Barrier Certificate based Safe Control for LiDAR-based Systems under Sensor Faults and Attacks

Hongchao Zhang, Shiyu Cheng, Luyao Niu, and Andrew Clark

***Abstract*— Autonomous Cyber-Physical Systems (CPS) fuse proprioceptive sensors such as GPS and exteroceptive sensors including Light Detection and Ranging (LiDAR) and cameras for state estimation and environmental observation. It has been shown that both types of sensors can be compromised by malicious attacks, leading to unacceptable safety violations. We study the problem of safety-critical control of a LiDAR-based system under sensor faults and attacks. We propose a framework consisting of fault tolerant estimation and fault tolerant control. The former reconstructs a LiDAR scan with state estimations, and excludes the possible faulty estimations that are not aligned with LiDAR measurements. We also verify the correctness of LiDAR scans by comparing them with the reconstructed ones and removing the possibly compromised sector in the scan. Fault tolerant control computes a control signal with the remaining estimations at each time step. We prove that the synthesized control input guarantees system safety using control barrier certificates. We validate our proposed framework using a UAV delivery system in an urban environment. We show that our proposed approach guarantees safety for the UAV whereas a baseline fails.**

I. Introduction

Autonomous Cyber-Physical Systems (CPS) are expected to satisfy safety property in different applications [1]. Safety violations can lead to severe economic loss and catastrophic damage to systems as well as human operators [1]. When the system can perfectly observe its state and the surrounding environment, safe control methodologies have been proposed including control barrier function (CBF) [2], Hamilton-Jacobi-Bellman-Isaacs (HJI) equation [3], and finite-state abstraction [4] -based approaches.

In real-world applications, system states and the environment are measured by sensors. As the environment becomes increasingly complex, modern CPS utilizes exteroceptive sensors including Light Detection and Ranging (LiDAR) and cameras to obtain richer perception of the operating space [5]. Fusion among the exteroceptive sensors and proprioceptive sensors such as GPS and odometer allows CPS to better understand the environment [6] and ensure safe operation.

Sensors have been shown to be vulnerable to faults and malicious attacks, under which ensuring CPS safety becomes more challenging. The navigation sensors can be spoofed by an adversary to cause crashes of autonomous vehicles [7], [8]. Reflections [9] and malicious attacks [10], [11] targeting LiDAR sensors can create a compromised description of the environment. These false sensor measurements bias the CPS state estimation and observations over the environment, rendering CPS to make erroneous control decisions and incur safety violations.

Modeling and detecting of sensor faults and attacks have been extensively studied [12], [13]. Secure system state estimation using measurements from proprioceptive sensors has been investigated in [14], [15]. Closed-loop safety-critical control under sensor faults and attacks has been recently studied in [16], [17]. However, these approaches are applicable to CPS using only proprioceptive sensors. When exteroceptive sensors such as LiDAR are adopted by CPS, the impact of attacks on the output of the nonlinear filters used to process LiDAR measurements are not incorporated into the aforementioned safety-critical control designs [16], [17], rendering them less effective.

In this paper, we study the problem of safety-critical control for a LiDAR-based system in the presence of sensor faults and attacks. We propose a fault tolerant safe control framework consisting of two components, namely fault tolerant estimation and fault tolerant control. Our proposed framework leverages the fact that only a narrow sector (normally within 8°) of LiDAR scans can be compromised by an adversary. Using these insights, we select system state estimations and sectors in LiDAR scans simultaneously so that they are aligned, while removing untrusted state estimates. We then use the selected state estimations and LiDAR measurements to compute a control input with safety guarantees. We make the following specific contributions:

- We propose a fault tolerant state estimation algorithm that is resilient to attacks against proprioceptive sensors and LiDAR measurements. Our approach reconstructs a simulated scan based on a state estimate and an precomputed map of the environment. We leverage this reconstruction to remove false sensor inputs as well as detect and remove spoofed LiDAR measurements.
- We propose a fault tolerant safe control design using control barrier certificates. We present a sum-of-squares program to compute a control barrier certificate, which verifies a given safety constraint in the presence of estimation errors due to noise and attacks. We prove bounds on the probability that our synthesized control input guarantees safety.
- We validate our proposed framework using a UAV delivery system equipped with multiple sensors including a LiDAR. We show that the UAV successfully avoids the obstacles when navigating in an urban environment using our synthesized control law, while crashes into the unsafe region using a baseline.

Hongchao Zhang, Shiyu Cheng, Luyao Niu, and Andrew Clark are with the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA 01609 USA. {hzhang9,scheng3,lniu,aclark}@wpi.edu

The remainder of this paper is organized as follows. Section II presents the related work. Section III presents the system model, threat model, and necessary background. Section IV presents our proposed fault tolerant safe control framework along with its safety guarantee. Section V gives a numerical case study on a UAV delivery system. Section VI concludes the paper.

II. Related Work

Ensuring CPS safety has attracted extensive research attention. Typical approaches include finite-state abstraction [4], HJI equation [3], and counterexample-guided synthesis [18]. Barrier function-based approaches, which formulate the safety constraint as a linear inequality over the control input, have been proposed to guarantee safety for CPS [2], [19]–[21]. These approaches are applicable to CPS estimating system state using proprioceptive sensors.

Safety-critical control for systems using exteroceptive sensors such as cameras and LiDAR have been recently investigated in [22]–[25]. CBFs designed for high-dimensional exteroceptive sensor measurements including measurement-robust CBF [23], observation-based neural CBF [24], and differentiable CBFs for learning systems [25] have been proposed to compute controllers with safety guarantees.

False data injection (FDI) attacks have been reported in different applications, including modern power systems [13] and unmanned aerial vehicle (UAV) [26]. To this end, modeling, mitigating, and detecting FDI [12]–[15] have been studied. LiDAR sensors have been demonstrated to be vulnerable to spoofing attacks in [11], [27]. The authors of [10] designed attacks that are capable of injecting false points at different locations in the point cloud. In [28], a stealthy attack against a perception-based controller equipped with an anomaly detector were proposed.

The existing literature on safe control in the presence of FDI attacks mainly focuses on systems with proprioceptive sensors. In [16], a barrier certificate based approach is proposed to ensure safety and reachability under FDI attack. A fault tolerant CBF is introduced in [17] to ensure joint safety and reachability under attacks targeting proprioceptive sensors. In [29], the authors have demonstrated that camera and LiDAR fusion is secure against naive attacks. For systems under attacks targeting both proprioceptive and exteroceptive sensors, how to synthesize a safety-critical control has been less studied.

III. Problem Formulation

In this section, we introduce the system and threat model. We then formulate the problem and give needed background.

