2020 59th IEEE Conference on Decision and Control (CDC)
Jeju Island, Republic of Korea, December 14-18, 2020

Control Barrier Functions for Safe CPS Under Sensor Faults and Attacks

Andrew Clark, Zhouchi Li, and Hongchao Zhang

Abstract— **CPS safety, defined as the system state remaining within a desired safe region, is a critical property in applications including medicine, transportation, and energy. Sensor faults and attacks may cause safety violations by introducing bias into the system state estimation, which in turn leads to erroneous control inputs. In this paper, we propose a class of Fault-Tolerant Control Barrier Functions (FT-CBFs) that provide provable guarantees on the safety of stochastic CPS. Our approach is to maintain a set of state estimators, each of which ignores a subset of sensor measurements that are affected by a particular fault pattern. We then introduce a linear constraint for each state estimator that ensures that the estimated state remains outside the unsafe region, and propose an approach to resolving conflicts between the constraints that may arise due to faults. We present sufficient conditions on the geometry of the safe region and the noise characteristics to provide a desired probability of maintaining safety. We then propose a framework for joint safety and stability by integrating FT-CBFs with Control Lyapunov Functions. Our approach is validated through a numerical study of a wheeled mobile robot.**

I. Introduction

A cyber-physical system (CPS) is safe if it remains within a predetermined safe region for all time. Safety is a fundamental requirement in critical applications including medicine, transportation, and energy, in which safety violations can cause catastrophic economic damage and loss of human life. The need to ensure safety in systems with dynamic environments, noisy and uncertain dynamics, and malicious attacks has resulted in substantial literature on design and verification of safe CPS [1], [2].

Safety is an especially challenging problem when the system dynamics are affected by faults and malicious attacks. Sensor faults occur when one or more sensors used to measure the system state provide arbitrary, inaccurate readings. Sensor faults affect safety in two ways. First, they may prevent the system from detecting and preventing safety violations. Second, they may bias estimates of the system state, leading to erroneous control signals that drive the true system state to an unsafe operating point. Both of these cases are especially damaging when sensors are compromised by malicious adversaries, who may deliberately design sensor signals to evade detection [3], [4]. There has been research attention on detecting sensor faults [5] and attacks [6], [7], as well as ensuring stability [8], [9] under sensor attacks.

In this paper, we propose sufficient conditions for a control policy to guarantee safety under one or more possible sensor faults. We develop our approach within the framework of *Control Barrier Functions (CBF)* [10]. An advantage of CBFs is that they can be readily integrated into existing control policies by adding linear constraints on the control input.

We propose a class of Fault-Tolerant Control Barrier Functions (FT-CBFs) for CPS with sensor faults, which we construct as follows. When there is one possible fault pattern, a straightforward approach is to add a CBF constraint on the state estimate produced by the other, non-faulty sensors. Typically, however, there are multiple possible fault patterns, each of which may cause a distinct set of sensors to fail. This shortcoming can be addressed by maintaining a set of state estimators, each omitting a set of sensors associated with one fault pattern, and then using CBFs to ensure that each of the estimated states remains within the safe region. Such an approach, however, may be impossible when faults occur and the state estimates deviate due to the presence of the fault, making it impossible to satisfy all CBF constraints.

In order to resolve such conflicts, we maintain a second set of estimators, each of which estimates the state using all sensors that do not belong to a given *pair* of fault patterns. Given two constraints that conflict with each other, we compare each state estimate to the corresponding estimator that excludes all sensors affected by both fault patterns. If the difference between the baseline exceeds a given threshold, we relax the corresponding constraint. We make the following specific contributions:

- We construct FT-CBFs and derive sufficient conditions to ensure that safety is satisfied with a desired probability.
- We consider half-plane and ellipsoidal safe regions and derive conditions on the problem geometry that ensure that there are no conflicts between CBFs.
- We compose CBFs with Control Lyapunov Functions (CLFs) to provide joint guarantees on the safety and stability of a desired goal set under faults.
- We evaluate our approach via a numerical study. The proposed control policy ensured convergence to a goal set without violating safety under sensor attack.

The paper is organized as follows. Section II reviews the related work. Section III states the problem formulation and gives background on CBFs. Section IV proposes a CBF-based control policy and gives sufficient conditions for safety. Section V proposes a framework for joint safety and stability via CBF-CLFs. Section VI presents simulation results. Section VII concludes the paper.

The authors are with the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA, USA 01609. Email: {aclark,zli4,hzhang9}@wpi.edu. This work was supported by NSF grant CNS-1941670 and ONR grant N00014-17-1-2946.

978-1-7281-7447-1/20/$31.00 ©2020 IEEE

II. Related Work

Fault detection and isolation in control systems has been studied for decades. See [11] for an in-depth treatment. Standard approaches include statistical hypothesis testing for stochastic systems [12], unknown input observers for deterministic systems [13], and sliding-mode control [14]. More recently, data-driven approaches to fault tolerance have shown promise [15]. Several of these works aim to guarantee stability in the presence of faults [16], which is related to but distinct from the safety criteria we consider. While the approach of using Kalman filter residues to identify potential faults is related to our conflict resolution approach, fault-tolerant control via CBFs has not been studied.

Related to fault-tolerant control is resilient control in the presence of sensor attacks, which differ from faults due to the adversary's ability to evade detection and bias the control to a desired operating point. A variety of schemes for detecting compromised sensors and computing state estimates in the presence of compromised sensor inputs have been proposed for deterministic and stochastic systems [3], [8], [17]–[22].

Safety verification of cyber-physical systems is an area of extensive research, with popular methods including finite-state approximations [23], barrier certificates [24], simulation-driven approaches [25], and counterexample-guided synthesis [26]. Among these methods, CBFs were proposed in [27]. CBFs for stochastic systems were investigated in [28], [29]. CBFs for high relative degree systems were presented in [30]–[34]. CBFs for safe reinforcement learning were introduced in [35]. Applications to domains such as multi-agent systems [36] and autonomous vehicles [37] have been considered. None of these existing works, however, incorporated the effects of faults and attacks.

III. Preliminaries and Problem Statement

This section presents the system model and problem statement. We then give background on the Extended Kalman Filter and control barrier functions.

A. System Model and Problem Statement

Notations. For a set S, let $\text{int}(S)$ and ∂S denote the interior and boundary of S, respectively. For any vector v, we let $[v]_i$ denote the i-th element of v. For a matrix c and set of indices S, we let $c(S)$ denote the matrix with rows indexed in S. We let $\overline{\lambda}(A)$ denote the magnitude of the largest eigenvalue of A, noting that this is equal to the largest eigenvalue when A is symmetric and positive definite. When the value of A is clear, we write $\overline{\lambda}$.

We consider a nonlinear control system with state $x_t \in \mathbb{R}^n$ and input $u_t \in \mathbb{R}^p$ at time t. The state dynamics are described by the stochastic differential equation

$$dx_t = (f(x_t) + g(x_t)u_t)\ dt + \sigma_t\ dW_t \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times p}$ are locally Lipschitz, $\sigma_t \in \mathbb{R}^{n\times n}$, and W_t is an n-dimensional Brownian motion.

