_s, \zeta)$ for (20)), and the corresponding Lagrange multiplier is $\bar{\lambda}_i \in \mathbb{R}_{\geq 0}$, then strict complementary slackness requires that $\bar{\lambda}_i^* G(\hat{x}, \bar{y}, u_{QP}^*) < 0$, where u_{QP}^* and $\bar{\lambda}_i^*$ denote the optimal solution and the corresponding optimal Lagrange multiplier, respectively. We are now ready to state the second main result of the paper.

Theorem 2. *Given system (2), suppose that Assumptions 1 and 2 hold. For the attack model (3), the observer (8), and a closed set $S \subset \mathbb{R}^n$, let X_0, $\hat{X}_0$, $\tilde{X}$ and $\bar{X}$ be given as in (13)-(16), and assume that the strict complementary slackness holds for the QPs (19) and (20) for all $\hat{x} \in \bar{X}$. The following holds:*

1) *If $S \setminus (\partial S + (1+\gamma_1(0)\gamma_2(T_a))\bar{E}\mathbb{B}) \neq \emptyset$, then, for each $\hat{x} \in \text{int}(\bar{X})$, the QPs (19) and (20) are feasible and their respective solutions $t \mapsto u_1^*(\hat{x}(t), \bar{y}(t)), t \mapsto u_2^*(\hat{x}(t), \bar{y}(t))$ are continuous on $\text{int}(\bar{X})$.*
2) *For each $x_0 \in X_0$ and $\hat{x}_0 \in \bar{X} \cap \hat{X}_0(x_0)$, each solution pair $t \mapsto (x(t), \hat{x}(t))$ to the closed-loop system resulting from assigning the input u of (2) and (8) to the observer-based feedback law κ in (7) with $\kappa_1(\hat{x}, \bar{y}) = u_1^*(\hat{x}, \bar{y})$ and $\kappa_2(\hat{x}, \bar{y}) = u_2^*(\hat{x}, \bar{y})$, satisfies $\hat{x}(t) \in \bar{X}$ and $x(t) \in S$ for all $t \geq 0$.*

VI. Numerical Example

Consider a system $\mathcal{S}$ as in (2), with state $x = (x_1, x_2) \in \mathbb{R}^2$, input $u \in \mathbb{R}$, and dynamics $\dot{x} = (x_2, u), y = (x_1, x_2)$ where $y_a = x_1$ is only available when there are no attacks. DoS attacks have maximum duration of $T_a = 1.6$ seconds and are launched only after at least $T_{na} = 0.047$ seconds without an attack. Here, u is designed such that every response $t \mapsto x(t)$ to $\mathcal{S}$ satisfies $x(t) \in S := \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + 2x_2^2 + 2x_1x_2 - 35 \leq 0\}$ for all $t \geq 0$, given that $x(0) \in X_0 := S \setminus (\partial S + \varepsilon\mathbb{B})$, with $\varepsilon = 2.01$.

An observer as in (8) is designed. Given that Assumption 1 is satisfied, and by setting $L = \left[\begin{smallmatrix} 32 & 0.5 \\ 0.5 & 32 \end{smallmatrix}\right]$ and $\tilde{L} = \left[\begin{smallmatrix} 0.05 \\ 3.2 \end{smallmatrix}\right]$, we have $\lambda(A - LC) = -31.75 \pm i0.43$, and $\lambda(A - \tilde{L}\tilde{C}) = \{0.5, -3.2\}$. Given $x_0 = (5.3, -2.4)$, $\hat{x}_0 = (4.9, -2.1)$, and $\bar{E} = 0.55$, we have that $|e(0)| = 0.5 \leq \bar{E}$, so Assumption 2 holds.

Thus, by applying Lemma 1, with $P = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right], Q = \left[\begin{smallmatrix} 63 & 0 \\ 0 & 64 \end{smallmatrix}\right]$, and given that every pair of subsequent attacks are separated by at least T_{na} seconds, the estimation error satisfies $|e(t)| \leq \gamma_1(t - t_2^i)e(t_2^i)$ for all $t \in [t_2^i, t_1^{i+1}]$, $i \in \mathbb{N}$, $\gamma_1(T_{na} = 0.047) = 0.226$, and is displayed in green[2] in Figure 1. Given that the growth rate of the exponential defining the function γ_1 is negative, the bound on the error norm decreases at each interval without attacks.