A. System Dynamics and Observation Model

Consider a discrete-time control-affine system given as:

$$x[k+1] = f(x[k]) + g(x[k])u[k] + w[k] \tag{1}$$

where $w[k]$ is a Gaussian process with mean zero and autocorrelation function $R_w(k,k') = Q_k\delta(k-k')$ with δ denoting the discrete-time delta function. We assume that there is a nominal controller $u = \pi(x)$, for some function $\pi : \mathcal{X} \to \mathbb{R}^m$. We let $x[k] \in \mathcal{X} \subseteq \mathbb{R}^n$ denote the system state and $u[k] \in \mathbb{R}^m$ denote a control signal at time k. Functions $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are assumed to be Lipschitz continuous.

System (1) uses a set of sensors $I_p := \{1, \dots, n_p\}$ to measure its states with observation $y[k] \in \mathbb{R}^z$ following the dynamics described as:

$$y[k] = o(x[k]) + v[k], \tag{2}$$

where $o : \mathbb{R}^n \to \mathbb{R}^z$ is the observation function and $v[k]$ is an independent Gaussian process with mean identically zero and autocorrelation function $R_v[k,k'] = R_k\delta(k-k')$ and R_k is a positive definite matrix.

The system is equipped with a LiDAR sensor that observes the environment by calculating the ranges and angles to objects. A LiDAR sensor fires and collects n_s laser beams to construct a scan $S := \{(s_i^r, s_i^a),\ 0 \le i \le n_s\}$, where s_i^r denotes the range of the i-th scan, and s_i^a denotes the angle of the i-th scan. We denote the Cartesian translated LiDAR scan S measured at pose x as $\mathcal{O}(x, S)$

We assume a 2D point-cloud map $\mathcal{M}$ is known by the system as prior knowledge. The map $\mathcal{M} := \{(m_i^x, m_i^y),\ 0 \le i \le n_{\mathcal{M}}\}$ is a collection of $n_{\mathcal{M}}$ points with tuples of object positions (m_i^x, m_i^y) in the world coordinate.

B. Threat Model

We assume that there exists an adversary that aims to cause collisions or other unsafe behaviors. The adversary has the capability to utilize any state-of-the-art spoofer for different sensors to conduct false data injection to perturb the observations. The injected false data can bias the system state estimation and cause the system to make incorrect control decisions. We denote the perturbed observations as

$$\bar{y}[k] = o(x[k]) + v[k] + a[k]. \tag{3}$$

The adversary can also compromise the LiDAR sensor by creating a near obstacle as demonstrated in [10]. The adversary fires laser beams to inject several artificial points into a LiDAR scan. We denote the compromised LiDAR scan as $S \oplus e'$, where $\oplus$ is a merge function introduced by [10]. However, due to the physical limitation of spoofer hardware, the injected point can only be within a very narrow spoofing angle, i.e. 8° horizontal angle.

We index the LiDAR sensor as the 0-th sensor and define $I = \{0\} \bigcup I_p$. We denote the set of sensors attacked by the adversary as $\mathcal{A} \subseteq I$. We assume that the system is uniformly observable from the sensors in $I \backslash \mathcal{A}$. We assume that, at each time k, the support of $a[k]$ is contained in $\mathcal{A}$.

C. Safety and Problem Formulation

We define the state space $\mathcal{X}$ and a safety set $\mathcal{C}$ as

$$\mathcal{X} = \{x : h(x) \ge 0\}, \quad \mathcal{C} = \{x \in \mathcal{X} : h_0(x) \ge 0\},$$

where $h, h_0 : \mathcal{X} \mapsto \mathbb{R}$. We say system (1) is safe with respect to $\mathcal{C}$ if $x[k] \in \mathcal{C}$ for all time $k = 0, 1, \dots$. We assume that

the safe region $\mathcal{C}$ is pre-defined and known by the system, and the initial state of the system is safe, i.e. $x_0 \in \mathcal{C}$.

Problem 1. *Given a map $\mathcal{M}$ and a safety set $\mathcal{C}$, we consider a nonlinear LiDAR-based system with dynamics* (1) *that is controlled by a nominal controller. The problem studied is to find a scheme to ensure system safety with desired probability $(1-\epsilon)$, where $\epsilon \in (0,1)$, when an adversary is present.*

D. Preliminaries

In what follows, we give background on discrete-time Extended Kalman Filter (EKF) and estimating pose from LiDAR scans

1) DT-EKF: For the system with dynamics (1) and observation (2), the state estimate $\hat{x}$ is computed via EKF as:

$$\hat{x}[k+1] = F(\hat{x}[k], u[k]) + K_k(y[k] - o(\hat{x}[k])), \quad (4)$$

where $F(x[k], u[k]) = f(x[k]) + g(x[k])u[k]$. The Kalman filter gain is

$$K_k = A_k P_k C_k^T (C_k P_k C_k^T + R_k)^{-1}, \quad (5)$$

where $A_k = \frac{\partial F}{\partial x}(\hat{x}[k], u[k])$, $C_k = \frac{\partial o}{\partial x}(\hat{x}[k])$, and P_k is defined by the Riccati difference equation:

$$P_{k+1} = A_k P_k A_k^T + Q_k - K_k(C_k P_k C_k^T + R_k)K_k^T.$$

The error bound of discrete-time EKF can be derived by Theorem 3.2 in [30] if Assumption 1 holds.

Assumption 1. *The system described by* (1) *and* (2) *satisfies the conditions:*

- *A_k is nonsingular for every $k \geq 0$.*
- *There are positive real numbers $\bar{a}, \bar{c}, \underline{p}, \bar{p} > 0$ such that the following bounds on various matrices are fulfilled for every $k \geq 0$:*

$$\|A_k\| \leq \bar{a};\ \|C_k\| \leq \bar{c};\ \underline{p}I \leq P_k \leq \bar{p}I;$$
$$\underline{q}I \leq Q_k;\ \underline{r}I \leq R_k.$$

- *Let ϕ and χ be defined as*

$$F(x[k], u[k]) - F(\hat{x}[k], u[k]) = A_k(x[k] - \hat{x}[k]) + \varphi(x[k], \hat{x}[k], u[k])$$
$$o(x[k]) - o(\hat{x}[k]) = C_k(x[k] - \hat{x}[k]) + \chi(x[k], \hat{x}[k])$$

Then there are positive real numbers $\epsilon_\varphi, \epsilon_\chi, \kappa_\varphi, \kappa_\chi > 0$ such that the nonlinear functions φ, χ are bounded via

$$\|\varphi(x, \hat{x}, u)\| \leq \kappa_\varphi \|x - \hat{x}\|^2, \quad \|\chi(x, \hat{x})\| \leq \kappa_\chi \|x - \hat{x}\|^2$$

for $x, \hat{x} \in R^n$ with $\|x - \hat{x}\| \leq \epsilon_\varphi$ and $\|x - \hat{x}\| \leq \epsilon_\chi$, respectively.