The system output is denoted as $y_t \in \mathbb{R}^q$. The output may be affected by one of m faults. The set of possible faults is indexed as $\{r_1, \ldots, r_m\}$. Each fault r_i maps to a set of affected observations $\mathcal{F}(r_i) \subseteq \{1, \ldots, q\}$. We assume that $\mathcal{F}(r_i) \cap \mathcal{F}(r_j) = \emptyset$ for $i \neq j$. Let $r \in \{r_1, \ldots, r_m\}$ denote the index of the fault experienced by the system. The observation vector y_t has dynamics

$$dy_t = (cx_t + a_t)\ dt + \nu_t\ dV_t \tag{2}$$

where $c \in \mathbb{R}^{q\times n}$, $a_t \in \mathbb{R}^q$, $\nu_t \in \mathbb{R}^{q\times q}$, and V_t is a q-dimensional Brownian motion. The vector a_t represents the impact of the fault and is constrained by $\text{supp}(a_t) \subseteq \mathcal{F}(r)$. Hence, if fault r occurs, then the outputs of any of the sensors indexed in $\mathcal{F}(r_i)$ can be arbitrarily modified by the fault. The sets $\mathcal{F}(r_1), \ldots, \mathcal{F}(r_m)$ are known, but the value of r is unknown. In other words, the set of possible faults is known, but the exact fault that has occurred is unknown to the controller. Define $\overline{f}(x, u) = f(x) + g(x)u$. We assume that the system is controllable. The detectability property is defined as follows.

Definition 1: The pair $[\frac{\partial \overline{f}}{\partial x}(x, u), c]$ is uniformly detectable if there exists a bounded, matrix-valued function $\Theta(x)$ and a real number $\eta > 0$ such that

$$w^T \left(\frac{\partial \overline{f}}{\partial x}(x, u) + \Theta(x)c \right) w \leq -\eta ||w||^2$$

for all w, u, and x.

We assume that, for each $i, j \in \{1, \ldots, m\}$, the pair $[\frac{\partial \overline{f}}{\partial x}(x, u), c(\{1, \ldots, q\} \setminus (\mathcal{F}(r_i) \cup \mathcal{F}(r_j)))]$ is uniformly detectable. In other words, if we compute an estimate that does not incorporate data from sensors affected by any pair of faults, then that estimate satisfies uniform detectability. The safe region of the system is a set $\mathcal{C} \subseteq \mathbb{R}^n$ defined by

$$\mathcal{C} = \{x : h(x) \geq 0\}, \quad \partial\mathcal{C} = \{x : h(x) = 0\} \tag{3}$$

where $h : \mathbb{R}^n \to \mathbb{R}$ is twice-differentiable on $\mathcal{C}$. We assume throughout the paper that $x_0 \in \text{int}(\mathcal{C})$, i.e., the system is initially safe.

Problem Statement: Given a set $\mathcal{C}$ defined as above and a parameter $\epsilon \in (0, 1)$, construct a control policy that, at each time t, maps the sequence $\{y_{t'} : t' \in [0, t)\}$ to an input u_t and, for any fault $r \in \{r_1, \ldots, r_m\}$, $Pr(x_t \in \mathcal{C}\ \forall t) \geq (1-\epsilon)$ when fault r occurs.

B. Background and Preliminary Results

The Extended Kalman Filter (EKF) for the system

$$dx_t = (f(x_t) + g(x_t)u_t)\ dt + \sigma_t\ dW_t \tag{4}$$

$$dy_t = cx_t\ dt + \nu_t\ dV_t \tag{5}$$

is defined by

$$d\hat{x}_t = (f(\hat{x}_t) + g(\hat{x}_t)u_t)dt + K_t(dy_t - c\hat{x}_t),$$

where $K_t = P_t c^T R_t^{-1}$ and $R_t = \nu_t \nu_t^T$. The matrix P_t is the positive-definite solution to

$$\frac{dP}{dt} = A_t P_t + P_t A_t^T + Q_t - P_t c^T R_t^{-1} c P_t$$

where $Q_t = \sigma_t \sigma_t^T$ and $A_t = \frac{\partial \overline{f}}{\partial x}(\hat{x}_t, u_t)$. We first introduce the following assumptions.

Assumption 1: The SDEs (1) and (2) satisfy the conditions:

1) There exist constants β_1 and β_2 such that $\mathbf{E}(\sigma_t\sigma_t^T) \geq \beta_1 I$ and $\mathbf{E}(\nu_t\nu_t^T) \geq \beta_2 I$ for all t.
2) The pair $[\frac{\partial \overline{f}}{\partial x}(x,u), c]$ is uniformly detectable.
3) Let ϕ be defined by

$$\overline{f}(x,u) - \overline{f}(\hat{x},u) = \frac{\partial \overline{f}}{\partial x}(x-\hat{x}) + \phi(x,\hat{x},u).$$

Then there exist real numbers k_ϕ and ϵ_ϕ such that

$$||\phi(x,\hat{x},u)|| \leq k_\phi ||x-\hat{x}||_2^2$$

for all x and $\hat{x}$ satisfying $||x-\hat{x}||_2 \leq \epsilon_\phi$.

The following result describes the accuracy of the EKF.

Theorem 1 ([38]): Suppose that the conditions of Assumption 1 hold. Then there exists $\delta > 0$ such that if $\sigma_t\sigma_t^T \leq \delta I$ and $\nu_t\nu_t^T \leq \delta I$, then for any $\epsilon > 0$, there exists $\gamma > 0$ such that

$$Pr\left(\sup_{t\geq 0} ||x_t - \hat{x}_t||_2 \leq \gamma\right) \geq 1-\epsilon.$$

For the remainder of the paper, we assume that the system satisfies the conditions of Theorem 1. We note that for observable linear systems, the conditions of Theorem 1 hold with $\delta = \infty$. We next provide background and preliminary results on control barrier functions. The following theorem provides sufficient conditions for safety.

Theorem 2: For a system (4)–(5) with safety region defined by (3), define

$$\overline{h}_\gamma = \sup\{h(x) : ||x-x^0||_2 \leq \gamma \text{ for some } x^0 \in h^{-1}(\{0\})\}$$

and $\hat{h}(x) = h(x) - \overline{h}_\gamma$. Let $\hat{x}_t$ denote the EKF estimate of x_t, and suppose that there exists a constant $\delta > 0$ such that whenever $\hat{h}(\hat{x}_t) < \delta$, u_t is chosen to satisfy

$$\frac{\partial h}{\partial x}(\hat{x}_t)\overline{f}(\hat{x}_t,u_t) - \gamma||\frac{\partial h}{\partial x}(\hat{x}_t)K_t c||_2 + \frac{1}{2}\mathbf{tr}\left(\nu_t^T K_t^T \frac{\partial^2 h}{\partial x^2}(\hat{x}_t)K_t\nu_t\right) \geq -\hat{h}(\hat{x}_t). \quad (6)$$

Then $Pr(x_t \in \mathcal{C}\ \forall t|\ ||x_t - \hat{x}_t||_2 \leq \gamma\ \forall t) = 1$.