In addition, by applying Lemma 2 with $\hat{c}_1 = 1.12, \hat{c}_2 = 1.19, \hat{\lambda}_1 = 3.2, \hat{\lambda}_2 = 0.5, \hat{P} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right], \hat{Q} = \left[\begin{smallmatrix} 6.4 & 0 \\ 0 & 6.4 \end{smallmatrix}\right], \Phi = \left[\begin{smallmatrix} 1 & -0.25 \\ 0 & 0.97 \end{smallmatrix}\right]$, and $\hat{A}_{22} = 0.5$, given that every attack has a maximum duration of T_a seconds, the estimation error

[2]Code at https://github.com/HybridSystemsLab/SafeRecovery-DoSAttacks

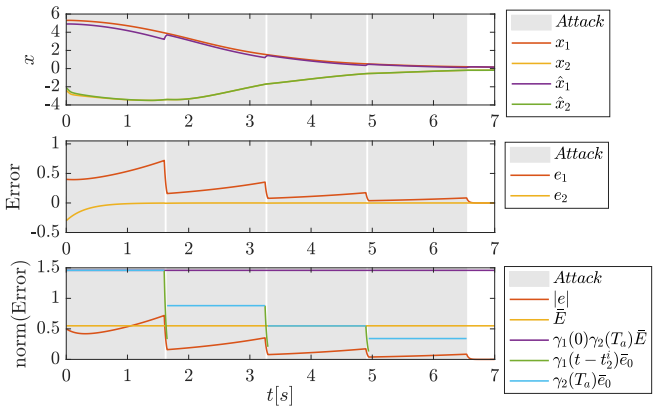

Fig. 1. Solutions to the 2D system and state estimation error during worst-case attacks of $T_a = 1.6s$, for $x_0 = (5.3, -2.4)$, $\hat{x}_0 = (4.9, -2.1)$, and $\bar{E} = 0.55$. In the third plot, the bound (purple) is defined as in Theorem 1.

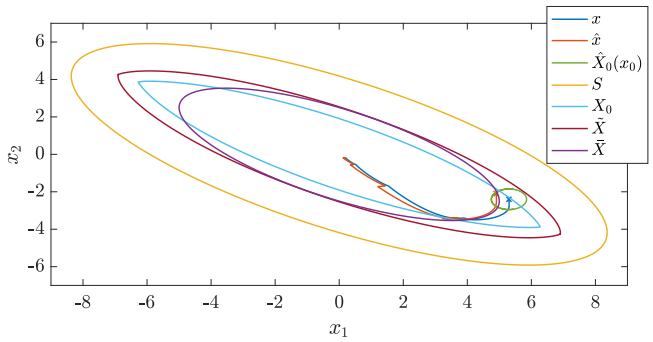

Fig. 2. Phase portrait of $\dot{x} = (x_2, u), y = (x_1, x_2)$ with state estimation for safe recovery of DoS attacks in the measurements of x_1. By initializing the estimation $\hat{x}$ in the $\bar{E}-$ball (green) around $x(0)$, the set $\bar{X}$ (purple) is rendered forward invariant for $\hat{x}$ (orange), and the safe set S (yellow) conditionally invariant for x (dark blue) with respect to the set of allowed initial states, namely X_0 (light blue), via the control barrier function. The set $\tilde{X}$ (scarlet) denotes the allowed initial observer states.

satisfies $|e(t)| \leq \gamma_2(T_a)|e(t_1^i)|$ for all $t \in [t_1^i, t_2^i]$, $i \in \mathbb{N}_{>0}$ where $\gamma_2(T_a) = 2.65$, and is displayed in light blue in Figure 1. Thanks to Theorem 1, given that $\gamma_1(T_{na})\gamma_2(T_1) \leq 1$, the error satisfies $|e(t)| \leq c_1\gamma_2(T_a)\bar{E} = 1.46$ for all $t \geq 0$.