If the conditions of Assumption 1 hold, the estimation error $\zeta_k = x[k] - \hat{x}[k]$ is exponentially bounded in mean square and bounded with probability one, provided that the initial estimation error satisfies $\|\zeta_0\| \leq \bar{\zeta}$ [30].

2) Estimating Pose By Comparing Scans: Pose refers to the position of the system in a Cartesian coordinate frame. Pose estimations with LiDAR scans have been extensively studied. NDT [31], as one of the widely-used approaches, models the distribution of all reconstructed 2D-Points of one laser scan by a collection of local normal distributions.

Consider two states $x_1, x_2 \in \mathcal{X}$ and the LiDAR scans $\mathcal{O}(x_1, S_1)$ and $\mathcal{O}(x_2, S_2)$ collected at x_1 and x_2, respectively. The NDT method estimates the relative pose change as $r = \mathcal{O}(x_1, S_1) \ominus \mathcal{O}(x_2, S_2)$, where $\ominus$ is a scan match operation. The scan match operation is implemented as follows. The NDT method first subdivides the surrounding space uniformly into cells with constant size. For each cell in $\mathcal{O}(x_1, S_1)$, the mean q and the covariance matrix Σ are computed to model the points contained in the cell as the normal distribution $N(q, \Sigma)$. Denote the points in $\mathcal{O}(x_2, S_2)$ as $p_i,\ i \in n_s$, where p_i is a position vector and n_s is the number of valid points. Define loss function $\mathcal{L}_s(r)$ as

$$\mathcal{L}_s(r) = \sum_i \exp\left(\frac{-((p_i - r) - q_i)^T \Sigma_i^{-1}((p_i - r) - q_i)}{2}\right) \quad (6)$$

Estimating r is to solve the minimization problem:

$$\min_r -\mathcal{L}_s(r) \quad (7)$$

with Newton's algorithm. We use r to denote the solution to (7) for the rest of the paper. The corresponding loss $\mathcal{L}_s(r)$ can be computed with the output of scan match r by (6).

IV. Fault Tolerant Safe Control Framework

In this section, we propose a framework for safe control that is compatible with existing LiDAR-based autonomous systems. We first give an overview and then describe each component in detail.

A. Overview of Framework

We consider a system with dynamics (1) and observation model (2) in the presence of an adversary, as described in Section III. To guarantee the system safety under attacks, we propose a fault tolerant framework to ensure safety at each time step. The framework consists of two parts, namely *fault tolerant estimation* and *fault tolerant control*.

The idea of fault tolerant estimation is to exclude compromised sensors in I_p by utilizing additional information contained in LiDAR sensor measurements. We maintain a set of state estimations $\hat{x}_i$ using EKF, where $i \in I_l \subseteq 2^{I_p}$ and each element of $i \in I_l$ is a collection of sensors in I_p such that system (1) is uniformly observable from the sensors in I_l. As shown in Fig. 1, a fault tolerant estimation reconstructs a LiDAR observation, denoted as $\mathcal{O}(\hat{x}_i, \mathcal{M})$, for each state estimation $\hat{x}_i$. The reconstruction is achieved by simulating the scan process on knowledge map $\mathcal{M}$ with state estimate $\hat{x}$ being the center. We propose a fault tolerant LiDAR estimation to compare the estimated LiDAR scan $\mathcal{O}(\hat{x}_i, \mathcal{M})$ with the actual LiDAR measurement $\mathcal{O}(x, S)$. The comparison then provides a pose estimation. Using the pose estimation, our proposed fault tolerant state

estimation excludes the conflicting state estimations, i.e., the state estimations that deviate from the LiDAR estimation. We will detail fault tolerant estimation in Section IV-B and IV-C.

After excluding the conflicting state estimations using fault tolerant estimation, we then design fault tolerant safe control to ensure safety of the system at each time step. Fault tolerant safe control computes an input u_o that does not deviate too far from the nominal controller $\pi(\hat{x}_i)$ for all i given by the fault tolerant estimation. The safety of u_o is certified by a discrete-time barrier certificate. We will present the details of fault tolerant safe control in Section IV-D

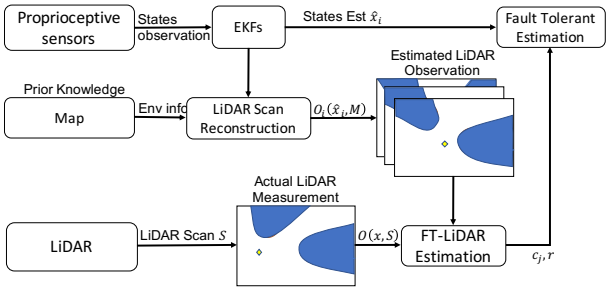


Fig. 1: Fault tolerant estimation for LiDAR-based system removes conflicting state estimations by comparing estimations of proprioceptive sensors with additional information from exteroceptive sensors measurements.

In what follows, we describe the fault tolerant estimation in two-fold, that is fault tolerant LiDAR estimation (Section IV-B) and fault tolerant state estimation (Section IV-C).

Algorithm 1 LiDAR Scan Reconstruction

1: **Input:** State estimate $\hat{x}_i$, point-cloud map $\mathcal{M}$
2: **Parameters:** Resolution of the LiDAR scan c_r, maximum LiDAR range r_{max}.
3: **Output:** Estimated LiDAR Observation $\mathcal{O}(\hat{x}_i, \mathcal{M})$
4: **Init:** Set $\hat{x}_i$ as the center of scan $S_{\mathcal{M}}$, set $l_k^r \leftarrow r_{max}$. Separate the scan equally into $\frac{2\pi}{c_r}$ sectors S_k with corresponding angle l_k^a.
5: Translate points $m_j \in \mathcal{M}$ into polar coordinate with the origin $\hat{x}_i$, and represent it with a tuple (m_j^r, m_i^a).
6: **for** $m_j \in \mathcal{M}$ and $k \in [0, c_r]$ **do**
7: **if** $m_j \in S_k$ and $m_j^r \leq l_k^r$ **then**
8: $l_k^r \leftarrow p_i^r$
9: **end if**
10: **end for**
11: **for** k s.t. $l_k^r = r_{max}$ **do**
12: $l_k^r \leftarrow NaN$
13: **end for**
14: Reconstruct $S_{\mathcal{M}} = \{(l_k^r, l_k^a)\}$
15: **Return** $\mathcal{O}(\hat{x}_i, \mathcal{M}) = \mathcal{O}(\hat{x}_i, S_{\mathcal{M}})$

B. Fault Tolerant LiDAR Estimations

In the following, we introduce fault tolerant LiDAR estimation. This procedure converts each state estimation $\hat{x}_i$ given by EKFs to an estimated LiDAR observation $\mathcal{O}(\hat{x}_i, \mathcal{M})$ using map $\mathcal{M}$. The estimated LiDAR observation is then compared with the actual LiDAR observation to exclude possible faults in state estimations.