The proof of Theorem 2 is very similar to the proof of Theorem 2 from [29] and is omitted due to space constraints.

We call a function h satisfying (6) a *Stochastic Control Barrier Function (SCBF)*. Intuitively, Eq. (6) implies that as the state approaches the boundary, the control input is chosen such that the rate of increase of the barrier function decreases to zero. Hence Theorem 2 implies that if there exists an SCBF for a system, then the safety condition is satisfied with probability $(1-\epsilon)$ when an EKF is used as an estimator and the control input is chosen at each time t to satisfy (6).

IV. Proposed Safe Control Strategy

This section presents our proposed CBF-based approach to safe control. We first describe the control policy and derive conditions for it to guarantee safety with the desired probability. We then analyze these conditions for special cases of the safe region. Finally, we discuss computation of parameters associated with the control policy.

A. Control Policy Definition

The intuition behind our approach is as follows. If the fault pattern r is known, then safety can be guaranteed with probability $(1-\epsilon)$ by constructing an estimator that ignores the sensor measurements from the set $\mathcal{F}(r)$, defining an SCBF, and then applying a linear constraint to the control input derived from Eq. (6). Since the fault pattern is unknown, we can instead maintain a set of m EKFs and m SCBFs, each corresponding to a different possible fault pattern in $\{r_1,\ldots,r_m\}$, and each resulting in a different linear constraint on the control input.

The potential drawback of this approach, however, is that it may be infeasible to select a control input that satisfies all m constraints simultaneously at time t, particularly when faulty sensor measurements cause the state estimates to diverge. We resolve conflicts between the constraints by defining a set of $\binom{m}{2}$ EKFs, each of which omits all sensors affected by either fault r_i or fault r_j for some $i,j \in \{1,\ldots,m\}$. These estimators are used to remove conflicting constraints.

The policy is defined formally as follows. Define $\overline{c}_i$ to be the c matrix with the rows indexed in $\mathcal{F}(r_i)$ removed, $\overline{y}_{t,i}$ to be equal to the vector y with the entries indexed in $\mathcal{F}(r_i)$ removed, and $\overline{\nu}_{t,i}$ to be the matrix ν_t with rows and columns indexed in $\mathcal{F}(r_i)$ removed. Let $\overline{R}_{t,i} = \overline{\nu}_{t,i}\overline{\nu}_{t,i}^T$ and $K_{t,i} = \overline{P}_{t,i}\overline{c}_i^T(\overline{R}_{t,i})^{-1}$. Here $\overline{P}_{t,i}$ is the solution to the Riccati differential equation

$$\frac{d\overline{P}_{t,i}}{dt} = A_{t,i}\overline{P}_{t,i} + \overline{P}_{t,i}A_{t,i}^T + Q_t - \overline{P}_{t,i}\overline{c}_i^T\overline{R}_{t,i}^{-1}\overline{c}_i\overline{P}_{t,i}$$

with $A_{t,i} = \frac{\partial \overline{f}}{\partial x}(\hat{x}_{t,i},u_t)$. Define a set of m EKFs with estimates denoted $\hat{x}_{t,i}$ via

$$d\hat{x}_{t,i} = (f(\hat{x}_{t,i}) + g(\hat{x}_{t,i})u_t)\ dt + K_{t,i}(d\overline{y}_{t,i} - \overline{c}_i\hat{x}_{t,i}\ dt). \quad (7)$$

Each of these EKFs represents the estimate obtained by removing the sensors affected by fault r_i. Furthermore, define $\overline{y}_{t,i,j}$, $\overline{\nu}_{t,i,j}$, $\overline{c}_{i,j}$, $\overline{R}_{t,i,j}$, and $K_{t,i,j}$ in an analogous fashion with entries indexed in $\mathcal{F}(r_i)\cup\mathcal{F}(r_j)$ removed. We assume that the $\overline{R}$ matrices are invertible. We then define a set of $\binom{m}{2}$ estimators $\hat{x}_{t,i,j}$ as

$$d\hat{x}_{t,i,j} = (f(\hat{x}_{t,i,j}) + g(\hat{x}_{t,i,j})u_t)\ dt + K_{t,i,j}(d\overline{y}_{t,i,j} - \overline{c}_{i,j}\hat{x}_{t,i,j}\ dt). \quad (8)$$

When $\mathcal{F}(r_i)\cup\mathcal{F}(r_j) = \{1,\ldots,q\}$, the open-loop estimator is used for $\hat{x}_{t,i,j}$.

We then select parameters $\gamma_1,\ldots,\gamma_m \in \mathbb{R}_+$, and $\{\theta_{ij} : i<j\} \subseteq \mathbb{R}_+$, $\delta > 0$. The set of feasible control actions is defined at each time t using the following steps:

1) Define $X_t(\delta) = \{i : \hat{h}_i(\hat{x}_{t,i}) < \delta\}$. Let $Z_t = X_t(\delta)$. Define a collection of sets Ω_i, $i \in Z_t$, by

$$\Omega_i \triangleq \left\{u : \frac{\partial h_i}{\partial x}(\hat{x}_{t,i})\overline{f}(\hat{x}_{t,i},u_t) - \gamma_i||\frac{\partial h}{\partial x}(\hat{x}_{t,i})K_{t,i}c||_2 + \frac{1}{2}\mathbf{tr}(\overline{\nu}_{t,i}^T K_{t,i}^T \frac{\partial^2 h_i}{\partial x^2}(\hat{x}_{t,i})K_{t,i}\overline{\nu}_{t,i}) \geq -\hat{h}_i(\hat{x}_{t,i})\right\}. \quad (9)$$

Select u_t satisfying $u_t \in \bigcap_{i \in X_t(\delta)} \Omega_i$. If no such u_t exists, then go to Step 2.

2) For each i, j with $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 > \theta_{ij}$, set $Z_t = Z_t \backslash \{i\}$ (resp. $Z_t = Z_t \setminus \{j\}$) if $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$ (resp. $||\hat{x}_{t,j} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$). If $\bigcap_{i \in Z_t} \Omega_i \neq \emptyset$, then select $u_t \in \bigcap_{i \in Z_t} \Omega_i$. Else go to Step 3.
3) Remove the indices i from Z_t corresponding to the estimators with the largest residue values $\overline{y}_{t,i} - \overline{c}_i \hat{x}_{t,i}$ until there exists $u_t \in \bigcap_{i \in Z_t} \Omega_i$.

This policy attempts to select a control input that guarantees safety regardless of the fault pattern that is experienced (Step 1). If no such input exists, then the set of constraints is pruned by looking for constraints Ω_i such that $\hat{x}_{t,i}$ deviates from $\hat{x}_{t,i,j}$ by more than a threshold value, since such deviations are likely to be due to faults. Meanwhile, if the fault pattern is r_i, then the estimates $\hat{x}_{t,i}$ and $\hat{x}_{t,i,j}$ will likely be close to one another for all t and $j \neq i$, since both estimators do not rely on the faulted sensor. We note that these constraints are compatible with feedback policies as well as more general history-based control policies.