In Figure 2, the set X_0 is a deflation of the set S by ε, and the set $\tilde{X}$ is an inflation of the set X_0 by $\bar{E}$. The set of initial estimations, $\hat{X}_0(x_0)$, is defined as the ball of radius $\bar{E}$ centered at x_0. Thus, the estimator $\hat{x}$ is initialized at $X_0(x_0) \subset \tilde{X}$. The set $\bar{X} := \{(x_1, x_2) \in \mathbb{R}^2 : h(x) \leq 0\} \subset \tilde{X}$ is defined by the barrier function $h(x) = x_1^2 + 2x_2^2 + 2x_1x_2 - 12.5$. Given that the set $S \subset \mathbb{R}^n$ is such that $S \setminus (\partial S + (1 + \gamma_1(0)\gamma_2(T_a))\bar{E}\mathbb{B}) \neq \emptyset$, by assigning $K = [2.3016, \ 2.3671]$ and solving the QPs (19) and (20) at every point of the trajectory $\hat{x}(t) \in \bar{X}$ to assign the input action, thanks to Theorem 2, we ensure that $\hat{x}(t) \in \bar{X}$ for all t, and consequently, $x(t) \in S$ for all t.

VII. Conclusion and Future Work

In this paper, we present a switched controller design that, together with a switched observer, ensures a linear time-invariant system to recover safely from finite-time DoS attacks in some of the system outputs. Conditional invariance of a set is guaranteed with respect to a subset of initial conditions by employing a barrier function approach and bounding the estimation error at all times. Future works include studying safe-recovery controllers under uncertainty in the model parameters, noise in the unattacked sensors, nonlinearities in the system dynamics, and only approximate information on the attack times. In addition, an implementation of a finite-time observer and a tighter bound to relax the conservatism of the conditions are to be considered.

References

[1] M. S. Chong, H. Sandberg, and A. M. Teixeira, "A tutorial introduction to security and privacy for cyber-physical systems," in *2019 18th European Control Conference (ECC)*. IEEE, 2019, pp. 968–978.

[2] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[3] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *2008 The 28th International Conference on Distributed Computing Systems Workshops*. IEEE, 2008, pp. 495–500.

[4] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Proceedings of the 12th International Conference on Hybrid Systems: Computation and Control*, vol. 5469. Springer, Berlin, Heidelberg, 2009, pp. 31–45.

[5] M. Krotofil, A. A. Cárdenas, B. Manning, and J. Larsen, "Cps: Driving cyber-physical systems to unsafe operating conditions by timing dos attacks on sensor signals," in *Proceedings of the 30th Annual Computer Security Applications Conference*, 2014, pp. 146–155.

[6] 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.

[7] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram, "Resilient asymptotic consensus in robust networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 766–781, 2013.

[8] S. H. Kafash, J. Giraldo, C. Murguia, A. A. Cardenas, and J. Ruths, "Constraining attacker capabilities through actuator saturation," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 986–991.

[9] Y. Yan, P. Antsaklis, and V. Gupta, "A resilient design for cyber physical systems under attack," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 4418–4423.

[10] C.-Z. Bai, V. Gupta, and F. Pasqualetti, "On kalman filtering with compromised sensors: Attack stealthiness and performance bounds," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6641–6648, 2017.

[11] A. D. Ames, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs with application to adaptive cruise control," in *53rd Conference on Decision and Control*. IEEE, 2014, pp. 6271–6278.

[12] A. Clark, Z. Li, and H. Zhang, "Control barrier functions for safe cps under sensor faults and attacks," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 796–803.

[13] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2009, pp. 31–45.

[14] M. Maghenem and R. G. Sanfelice, "Characterizations of safety and conditional invariance in dynamical systems," in *Proceedings of the American Control Conference*, July 2019, pp. 5039–5044.

[15] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2017.

[16] Y. Shen and X. Xia, "Semi-global finite-time observers for nonlinear systems," *Automatica*, vol. 44, no. 12, pp. 3152–3156, 2008.

[17] K. Garg, E. Arabi, and D. Panagou, "Fixed-time control under spatiotemporal and input constraints: A quadratic programming based approach," *Automatica*, vol. 141, p. 110314, 2022.