Fault tolerant LiDAR estimation is presented in Alg. 1. Given parameters on the resolution of the LiDAR scan c_r and maximum LiDAR range r_{max}, we initialize the estimated scan $S_{\mathcal{M}} = \{(l_k^r, l_k^a)\}$ with a circle centered at $\hat{x}_i$ and radius as r_{max}. We equally divide the circle and assign sectors S_k to the corresponding l_k^a. Next, we represent the points in map $\mathcal{M}$ using polar coordinates with the origin at $\hat{x}_i$. To simulate the scan, we assign the closest point to the scan from line 6 to 10. We iterate through all points $m_j \in \mathcal{M}$. For the point in sector S_k, we replace l_k^r with m_j^r if $m_j^r \leq l_k^r$. Then we remove the points that have never been updated. Finally, we output estimated observation $\mathcal{O}(\hat{x}_i, \mathcal{M}) = \mathcal{O}(\hat{x}_i, S_{\mathcal{M}})$. Intuitively, this estimated observation can be viewed as the output of a LiDAR scan centered at state $\hat{x}_i$ with object locations given in the map $\mathcal{M}$. Hence, any deviation of the estimated and actual scans indicates either an error in the state estimate or a spoofing attack on the scan.

Next, we consider the case where the adversary not only injects false data into prorioceptive sensors but also spoofs LiDAR sensors. The intuitive countermeasure is to remove the region of the scan that is impacted by false data. Since the adversary is only capable of modifying points in the scan within a narrow spoofing angle, our approach is to partition the scan and map into regions c_j and attempt to identify which region has been impacted by spoofing. That region is then removed from the scan and the estimated scan. Since the adversary tries to bias the state estimation, we model the problem of choosing a set of observations to ignore in order to mitigate the impact of false data as a minimax optimization

$$\min_{e_I'} \max_{c_j} \mathcal{L}_s(\tilde{r}) \tag{8}$$

where $\tilde{r} = \mathcal{O}(\hat{x}_i, \mathcal{M} \backslash c_j) \ominus \mathcal{O}(x, S \oplus e_I' \backslash c_j)$. We search for subdivision c_j through the LiDAR observation space with Alg. 2, which is detailed as follows.

Algorithm 2 FT-LiDAR Estimation

1: **Input:** State estimation $\hat{x}$, number of sector n_j, Map $\mathcal{M}$ and LiDAR scan S
2: **Output:** r_j, c_j
3: **Init:** Equally separate scan S into n_j sectors $c_j \in S$
4: **for** $c_j \in S$ **do**
5: Scan Reconstruction $\mathcal{O}_j(\hat{x}, \mathcal{M} \backslash c_j)$
6: Scan Reconstruction $\mathcal{O}_j(x, S \backslash c_j)$
7: Compute n_s^j the number of points in $S \backslash c_j$.
8: Compute $\tilde{r}_j = \mathcal{O}_j(\hat{x}, \mathcal{M} \backslash c_j) \ominus \mathcal{O}_j(x, S \backslash c_j)$
9: Compute $\zeta_s^j = n_s^j - \mathcal{L}_s(r_j)$
10: **if** $\zeta_s^j \leq \bar{\zeta}_s$ **then return** $\tilde{r}_j, c_j$
11: **end if**
12: **end for**

The adversary compromises the LiDAR scan S by merging it with false data e_I', denoted as $S \oplus e_I'$. As shown in Alg. 2,

we take in state estimation $\hat{x}_i$, number of sectors n_j, map $\mathcal{M}$, and scan S to search for sector c_j over scan S. The algorithm outputs the corresponding estimated relative pose $\tilde{r}_j$. For each sector c_j, we estimate observations $\mathcal{O}(\hat{x}_i, \mathcal{M}\backslash c_j)$ with Alg. 1 and reconstruct the corresponding LiDAR observation $\mathcal{O}(x, S\oplus e'_I\backslash c_j)$. Next, we compute n_s^j, the number of points contained in $S\backslash c_j$, and perform scan match to obtain $\tilde{r}$ by

$$\tilde{r}_j = \mathcal{O}(\hat{x}_i, \mathcal{M}\backslash c_j) \ominus \mathcal{O}(x, S \oplus e'_I\backslash c_j). \tag{9}$$

Then, we compute the loss function $\mathcal{L}_s(\tilde{r})$ and the performance degradation $\zeta_s^j = n_s^j - \mathcal{L}_s(\tilde{r})$. Finally, we output $\tilde{r}_j$ and c_j for $\zeta_s^j \leq \bar{\zeta}_s$. In what follows, we compute the upper bound $\bar{\zeta}_s$ of the degradation of the loss $\mathcal{L}_s$ brought by noise as the criteria of whether LiDAR sensor is affected by factors other than noise.

We consider a point p_i sampled in the LiDAR scan collected at state x with a zero-mean disturbance w_i whose norm is bounded as $\|w_i\| \leq \bar{w}_i$.

Theorem 1. *Consider a state x and its state estimation $\hat{x}$. Let $\mathcal{O}(x, S)$ and $\mathcal{O}(\hat{x}, \mathcal{M})$ be LiDAR scan and estimated LiDAR observation. Let $r = \mathcal{O}(\hat{x}_i, \mathcal{M}) \ominus \mathcal{O}(x, S)$ and $\tilde{r}$ be computed by* (9) *when adversary present. In the case where the LiDAR sensor is not attacked, we have the performance degradation ζ_s is bounded by*

$$\begin{aligned}\zeta_s &:= \mathcal{L}_s^{max}(r) - \mathcal{L}_s(r) \\ &\leq n_s - \sum_i \exp\left(\frac{-\bar{w}_i^2 \lambda(\Sigma_i^{-1})}{2}\right) =: \bar{\zeta}_s, \end{aligned} \tag{10}$$

where $\mathcal{L}_s^{max}(r)$ is the maximum of (6)*, n_s is the number of points contained in S, and $\lambda(\Sigma_i^{-1})$ is the maximum eigenvalue of Σ_i^{-1}.*