At each time t, this policy requires maintaining $m + \binom{m}{2}$ EKFs, checking $\binom{m}{2}$ inequalities of the form $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$ in the worst-case, and checking the feasibility of m linear inequalities. The following result gives sufficient conditions for this control policy to guarantee safety.

Theorem 3: Define

$$\overline{h}_{\gamma_i} = \sup \big\{ h(x) : ||x - x^0||_2 \leq \gamma_i \quad \text{for some } x^0 \in h^{-1}(\{0\}) \big\}$$

and $\hat{h}_i(x) = h(x) - \overline{h}_{\gamma_i}$. Suppose $\gamma_1, \ldots, \gamma_m$, and θ_{ij} for $i < j$ are chosen such that the following conditions are satisfied:

1) Define $\Lambda_i(\hat{x}_{t,i}) = \frac{\partial h_i}{\partial x}(\hat{x}_{t,i}) g(\hat{x}_{t,i})$. There exists $\delta > 0$ such that for any $X_t' \subseteq X_t(\delta)$ satisfying $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 \leq \theta_{ij}$ for all $i, j \in X_t'$, there exists u such that
$$\Lambda_i(\hat{x}_{t,i}) u > 0 \tag{10}$$
for all $i \in X_t'$.
2) For each i, when $r = r_i$,
$$Pr(||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \leq \theta_{ij}/2 \, \forall j, ||\hat{x}_{t,i} - x_t||_2 \leq \gamma_i \, \forall t) \geq 1 - \epsilon. \tag{11}$$

Then $Pr(x_t \in \mathcal{C} \ \forall t) \geq 1 - \epsilon$ for any fault pattern $r \in \{r_1, \ldots, r_m\}$.

Proof: Suppose that $r = r_i$. We will show that if $||\hat{x}_{t,i} - x_t||_2 \leq \gamma_i$ and $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \leq \theta_{ij}/2$ for all t, then $u_t \in \Omega_i$ holds whenever $\hat{h}_i(\hat{x}_{t,i}) < \delta$. Hence $x_t \in \mathcal{C}$ for all t by Theorem 2.

At time t, suppose that $\hat{h}_i(\hat{x}_{t,i}) < \delta$, so that $i \in X_t(\delta)$, and that $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \leq \theta_{ij}/2$. We consider three cases, namely (i) $||\hat{x}_{t,j} - \hat{x}_{t,k}||_2 \leq \theta_{jk}$ for all $j, k \in X_t(\delta)$, (ii) $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 \leq \theta_{ij}$ for all $j \in X_t(\delta)$, but there exist $j, k \in X_t(\delta) \setminus \{i\}$ such that $||\hat{x}_{t,j} - \hat{x}_{t,k}||_2 > \theta_{jk}$, and (iii) $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 > \theta_{ij}$ for some $j \in X_t(\delta)$.

Case (i): We will show that there exists $u \in \cap_{j \in X_t(\delta)} \Omega_j$, and hence in particular u_t satisfies Ω_i. Each Ω_j can be written in the form

$$\Omega_j = \{u : \Lambda_j(\hat{x}_{t,j}) u_t \geq \overline{\omega}_j\} \tag{12}$$

where $\overline{\omega}_j$ is a real number that does not depend on u_t. Under the assumption 1) of the theorem, there exists u satisfying (10) for all $i \in X_t(\delta)$. Choose

$$u_t = \left(\max_j \{|\overline{\omega}_j|\} / ||u||_2 \right) u.$$

This choice of u_t satisfies $u_t \in \bigcap_{j \in X_t(\delta)} \Omega_j$, in particular $u_t \in \Omega_i$.

Case (ii): In this case, Step 2 of the procedure is reached and constraints Ω_j are removed until all indices in Z_t satisfy $||\hat{x}_{t,j} - \hat{x}_{t,k}||_2 \leq \theta_{jk}$. Since $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 \leq \theta_{ij}$ already holds for all $j \in X_t(\delta)$, i will not be removed from Z_t during this step. After Step 2 is complete, the analysis of Case (i) holds and there exists a u which satisfies all the remaining constraints, including Ω_i.

Case (iii): Suppose j satisfies $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 > \theta_{ij}$. We have

$$\begin{aligned} \theta_{ij} &< ||\hat{x}_{t,i} - \hat{x}_{t,i,j} + \hat{x}_{t,i,j} - \hat{x}_{t,j}||_2 \\ &\leq ||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 + ||\hat{x}_{t,i,j} - \hat{x}_{t,j}||_2 \qquad (13) \\ &\leq \theta_{ij}/2 + ||\hat{x}_{t,i,j} - \hat{x}_{t,j}||_2 \qquad (14) \end{aligned}$$

where Eq. (13) follows from the triangle inequality and (14) follows from the assumption that $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \leq \theta_{ij}/2$. Hence $||\hat{x}_{t,j} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$ and j is removed from Z_t. By applying this argument to all such indices j, we have that i is not removed during Step 2 of the procedure, and thus the analyses of Cases (i) and (ii) imply that $u_t \in \Omega_i$.

From these cases, we have that Ω_i holds whenever $\hat{h}_i(\hat{x}_{t,i}) < \delta$. Therefore, by Theorem 2,

$$Pr(x_t \in \mathcal{C} \ \forall t | ||\hat{x}_{t,i} - \hat{x}_t||_2 \leq \gamma_i, ||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \leq \theta_{ij}/2 \, \forall t) = 1$$

and $Pr(x_t \in \mathcal{C} \ \forall t) > 1 - \epsilon$ by (11). ■

If functions $h_1, \ldots, h_m$ that satisfy the conditions of Theorem 3, then they are referred to as *Fault-Tolerant Control Barrier Functions (FT-CBF).*

B. FT-CBF Construction

The conditions of Theorem 3 are not guaranteed to hold and depend on the system dynamics, level of noise, and the geometry of the safe region. In what follows, we develop sufficient conditions for LTI systems with dynamics

$$dx_t = (Fx_t + Gu_t) \, dt + \sigma dW_t. \tag{15}$$

We consider two cases of the safe region, namely, safe regions defined by half-planes and safe regions defined by ellipsoids. Since Eq. (11) depends on the accuracy of the estimator instead of the set $\mathcal{C}$, we will focus on satisfying constraint (10).

1) Half-Plane Constraint with LTI System: We first consider constraints of the form $h(x) = a^T x - b$. In this case, $\nabla \hat{h}_i(x) = a^T$ for all i and x, and hence $\Lambda_i(\hat{x}_{t,i}) = a^T G$.

Lemma 1: Suppose that $a^T G \neq 0$. Then at each time t, there exists u satisfying (10).

Proof: For any values of $\hat{x}_{t,i}$, we can choose an index $l \in \{1, \ldots, p\}$ such that $[a^T G]_l \neq 0$, set $[u]_s = 0$ for $s \neq l$ and select $[u]_l > 0$ if $[a^T G]_l > 0$ and $[u]_l < 0$ if $[a^T G]_l < 0$. Hence, for all $\hat{x}_{t,i}$ we can choose u satisfying (10). ■

Next, we consider the case where $a^T G = 0$. If the LTI system is controllable, then there exists a minimum i such that $a^T F^i G \neq 0$. Define a set of functions $h_0, \ldots, h_i$ as $h_0 = h(x)$,

$$h_{k+1}(x) = \frac{\partial h_k}{\partial x} F x + \frac{1}{2} \mathbf{tr}\left(\sigma^T \left(\frac{\partial^2 h_k}{\partial x^2}\right) \sigma\right) - \gamma || \frac{\partial h_k}{\partial x}(x) K c||_2 + h_k(x).$$

Define $\mathcal{C}_k = \{x : h_k(x) \geq 0\}$. The following result gives a sufficient condition for safety in this case.

Theorem 4 ([29]): Suppose that $x_0 \in \bigcap_{k=0}^{i} \mathcal{C}_k$ and, for all t,

$$\frac{\partial h_i}{\partial x} g(x) u \geq -\frac{\partial h_i}{\partial x} f(x) - \frac{1}{2} \mathbf{tr}\left(\sigma^T \frac{\partial^2 h_i}{\partial x^2} \sigma\right) - \gamma || \frac{\partial h_k}{\partial x}(x) K c||_2 - h_i(x). \quad (16)$$

Then $Pr(x_t \in \mathcal{C} \ \forall t) = 1$. Furthermore, $\frac{\partial h_i}{\partial x} G u = a^T F^i G u$. As a corollary to Theorem 4, we can choose an index $l \in \{1, \ldots, p\}$ such that $[a^T F^i G]_l \neq 0$, set $[u]_s = 0$ for $s \neq l$, and select $[u]_l > 0$ if $[a^T F^i G]_l > 0$ and $[u]_l < 0$ if $[a^T F^i G]_l < 0$. Hence a high relative degree half-plane constraint can be satisfied with the desired probability.

2) Ellipsoid Constraint with LTI System: We next consider an ellipsoid constraint of the form $\mathcal{C} = \{x : (x - x')^T \Phi (x - x') \leq 1\}$ for some positive definite matrix Φ and $x' \in \mathbb{R}^n$, so that $h(x) = 1 - (x - x')^T \Phi (x - x')$. We therefore have $\hat{h}_i(\hat{x}) = 1 - \overline{h}_{\gamma_i} - (x - x')^T \Phi (x - x')$. The gradient of h_i is then given by $\nabla h_i(x) = -2((x - x')^T \Phi)$. In this case,

$$\Lambda_i(\hat{x}_{t,i}) = -2(\hat{x}_{t,i} - x')^T \Phi G.$$

We first consider the case where $\text{rank}(G) = n$. Define $\overline{\theta} = \max\{\theta_{ij} : i < j\}$ and $\overline{h} = \max\{\overline{h}_{\gamma_i} : i = 1, \ldots, m\}$.

Proposition 1: Suppose $\overline{\theta} \leq \sqrt{\frac{2(1-\overline{h})}{\lambda(\Phi)}}$ and $\text{rank}(G) = n$. Then there exists δ such that (10) is satisfied.

Proof: Choose $\delta < 1 - \overline{h} - \frac{1}{2}\overline{\lambda\theta}^2$. We select u such that $Gu = (\hat{x}_{t,i} - x')$ for some $i \in X_t'$. For this choice of u, we have that $\Lambda_j(\hat{x}_{t,j})u$ is proportional to $(\hat{x}_{t,j} - x')^T \Phi (\hat{x}_{t,i} - x')$. We therefore need to show $(\hat{x}_{t,j} - x')^T \Phi (\hat{x}_{t,i} - x') > 0$.

Let $z_{t,i} = \Phi^{1/2}(\hat{x}_{t,i} - x')$ and $z_{t,j} = \Phi^{1/2}(\hat{x}_{t,j} - x')$. We have $||z_{t,i}||_2^2 \in (1 - \delta - \overline{h}, 1 - \overline{h})$, $||z_{t,j}||_2^2 \in (1 - \delta - \overline{h}, 1 - \overline{h})$, and

$$||z_{t,i} - z_{t,j}||_2 = ||\Phi^{1/2}(\hat{x}_{t,i} - \hat{x}_{t,j})||_2 \leq ||\Phi^{1/2}||_2 ||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 \leq \overline{\theta}\sqrt{\lambda}$$

Furthermore,

$$(\hat{x}_{t,i} - x')^T \Phi (\hat{x}_{t,j} - x') = z_{t,i}^T z_{t,j} = ||z_{t,i}||_2 ||z_{t,j}||_2 \cos\zeta,$$

where ζ is the angle between $z_{t,i}$ and $z_{t,j}$. By the law of cosines,

$$\cos\zeta = \frac{||z_{t,i}||_2^2 + ||z_{t,j}||_2^2 - ||z_{t,i} - z_{t,j}||_2^2}{2||z_{t,i}||_2 ||z_{t,j}||_2} \geq \frac{2(1 - \delta - \overline{h}) - \overline{\lambda\theta}^2}{2||z_{t,i}||_2 ||z_{t,j}||_2} > 0$$

due to the choice of δ. Hence $\Lambda_j(\hat{x}_{t,j})u > 0$ for all j. ■

When $\text{rank}(G) < n$, the above approach may be insufficient to ensure the existence of an FT-CBF, since $-(\hat{x}_{t,i} - x')$ might not be in the span of G. We next propose two sufficient conditions for an FT-CBF to guarantee safety. The first is a condition on the state trajectory, which can be used to guide offline trajectory planning. The second approach imposes additional constraints that reduce the size of $\mathcal{C}$ but ensure the existence of an FT-CBF. Define H as the projection matrix onto the span of $\Phi^{1/2}G$ and define $\overline{H}$ as the projection onto the orthogonal space to the span of $\Phi^{1/2}G$.

Proposition 2: Let $r = r_i$ and suppose there exist ϕ and δ such that

$$\frac{(\hat{x}_{t,i} - x')^T \Phi^{1/2} H \Phi^{1/2} (\hat{x}_{t,i} - x')}{(\hat{x}_{t,i} - x')^T \Phi (\hat{x}_{t,i} - x')} \geq \phi \quad (17)$$

whenever $i \in X_t(\delta)$. If $\overline{\theta}^2 \leq \frac{(1-\delta-\overline{h})\phi}{\overline{\lambda}}$, then at each time t with $i \in X_t(\delta)$ there exists u satisfying (10).

The proof of Proposition 2 is omitted due to space constraints.

One approach to ensuring safety in the presence of faults when $\text{rank}(G) < n$ is to introduce an auxiliary half-plane constraint of the form $(x - x')^T \Phi v < 0$, which changes the safe region from $\mathcal{C}$ to $\overline{\mathcal{C}} \triangleq \mathcal{C} \cap \{x : (x - x')^T \Phi v < 0\}$. This constraint ensures that there is a control input u satisfying (10) at each time step, as shown by the following theorem.