When the LiDAR sensor is attacked, if a subdivision $c_j \supseteq e'_I$ can be found by Alg. 2, we have the performance degradation of scan match is bounded as (10)*, where n_s is the number of points contained in $S\backslash c_j$ and the summation is over all points in $S \setminus c_j$.*

Proof. We first show that $L_s^{max}(r) = n_s$. Then we derive a lower bound for $\mathcal{L}_s(r)$. Since covariance Σ_i is positive definite, using (6) we have

$$\begin{aligned} &L_s^{max}(r) \\ &= \sum_i \exp\left(\frac{-((p_i - r) - q_i)^T \Sigma_i^{-1} ((p_i - r) - q_i)}{2}\right) \\ &\leq \sum_i \exp(0) = n_s. \end{aligned}$$

Let p_i be a point sampled in LiDAR scan. We have that $((p_i - r) - q_i) \leq = w_i$ with w_i being the realized disturbance when sampling p_i. Since $\|w_i\| \leq \bar{w}_i$ and Σ_i^{-1} is Hermitian, we then have

$$\begin{aligned} &\sum_i \exp\left(\frac{-((p_i - r) - q_i)^T \Sigma_i^{-1} ((p_i - r) - q_i)}{2}\right) \\ &\geq \sum_i \exp\left(\frac{-\bar{w}_i^2 \lambda(\Sigma_i^{-1})}{2}\right). \end{aligned}$$

Hence, we have that ζ_s is bounded as (10).

When the LiDAR sensor is spoofed, there always exists a subdivision c_j such that the false data e'_I satisfies $e'_I \subseteq c_j$. If c_j is successfully identified by Alg. 2, then the subdivision c_j along with the false data e'_I are ignored. In this case, our analysis for the scenario where the LiDAR sensor is not attacked can be applied, yielding the bound in (10) with n_s being the number of points contained in $S \setminus c_j$. If c_j containing e'_I is not identified and is not ignored, then by line 10 of Alg. 2, we have that $\zeta_s^j \leq \bar{\zeta}_s$ and thus the bound in (10) follows. □

C. Fault Tolerant State Estimation

We next propose the criteria to develop an algorithm for a fault tolerant state estimation that provides bounded estimation error under false data attacks on the proprioceptive sensors. Our approach computes a set of indices $I_a \subseteq I_l$ that are removed to ensure that the state estimation error is bounded. A state estimate is *not removed* (i.e. $i \notin I_a$) if either of the following criteria holds.

- Case I: $i \notin I_a$ for estimation indexed $i \in I_l$, if $\|r_i\| \leq \theta_h$ and $\zeta_s^i \leq \bar{\zeta}_s$.
- Case II: $i \notin I_a$ for estimation indexed $i \in I_l$, if $\|\tilde{r}_i\| \leq \theta_h$ and $\tilde{\zeta}_s^i \leq \bar{\zeta}^s$.

We consider LiDAR observation is trusted, if for all $i \in I_l$ estimated LiDAR observation, the scan match degradation $\zeta_s^i \leq \bar{\zeta}_s$. In Case I, we have the scan match degradation $\zeta_s^i \leq \bar{\zeta}_s$, and the pose deviation $\|r_i\| \leq \theta_h$. We draw the conclusion that $\hat{x}_i$ agrees with the LiDAR observation, and hence $i \in I\backslash I_a$. When the LiDAR observation is not trusted, we reconstruct estimated and actual LiDAR observation with Alg. 2 to exclude section c_j. In Case II, we have the reconstructed scan match degradation $\|\tilde{r}_i\|_ \leq \theta_h$, and the pose deviation within tolerance with $\tilde{\zeta}_s^i \leq \bar{\zeta}^s$. We draw the conclusion that $i \in I\backslash I_a$.

In what follows, we show that sensor $i \in I\backslash I_a$ selected by criteria is attack-free and we can further have the deviation of FT-Estimation bounded by the EKF error bound of selected sensors.

Theorem 2. *Given scan match results r_i, $\tilde{r}_i$ and $\bar{\zeta}_s$, for sensor $i \in I\backslash I_a$ given by criteria I and II, we have estimation error bounded as $\|x - \hat{x}_i\| \leq \bar{\zeta}_i$.*

Proof. We prove by contradiction. We suppose that there exists a sensor $b \in I\backslash I_a$, whose estimation $\hat{x}_b$ satisfies $\|x - \hat{x}_b\| > \bar{\zeta}_b$. We next show contradictions for Case I and II.

In Case I, set $\theta_h = \min_i \bar{\zeta}_i$. Since $\zeta_s^b \leq \bar{\zeta}_s$, we have that LiDAR scan matches with estimated scan with relative pose change $r = x - \hat{x}_b$. If sensor b is included in $I\backslash I_a$, we have $\|x - \hat{x}_b\| \leq \theta_h \leq \zeta_b$, which contradicts to $\|x - \hat{x}_b\| > \bar{\zeta}_b$.

In Case II, set $\theta_h = \min_i \bar{\zeta}_i$. Since $\tilde{\zeta}_s^b \leq \bar{\zeta}_s$, we have that LiDAR scan matches with estimated scan with relative pose change $\tilde{r} = x - \hat{x}_b$. If sensor b is included in $I\backslash I_a$, we have $\|x - \hat{x}_b\| \leq \theta_h \leq \zeta_b$, which contradicts to $\|x - \hat{x}_b\| > \bar{\zeta}_b$.

Otherwise, sensor b will be excluded into set I_a and hence for any sensor $i \in I\backslash I_a$ we have the error bounded. □

D. Fault-Tolerant Safe Control

We next present the fault tolerant control synthesis to ensure safety of the system. We set the state estimation as $\hat{x}_\alpha[k] = \hat{x}_i$, for some $i \in I \backslash I_a$. We define the control input signal as $u_o[k] = \pi(\hat{x}_\alpha[k]) + \hat{u}[k]$. In what follows, we assume the nominal controller is of the form $\pi(x) = \pi_0 + K_c\hat{x}_\alpha$ for some $\pi_0 \in \mathbb{R}^m$ and matrix K_c. Since we have $||x[k]-\hat{x}_\alpha[k]|| \leq \bar{\zeta}_\alpha$ by Theorem 2, the nominal control input for the estimated state satisfies

$$||\pi(\hat{x}_\alpha[k]) - \pi(x[k])|| = ||K_c(x[k] - \hat{x}_\alpha[k])|| \leq ||K_c||\bar{\zeta}_\alpha.$$

Hence, if we choose $u_o[k]$ such that $||\hat{u}[k]||_2 \leq \xi - ||K_c||\bar{\zeta}_\alpha$ for some $\xi \geq 0$, then we can guarantee that the chosen control input is within a bounded distance of the nominal control input corresponding to the true state value.