Theorem 5: At each time t, if $\hat{x}_{t,i} \in \overline{\mathcal{C}}$ for all i, then there exists u_t such that

$$\frac{\partial h_i}{\partial x}(\hat{x}_{t,i}) G u_t < 0 \quad (18)$$

$$-v^T G u_t < 0 \quad (19)$$

Furthermore, if u_t satisfies (18) and (19) at each time step, then $Pr(x_t \in \mathcal{C}) > 1 - \epsilon$.

Proof: Select u_t such that $Gu_t = v$. The gradient associated with the auxiliary half-plane constraint is given by $-v^T v < 0$, and hence (19) is satisfied. The gradient for the ellipsoid constraint is equal to $(\hat{x}_{t,i} - x')^T \Phi v < 0$ by choice of v for each i, and hence (18) is satisfied. Both equations imply that Eq. (10) holds for both the set $\mathcal{C}$ and the set $\{x : (x - x')^T \Phi v < 0\}$. Thus $x_t \in \overline{\mathcal{C}}$ for all t with probability at least $(1 - \epsilon)$ by Theorem 3. ■

C. Computation of θ, γ, and $\overline{h}_\gamma$

The computation of γ_i, $i = 1, \ldots, m$, and θ_{ij} for $i < j$ is briefly considered as follows. For the parameter γ_i, we observe that for an LTI system, $(x_t - \hat{x}_{t,i})$ is a Gaussian random process with mean 0 and covariance matrix P_t, where P_t is the solution to the Riccati equation

$$\frac{dP}{dt} = FP_t + P_tF^T + \sigma_t - P_tC^T\nu_{t,i}^{-1}CP_t.$$

The minimum γ satisfying $Pr(||\hat{x}_{t,i} - x_t||_2 > \gamma) < 1 - \epsilon$ can be computed based on this distribution.

In the case of θ_{ij}, a simple bound can be obtained by using the fact that $(\hat{x}_{t,i} - x_t)$ and $(\hat{x}_{t,i,j} - x_t)$ are both Gaussian processes described above. Hence, $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \leq ||\hat{x}_{t,i} - x_t||_2 + ||\hat{x}_{t,i,j} - x_t||_2$, and $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2$ can be bounded above by deriving bounds on each of the two terms.

Computation of $\overline{h}_\gamma$ is described for half-plane constraints in [28]. For ellipsoid constraints, we have the following closed form for $\overline{h}_\gamma$. The proof is omitted.

Proposition 3: Suppose that $h(x) = 1-(x-x')^T\Phi(x-x')$ where Φ is a positive definite matrix and $x' \in \mathbb{R}^n$. Then

$$\overline{h}_\gamma = \begin{cases} 1, & \gamma \geq \frac{1}{\sqrt{\overline{\lambda}(\Phi)}} \\ 1 - \left(1 - \gamma\sqrt{\overline{\lambda}(\Phi)}\right)^2, & \text{else} \end{cases}$$

V. Joint Safety and Stability

This section presents a framework for jointly ensuring safety and stability in systems with faults via Control Lyapunov Functions (CLFs) and CBFs. Such an approach has been widely used in fault-free scenarios. We first give the problem statement, followed by our proposed joint CBF-CLF based policy and results on the CBF-CLF construction.

The stability problem is stated as follows. Define the goal set $\mathcal{G}$ by $\mathcal{G} = \{x : w(x) \geq 0\}$ for some function w. The goal of the system is to asymptotically approach the set $\mathcal{G}$ with some desired probability. Our approach towards satisfying this constraint is through the use of *stochastic Control Lyapunov Functions*. A function $V : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is a stochastic CLF for the SDE (1) if, for each x, we have

$$\inf_u \left\{ \frac{\partial V}{\partial x} f(x) + \frac{\partial V}{\partial x} g(x)u + \frac{1}{2}\mathbf{tr}\left(\sigma^T \frac{\partial^2 V}{\partial x^2}\sigma\right)\right\} < -\rho V(x_t) \tag{20}$$

for some $\rho > 0$. The parameter ρ can be chosen to increase the convergence rate of the algorithm, at the cost of potentially making the condition (20) infeasible.

We next state a control policy that combines CLFs and CBFs to ensure safety and stability. Define a threshold parameter $\overline{V}$. At each time t, the set of feasible control actions is defined as follows:

1) Define $Y_t(\overline{V}) = \{j : V(\hat{x}_{t,j}) > \overline{V}\}$, and initialize $U_t = Y_t(\overline{V})$. Define a collection of sets Υ_i, $i \in U_t$, by

$$\Upsilon_i \triangleq \left\{ u : \frac{\partial V_i}{\partial x}\overline{f}(\hat{x}_{t,i}, u) + \gamma_i||\frac{\partial V_i}{\partial x}(\hat{x}_{t,i})K_tc||_2 + \frac{1}{2}\mathbf{tr}\left(\overline{\nu}_{t,i}^T K_{t,i}^T \frac{\partial^2 V}{\partial x^2}(\hat{x}_{t,i})K_{t,i}\overline{\nu}_{t,i}\right) < -\rho V(\hat{x}_{t,i})\right\} \tag{21}$$

for $i = 1, \ldots, m$. Select any

$$u_t \in \left(\bigcap_{i \in Z_t} \Omega_i\right) \cap \left(\bigcap_{j \in U_t} \Upsilon_j\right),$$

where Ω_i is defined as in (9). If no such u_t exists, go to Step 2.

2) For each i, j with $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 > \overline{\theta}_{ij}$, set $Z_t = Z_t \backslash \{i\}$ and $U_t = U_t \setminus \{i\}$ (resp. $Z_t = Z_t \setminus \{j\}$ and $U_t = U_t \setminus \{j\}$) if $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$ (resp. $||\hat{x}_{t,j} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$). If

$$\left(\bigcap_{i \in Z_t} \Omega_i\right) \cap \left(\bigcap_{j \in U_t} \Upsilon_j\right) \neq \emptyset,$$

then select u_t from this set. Else go to Step 3.

3) Remove the sets Ω_i and Υ_i corresponding to the estimators with the largest residue values until there exists a feasible u_t.

This policy is similar to the CBF-based approach of Section IV, with additional constraints to satisfy the stability condition. This leads to another m linear inequalities.

We omit the analysis of this scheme due to space constraints. A controller that reaches a goal set defined by a function V while satisfying a safety constraint $\mathcal{C} = \{x : h(x) \geq 0\}$ can be obtained by solving

$$\begin{array}{ll} \text{minimize} & u_t^T R u_t \\ \text{s.t.} & u_t \in \bigcap_{i \in Z_t} \Omega_i \quad \text{(CBF)} \\ & u_t \in \bigcap_{j \in U_t} \Upsilon_j \quad \text{(CLF)} \end{array} \tag{22}$$

at each time step, where R is a positive definite matrix representing the cost of exerting control.

VI. Case Study

A simulation of our approach on a wheeled mobile robot (WMR) is described as follows. We first describe the system model. We then present the results of the simulation.

A. System Model

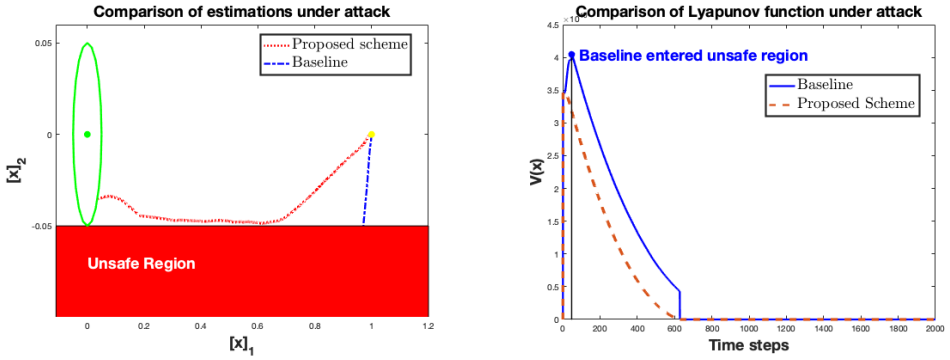
We consider a WMR with dynamics

$$\begin{pmatrix} [\dot{x}_t]_1 \\ [\dot{x}_t]_2 \\ \dot{\theta}_t \end{pmatrix} = \begin{pmatrix} \cos\theta_t & 0 \\ \sin\theta_t & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} [\omega_t]_1 \\ [\omega_t]_2 \end{pmatrix} + \mathbf{w}_t \tag{23}$$

where $([x_t]_1, [x_t]_2, \theta_t)^T$ is the vector of the horizontal, vertical, and orientation coordinates for the wheeled mobile robot, $([\omega_t]_1, [\omega_t]_2)^T$ (the linear velocity of the robot and the angular velocity around the vertical axis) is taken as the control input, and $\mathbf{w}_t$ is the process noise.

The feedback linearization [39] is utilized to transform the original state vector and the WMR model into the new state variable $x_t = ([x_t]_1, [x_t]_2, [\dot{x}_t]_1, [\dot{x}_t]_2)^T$ and the controllable linearized model

$$\begin{pmatrix} [\dot{x}_t]_1 \\ [\dot{x}_t]_2 \\ [\ddot{x}_t]_1 \\ [\ddot{x}_t]_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} [x_t]_1 \\ [x_t]_2 \\ [\dot{x}_t]_1 \\ [\dot{x}_t]_2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} [u_t]_1 \\ [u_t]_2 \end{pmatrix} + \mathbf{w}_t' \tag{24}$$



(a) (b)

Fig. 1: Evaluation of our proposed approach on a linearized wheeled mobile robot model case study. (a) The robot trajectory converges to the goal set (green circular region) without reaching the unsafe region (red rectangular region) in spite of a constant error of $a = 1$. A baseline scheme that computes CBF and CLF constraints including the faulty sensor violates safety. (b) Lyapunov function of the proposed approach shows lower values than Lyapunov function of the baseline, and converges to zero, proving stability.

where $\mathbf{w}_t^{'}$ is the process noise. The following compensator is used to calculate the input $[\omega_t]_1$ and $[\omega_t]_2$ into (23)

$$[\omega_t]_1 = \int_{t^-}^{t^+} [u_t]_1 \cos\theta_t + [u_t]_2 \sin\theta_t \; dt \quad (25)$$

$$[\omega_t]_2 = ([u_t]_2 \cos\theta_t - [u_t]_1 \sin\theta_t)/[\omega_t]_1. \quad (26)$$

Here we assume that the observation for the orientation coordinate θ_t is attack-free and noise-free, which enables feedback linearization based on the variable θ_t.

In the linearized model, we use the observation equation

$$\begin{pmatrix} [y_t]_1 \\ [y_t]_2 \\ [y_t]_3 \\ [y_t]_4 \\ [y_t]_5 \\ [y_t]_6 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} [x_t]_1 \\ [x_t]_2 \\ [\dot{x}_t]_1 \\ [\dot{x}_t]_2 \end{pmatrix} + \mathbf{a}_t + \mathbf{v}_t \quad (27)$$

where $\mathbf{a}_t$ and $\mathbf{v}_t$ describe the impact of the attack and the measurement noise. Note that there is one redundant sensor each for the horizontal and vertical coordinates.

Here we let the safe region $\mathcal{C} = \{x_t : h(x_t) = [x_t]_2 + 0.2[\dot{x}_t]_2 + 0.05 \geq 0, t \geq 0\}$ and the goal region $\mathcal{G} = \{x_t : \omega(x_t) = d - ||x_t - x_g||_2 \geq 0\}$, where x_g is the center and $d = 0.05$ is the radius of the goal region. In order to reach the goal region, we choose the CLF

$$V(x) = (x_t - x_g)^T P_d (x_t - x_g) \quad (28)$$

where $P_d = 10^8 \begin{pmatrix} \frac{1}{d}I & 0 \\ 0 & I \end{pmatrix} P_L \begin{pmatrix} \frac{1}{d}I & 0 \\ 0 & I \end{pmatrix}$, P_L is the solution of the Lyapunov equation $A^T P_L + P_L A = -I$, and I is the identity matrix [30], [40]. We set $\rho = 1/(d\bar{\lambda}(P_d))$ in the linear constraints corresponding to CLF. The control input u_t is computed at each time step by solving (22) with $R = I$.

B. Numerical Study

A numerical study of the proposed algorithm was performed using Matlab. We set $m = 2$ and $\mathcal{F}(r_1) = 2$, $\mathcal{F}(r_2) = 4$, which correspond to the redundant horizontal and vertical observations. In order to test the safety and stability of the system, an attack given by $a_t = (0, 0, 0, 1, 0, 0)^T$ is injected into the redundant vertical observation $[y]_4$. This input caused the robot to appear to be farther from the safe region than it actually was, which could potentially cause the controller to violate the safety constraint. We test the algorithm under the attack with start point $x_0 = (1, 0, 0, 0)^T$ and goal region $\mathcal{G} = \{x_t : \omega(x_t) = 0.05 - ||x_t - (0, 0, 0, 0)^T||_2 \geq 0\}$. The noises $\mathbf{w}_t^{'}$ and $\mathbf{v}_t$ are Gaussian processes with means identically zero and covariances $10^{-3}I$. The values of γ_i and $\theta_{i,j}$ are 0.001 and 0.045.

The results are shown in Fig. 1. In Fig. 1(a), we plot the first two dimensions of the state, which describe the horizontal and vertical coordinates. Note that the robot stays in the safe region and eventually reaches the goal region, and hence satisfies safety and stability. For comparison, the baseline based on all sensors (including the faulty sensors) defined in [28] resulted in a safety violation. As shown in Fig. 1(b), the value of $V(x_t)$ for our proposed scheme is always smaller than the value of $V(x_t)$ for the baseline, and $V(x_t)$ converges to zero under our approach.