Proposition 1. *Consider a discrete-time system described by* (1) *and sets* $\mathcal{C}, \mathcal{D} \subseteq \mathcal{X}$. *If there exist a function* $B : \mathcal{X} \to \mathbb{R}_0^+$, *a constant* $c \geq 0$, *a linear controller* $u = K_c x$, *and a constant* $\gamma \in [0, 1)$ *such that*

$$B(x) \leq \gamma, \qquad \forall x \in \mathcal{C} \tag{11}$$

$$B(x) \geq 1, \qquad \forall x \in \mathcal{D} \tag{12}$$

$$\mathbb{E}[B(f(x) + g(x)(K_c x + \hat{u}) + w) \mid x] \leq B(x) + c, \quad \forall x \in \mathcal{X}, \forall \|\hat{u}\| \leq \xi \tag{13}$$

then for any initial state $x_0 \in \mathcal{C}$, *we have the* $Pr(x[k] \in C, 0 \leq k \leq T_d) \geq 1 - \gamma - cT_d$.

Proof. We have $u_o - u = K_c x - K_c\hat{x}_\alpha + \hat{u}$. Since $\|\hat{u}\| \leq \xi - K_c\bar{\zeta}_\alpha$ and $\|K_c x - K_c\hat{x}_\alpha\| \leq K_c\bar{\zeta}_\alpha$. By triangle inequality, we can have $\|u_o - u\| \leq \xi$. Since there exists a function $B(x)$ satisfying (11) to (13), $B(x)$ is a control barrier certificate for system (1). According to [32] and (12), we have

$$\begin{aligned} &Pr\{x[k] \in \mathcal{D} \text{ for some } \ 0 \leq k < T_d \mid x(0) = x_0\} \\ &\quad \leq P\left\{\sup_{0 \leq k < T_d} B(x[k]) \geq 1 \mid x(0) = x_0\right\} \\ &\quad \leq B(x_0) + cT_d \leq \gamma + cT_d. \end{aligned}$$

□

The system has continuous state space $\mathcal{X}$ and action space U, we can follow the standard procedure to compute control barrier certificate $B(x)$ by solving an SOS programming. We define $h_B^\xi(\hat{u}) = (\xi - K_c\bar{\zeta}_\alpha)^2 - \|\hat{u}\|_2^2$. The SOS programming is given as follows:

Proposition 2. *Suppose there exist a function* $B(x)$ *and polynomials* $\lambda_0(x)$, $\lambda_1(x)$, $\lambda_x(x, \hat{u})$ *and* $\lambda_{\hat{u}}(x, \hat{u})$ *such that*

$$-B(x) - \lambda_0(x)h_0(x) + \gamma \text{ is SOS} \tag{14}$$

$$B(x) + \lambda_1(x)h_0(x) - 1 \text{ is SOS} \tag{15}$$

$$-\mathbb{E}[B(f(x) + g(x)(K_c x + \hat{u}) + w) \mid x] + B(x) - \lambda(x)h(x) - \lambda_{\hat{u}}(x, \hat{u})h_B^\xi(\hat{u}) + c \text{ is SOS} \tag{16}$$

then for any initial state $x_0 \in \mathcal{C}$, *we have the* $Pr(x[k] \in C, 0 \leq k \leq T_d) \geq 1 - \gamma - cT_d$.

Proof. Since the entries $B(x)$ and $\lambda_0(x)$ in $-B(x) - \lambda_0(x)h_0(x) + \gamma$ are SOS, we have $0 \leq B(x) + \lambda_0(x)h_0(x) \leq \gamma$. Since the term $\lambda_0(x)h_0(x)$ is nonnegative over $\mathcal{C}$, (14) and (15) implies (11) and (12) in Proposition 1. Since the terms $\lambda_{\hat{u}}(x)h_B^\xi(\hat{u})$ and $\lambda(x)h(x)$ are nonnegative over set $\mathcal{X}$, we have (13) holds, which implies that the function $B(x)$ is a control barrier certificate. □

The choice of ξ uses a similar approach as [16].

We propose Alg. 3 to compute feasible control inputs to ensure safety at each time-step k. We initialize $I_a \leftarrow \emptyset$ and define $\Omega_{i \in I \backslash I_a} := \{u_o : (u_o - u_i)^T(u_o - u_i) \leq \xi\}$. At each time-step k we maintain n_l state estimations for sensors in I_l and compute control input $u_i := \pi(\hat{x}_i)$ with nominal controller. We compute u_o by solving (17), where J is some cost function. If no such u_o exists, then we perform Alg. 2 and fault tolerant state estimation to exclude conflicting sensors into I_a.

Algorithm 3 Fault Tolerant Control

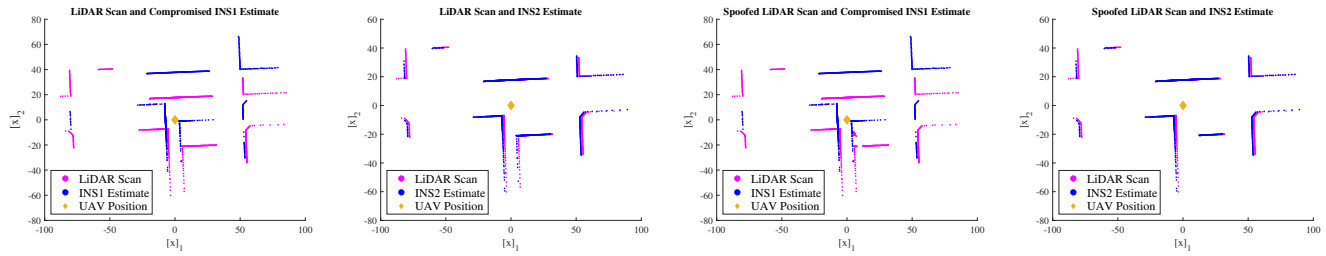
Init: $I_a \leftarrow \emptyset$ and $\Omega_{i \in I \backslash I_a} := \{u_o : (u_o - u_i)^T(u_o - u_i) \leq \xi\}$
Maintain n_l EKFs for each sensor to estimate state $\hat{x}_i,\ i \in I_l = \{1, 2, \ldots, n_l\}$.
Compute control input $u_i := \pi(\hat{x}_i)$.
if control input $u \in \bigcap_{i \in I \backslash I_a} \Omega_i$ **then**
    set $\hat{u} = 0$ and $u_o = u + \hat{u}$
else ▷ STEP 1
    Compute control input $\hat{u}$ such that $u_o := u + \hat{u}$ is the solution to the following problem.