VII. Conclusions and Future Work

This paper proposed a new class of Control Barrier Functions (CBFs) for safety and stability of stochastic systems under sensor faults and attacks. Under our model, the set of possible fault patterns is known, but the specific fault pattern experienced by the system is unknown. Our approach was to compute a set of state estimators, each of which excluded a set of possibly faulted sensors in order to mitigate the impact of a particular fault pattern. We then constructed a CBF for each state estimator, which guaranteed safety provided that a linear constraint on the control input was satisfied at each time step. We proposed a scheme for using additional state estimators to resolve conflicts between these constraints and derived sufficient conditions for ensuring

safety with a desired probability for linear systems under different geometries of the safe region, including half-plane and ellipsoidal regions. We then showed how to compose our proposed CBFs with Control Lyapunov Functions (CLFs) to achieve joint safety and stability under faults and attacks. Our approach was validated using a numerical study of a wheeled mobile robot. Future work will include attacks that affect sensors and actuators, as well as analysis under arbitrary geometries and nonlinear dynamics.

REFERENCES

[1] S. Mitra, T. Wongpiromsarn, and R. M. Murray, "Verifying cyber-physical interactions in safety-critical systems," *IEEE Security & Privacy*, vol. 11, no. 4, pp. 28–37, 2013.

[2] A. Banerjee, K. K. Venkatasubramanian, T. Mukherjee, and S. K. S. Gupta, "Ensuring safety, security, and sustainability of mission-critical cyber–physical systems," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 283–299, 2011.

[3] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *Preprints of the 1st workshop on Secure Control Systems*, 2010, pp. 1–6.

[4] Y. Guan and X. Ge, "Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 48–59, 2017.

[5] S. Wang and F. Xiao, "AHU sensor fault diagnosis using principal component analysis method," *Energy and Buildings*, vol. 36, no. 2, pp. 147–160, 2004.

[6] Y. H. Chang, Q. Hu, and C. J. Tomlin, "Secure estimation based Kalman filter for cyber–physical systems against sensor attacks," *Automatica*, vol. 95, pp. 399–412, 2018.

[7] Y. Shoukry, M. Chong, M. Wakaiki, P. Nuzzo, A. Sangiovanni-Vincentelli, S. A. Seshia, J. P. Hespanha, and P. Tabuada, "SMT-based observer design for cyber-physical systems under sensor attacks," *ACM Transactions on Cyber-Physical Systems*, vol. 2, no. 1, p. 5, 2018.

[8] 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.

[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] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: Theory and applications," *arXiv preprint arXiv:1903.11199*, 2019.

[11] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, and J. Schröder, *Diagnosis and Fault-Tolerant Control*. Springer, 2006, vol. 2.

[12] Z. Chen, Y. Cao, S. X. Ding, K. Zhang, T. Koenings, T. Peng, C. Yang, and W. Gui, "A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2710–2720, 2019.

[13] L. Li, H. Luo, S. X. Ding, Y. Yang, and K. Peng, "Performance-based fault detection and fault-tolerant control for automatic control systems," *Automatica*, vol. 99, pp. 308–316, 2019.

[14] H. Yang, Y. Jiang, and S. Yin, "Fault-tolerant control of time-delay Markov jump systems with Ito stochastic process and output disturbance based on sliding mode observer," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5299–5307, 2018.

[15] D. Jung and E. Frisk, "Residual selection for fault detection and isolation using convex optimization," *Automatica*, vol. 97, pp. 143–149, 2018.

[16] H. Yang, C. Huang, B. Jiang, and M. M. Polycarpou, "Fault estimation and accommodation of interconnected systems: a separation principle," *IEEE Transactions on Cybernetics*, vol. 49, no. 12, pp. 4103–4116, 2018.

[17] Y. Chen, S. Kar, and J. M. Moura, "Resilient distributed estimation through adversary detection," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2455–2469, 2018.

[18] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on scada systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2013.

[19] 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.

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

[21] S. Z. Yong, M. Zhu, and E. Frazzoli, "Simultaneous input and state estimation for linear time-varying continuous-time stochastic systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2531–2538, 2016.

[22] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *2015 American Control Conference (ACC)*. IEEE, 2015, pp. 2439–2444.

[23] A. Girard, "Controller synthesis for safety and reachability via approximate bisimulation," *Automatica*, vol. 48, no. 5, pp. 947–953, 2012.

[24] S. Prajna, A. Jadbabaie, and G. J. Pappas, "A framework for worst-case and stochastic safety verification using barrier certificates," *IEEE Transactions on Automatic Control*, vol. 52, no. 8, pp. 1415–1428, 2007.

[25] H. A. Blom, J. Krystul, and G. Bakker, "A particle system for safety verification of free flight in air traffic," in *Proceedings of the 45th IEEE Conference on Decision and Control*. IEEE, 2006, pp. 1574–1579.

[26] G. Frehse, S. K. Jha, and B. H. Krogh, "A counterexample-guided approach to parameter synthesis for linear hybrid automata," in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2008, pp. 187–200.

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

[28] A. Clark, "Control barrier functions for complete and incomplete information stochastic systems," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 2928–2935.

[29] ——, "Control barrier functions for stochastic systems," *arXiv preprint arXiv:2003.03498*, 2020.

[30] Q. Nguyen and K. Sreenath, "Exponential control barrier functions for enforcing high relative-degree safety-critical constraints," in *2016 American Control Conference (ACC)*. IEEE, 2016, pp. 322–328.

[31] X. Xu, "Constrained control of input–output linearizable systems using control sharing barrier functions," *Automatica*, vol. 87, pp. 195–201, 2018.

[32] W. Xiao and C. Belta, "Control barrier functions for systems with high relative degree," *arXiv preprint arXiv:1903.04706*, 2019.

[33] S. Yaghoubi, G. Fainekos, and S. Sankaranarayanan, "Training neural network controllers using control barrier functions in the presence of disturbances," *arXiv preprint arXiv:2001.08088*, 2020.

[34] W. Xiao, C. Belta, and C. G. Cassandras, "Adaptive control barrier functions for safety-critical systems," *arXiv preprint arXiv:2002.04577*, 2020.

[35] R. Cheng, G. Orosz, R. M. Murray, and J. W. Burdick, "End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks," *arXiv preprint arXiv:1903.08792*, 2019.

[36] L. Wang, A. D. Ames, and M. Egerstedt, "Safety barrier certificates for collision-free multirobot systems," *IEEE Transactions on Robotics*, vol. 33, no. 3, pp. 661–674, 2017.

[37] X. Xu, J. W. Grizzle, P. Tabuada, and A. D. Ames, "Correctness guarantees for the composition of lane keeping and adaptive cruise control," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 3, pp. 1216–1229, 2017.

[38] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the continuous-time extended Kalman filter," *IEE Proceedings-Control Theory and Applications*, vol. 147, no. 1, pp. 45–52, 2000.

[39] Z. Chen, L. Li, and X. Huang, "Building an autonomous lane keeping simulator using real-world data and end-to-end learning," *IEEE Intelligent Transportation Systems Magazine*, 2018.

[40] A. D. Ames, K. Galloway, K. Sreenath, and J. W. Grizzle, "Rapidly exponentially stabilizing control Lyapunov functions and hybrid zero dynamics," *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 876–891, 2014.