$$\min_{u_o} J(\hat{x}_i, u_o) \ s.t. \ u_o \in \cap_{i \in I \backslash I_a} \Omega_i \tag{17}$$

    if no such u_o can be found **then** ▷ STEP 2
        Perform FT-LiDAR Estimation (Alg. 2).
        Exclude false sensors into I_a by criteria I and II.
        Compute $\hat{u}$ by solving (17).
        if no such u_o can be found **then** ▷ STEP 3
            for $u \notin \bigcap_{i \in I \backslash I_a} \Omega_i$ **do**
                Compute residue values $y_i - o(\hat{x}_i)$
                Include i into I_a with the largest residue.
            end for
        end if
    end if
end if

Theorem 3. *Given a safe set* $\mathcal{C}$ *and* $\bar{\zeta}_s$, *if the following conditions hold: (i) Assumption 1 holds, and (ii) scan match results* r *and* $\tilde{r}$ *can be found at each time step* k, *and (iii) there exists a function* $B(x)$ *satisfying the conditions in Proposition 1, then we have* $Pr(x_k \in \mathcal{C},\ \forall 0 \leq k \leq T) \geq 1 - \gamma - cT$ *when the adversary is present.*

Proof. Given condition (i), (ii), and $\bar{\zeta}_s$, by Theorem 2, $\|x - \hat{x}_i\| \leq \bar{\zeta}_i$ for each sensor $i \in I \backslash I_a$. In Alg. 3, u is computed by a nominal controller and $\hat{u}$ is computed by solving (17). By condition (iii) and Proposition 1, we have $Pr(x[k] \in C, 0 \leq k \leq T_d) \geq 1 - \gamma - cT_d$. □



(a) (b) (c) (d)

Fig. 2: Comparison between the estimated LiDAR observations (blue lines) and actual LiDAR observations (pink lines). Fig. 2a to 2b compares the estimated and actual LiDAR observations under attack Scenario I (INS1 compromised). The estimate based on INS1 deviates from the actual scan, causing the compromised sensor INS1 to become untrusted. Fig. 2c to 2d compares the estimated and actual LiDAR observations under attack Scenario II (INS1 and LiDAR compromised). Fig. 2a and Fig. 2c estimate the LiDAR scan using the compromised measurements from INS1. Fig. 2b and Fig. 2d estimate the LiDAR estimation using the measurements from INS2. The proposed approach removes the spoofed obstacle and aligns with the non-compromised sensor INS2.

V. Case Study

This section evaluates our proposed approach on a UAV delivery system in an urban environment. The UAV system is based on MATLAB UAV Package Delivery Example [33]. The UAV adopts stability, velocity and altitude control modules, rendering its position control dynamics to be:

$$\begin{bmatrix}[x]_1\\ [x]_2\end{bmatrix}_{k+1} = \begin{bmatrix}1 & -4.29\times10^{-5}\\ -1.47\times10^{-5} & 1\end{bmatrix}\begin{bmatrix}[x]_1\\ [x]_2\end{bmatrix}_k + \begin{bmatrix}0.0019 & -1.93\times10^{-5}\\ -2.91\times10^{-4} & 0.0028\end{bmatrix}\begin{bmatrix}[u]_1\\ [u]_2\end{bmatrix}_k, \quad (18)$$

where $x[k] = [[x]_1, [x]_2]^T$ is the UAV position, $[x]_1$ and $[x]_2$ represent the position of UAV on X-axis and Y-axis, respectively. The UAV has one LiDAR sensor and two inertial navigation system (INS) sensors, denoted as INS1 and INS2. The UAV maintains two EKFs associated with each INS sensor to estimate its position at each time k, denoted as $\hat{x}_1[k]$ and $\hat{x}_2[k]$, respectively.

The system operates in the presence of an adversary who can compromise one of the INS sensors and spoof the LiDAR sensor [10]. We compare our proposed approach with a baseline utilizing a PID controller with state estimations given by INS1.

We first demonstrate how our proposed approach selects sensors via Alg. 1 and Alg. 2 to obtain an accurate state estimation. We consider two attack scenarios. In Scenario I, the adversary compromises INS1 to deviate the measurement by -20 meters along the X-axis. In Scenario II, the adversary spoofs both the LiDAR sensor and INS1. The adversary biases INS1 sensor by -20 meters on X-axis and generates a random obstacle in the LiDAR scan within range of $[10, 15]$ meters and angle of $[-70, -60]$ degrees.

We present the estimated and actual LiDAR observations under Scenario I in Fig. 2a-2b. In Fig. 2a, we note that the estimated LiDAR observations $\mathcal{O}(\hat{x}_1, \mathcal{M})$ generated using state estimation $\hat{x}_1$ from INS1 significantly deviates from the actual LiDAR observations (the scan in pink color).

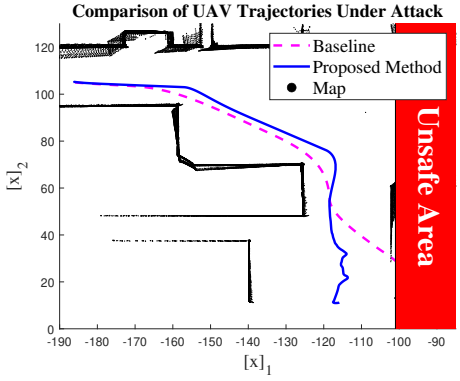


Fig. 3: Comparison of trajectories of the UAV when controlled using our proposed approach and the baseline.

The estimated LiDAR observations $\mathcal{O}(\hat{x}_2, \mathcal{M})$ align with the actual one as shown in Fig. 2b, which satisfies the criteria given in Section IV-C. Therefore, we treat INS2 as a trusted sensor while ignoring the measurements from INS1 when computing control input to the UAV.

We next compare the estimated and actual LiDAR observations under Scenario II in Fig. 2c-2d. The adversary manipulates the LiDAR observations by injecting a set of false points around position $(5.5, -11.6)$. In Fig. 2c, we observe a significant drift between the estimated LiDAR observations $\mathcal{O}(\hat{x}_1, \mathcal{M})$ and actual LiDAR observations $\mathcal{O}(x, S)$. In Fig. 2b, the obstacle points contained in sector c generated by the LiDAR spoofing attack are eliminated by Alg. 2, and thus the estimated LiDAR observations $\mathcal{O}(\hat{x}_2, \mathcal{M}\backslash c)$ aligns with the LiDAR observations $\mathcal{O}(x, S\backslash c)$. In this case, our proposed fault tolerant estimation indicates that INS1 should be ignored and INS2 can be trusted.

We finally present the trajectories of the UAV when using our proposed fault tolerant control (Alg. 3) and using the baseline. We present the trajectory generated using our proposed approach in Fig. 3 as the solid blue line, and the trajectory generated via the baseline using the dashed pink

line. We observe that our proposed approach ensures the UAV to successfully avoid all obstacles (the solid black lines) and the unsafe region (the region in red color), whereas the baseline controller leads to safety violation.

VI. Conclusion

In this paper, we studied the problem of safety-critical control for a LiDAR-based system in the presence of sensor faults and attacks. We considered the class of systems that is equipped with a set of sensors for state measurements and environment observations. We proposed a fault tolerant safe control framework for such systems to estimate their states and synthesize a control signal with safety guarantee. To obtain an accurate state estimate, we maintain a set of extended Kalman filters computed from different subsets of sensor measurements. For each estimate, we construct a simulated LiDAR scan based on the state estimate and an *a priori* known map of the environment, and exclude the state estimates that conflict with LiDAR measurements. When the LiDAR scan deviates from all of the state estimates, we remove the sector of the scan with the largest deviation. We proposed a control policy that selects a control input based on the fault tolerant estimate, and proved that the policy guarantees safety with a bounded probability using a control barrier certificate. We validated our proposed method with simulation studies on a UAV delivery system in an urban environment. We showed that our proposed approach ensures the UAV to be safe whereas a baseline controller causes it to reach an unsafe region.

References

[1] J. C. Knight, "Safety critical systems: challenges and directions," in *24th International Conference on Software Engineering*, 2002, pp. 547–550.

[2] 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, 2016.

[3] C. Tomlin, G. J. Pappas, and S. Sastry, "Conflict resolution for air traffic management: A study in multiagent hybrid systems," *IEEE Transactions on Automatic Control*, vol. 43, no. 4, pp. 509–521, 1998.

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

[5] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," *Sensors*, vol. 21, no. 6, p. 2140, 2021.

[6] C. Debeunne and D. Vivet, "A review of visual-LiDAR fusion based simultaneous localization and mapping," *Sensors*, vol. 20, no. 7, p. 2068, 2020.

[7] J. Petit and S. E. Shladover, "Potential cyberattacks on automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 546–556, 2014.

[8] A. J. Kerns, D. P. Shepard, J. A. Bhatti, and T. E. Humphreys, "Unmanned aircraft capture and control via GPS spoofing," *Journal of Field Robotics*, vol. 31, no. 4, pp. 617–636, 2014.

[9] A. Tatoglu and K. Pochiraju, "Point cloud segmentation with LiDAR reflection intensity behavior," in *IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 786–790.

[10] Y. Cao, C. Xiao, B. Cyr, Y. Zhou, W. Park, S. Rampazzi, Q. A. Chen, K. Fu, and Z. M. Mao, "Adversarial sensor attack on LiDAR-based perception in autonomous driving," in *ACM SIGSAC conference on Computer and Communications Security*, 2019, pp. 2267–2281.

[11] H. Shin, D. Kim, Y. Kwon, and Y. Kim, "Illusion and dazzle: Adversarial optical channel exploits against LiDAR for automotive applications," in *International Conference on Cryptographic Hardware and Embedded Systems*. Springer, 2017, pp. 445–467.

[12] 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.

[13] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, p. 13, 2011.

[14] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 4917–4932, 2017.

[15] 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.

[16] L. Niu, Z. Li, and A. Clark, "LQG reference tracking with safety and reachability guarantees under false data injection attacks," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 2950–2957.

[17] 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.

[18] 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.

[19] A. Clark, "Control barrier functions for stochastic systems," *Automatica*, vol. 130, p. 109688, 2021.

[20] J. Usevitch, K. Garg, and D. Panagou, "Strong invariance using control barrier functions: A Clarke tangent cone approach," in *59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 2044–2049.

[21] W. Xiao, C. G. Cassandras, C. A. Belta, and D. Rus, "Control barrier functions for systems with multiple control inputs," *arXiv preprint arXiv:2203.07978*, 2022.

[22] S. Dean, N. Matni, B. Recht, and V. Ye, "Robust guarantees for perception-based control," in *Learning for Dynamics and Control*. PMLR, 2020, pp. 350–360.

[23] S. Dean, A. J. Taylor, R. K. Cosner, B. Recht, and A. D. Ames, "Guaranteeing safety of learned perception modules via measurement-robust control barrier functions," in *CoRL*, 2020.

[24] C. Dawson, B. Lowenkamp, D. Goff, and C. Fan, "Learning safe, generalizable perception-based hybrid control with certificates," *IEEE Robotics and Automation Letters*, 2022.

[25] W. Xiao, T.-H. Wang, M. Chahine, A. Amini, R. Hasani, and D. Rus, "Differentiable control barrier functions for vision-based end-to-end autonomous driving," *arXiv preprint arXiv:2203.02401*, 2022.

[26] S. Wei, L. Ge, W. Yu, G. Chen, K. Pham, E. Blasch, D. Shen, and C. Lu, "Simulation study of unmanned aerial vehicle communication networks addressing bandwidth disruptions," in *Sensors and Systems for Space Applications VII*, vol. 9085. International Society for Optics and Photonics, 2014, p. 90850O.

[27] J. Liu and J.-M. Park, ""Seeing is not always believing": Detecting perception error attacks against autonomous vehicles," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2209–2223, 2021.

[28] A. Khazraei, H. Pfister, and M. Pajic, "Resiliency of perception-based controllers against attacks," https://cpsl.pratt.duke.edu/sites/cpsl.pratt.duke.edu/files/docs/khazraei_ld4c22.pdf, Tech. Rep.

[29] R. S. Hallyburton, Y. Liu, Y. Cao, Z. M. Mao, and M. Pajic, "Security analysis of camera-LiDAR fusion against black-box attacks on autonomous vehicles," *arXiv preprint arXiv:2106.07098*, 2021.

[30] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter," *IEEE Transactions on Automatic control*, vol. 44, no. 4, pp. 714–728, 1999.

[31] P. Biber and W. Straßer, "The normal distributions transform: A new approach to laser scan matching," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3. IEEE, 2003, pp. 2743–2748.

[32] P. Jagtap, S. Soudjani, and M. Zamani, "Formal synthesis of stochastic systems via control barrier certificates," *IEEE Transactions on Automatic Control*, vol. 66, no. 7, pp. 3097–3110, 2020.

[33] "Matlab UAV package delivery," https://www.mathworks.com/help/uav/ug/uav-package-delivery.html.