Control Barrier Functions for Safe CPS Under Sensor Faults and Attacks

Hongchao Zhang
Advisor: Prof. Clark
Committee: Prof. Fu and Prof. Zhang

SCPSLab
Electrical and Computer Engineering
Worcester Polytechnic Institute

July 22, 2020

To appear as: Andrew Clark, Zhouchi Li, Hongchao Zhang, Control Barrier Functions for Safe CPS Under Sensor Faults and Attacks 2020 59th IEEE Conference on Decision and Control (CDC)

Outline

- Introduction
- Related Work
- Preliminaries
- Proposed CBF Construction
- Joint Safety and Stability
- Case Study
- Conclusion and Future Work

Introduction: Motivation

- Safety is a fundamental requirement in critical applications.
- Safety is an especially challenging problem when sensors are affected by faults and malicious attacks.
    - Preventing the system from detecting and preventing safety violations.
    - Biasing estimates of the system state.

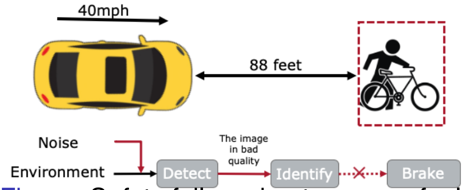


Figure: Safety failure due to sensor fault.[1]

[1] Phil McCausland, Nov. 9, 2019, Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk, https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281

Introduction: Problem Statement

How to design a control policy that guarantees that the system remains in a safe region with a desired probability when one or more sensor faults occur?

Related Work

- Detecting sensor faults [Wang et al‘04] and attacks [Chang et al‘18].
- Control Barrier Function(CBF) is proposed and used to verify and enforce safety properties[Ames et al‘14].
- CBFs for stochastic systems [Clark 19], high relative degree systems [Xiao et al‘19] and safe reinforcement learning [Cheng et al‘19] are investigated.
- CBFs for scenarios with sensor faults and attacks have not been considered.

Contributions

- Propose a class of Fault Tolerant Control Barrier Functions (FT-CBFs) for CPS with sensor faults.
- Derive sufficient conditions to ensure that safety is satisfied with a desired probability.
- Compose CBFs with Control Lyapunov Functions (CLFs) to provide joint guarantees on safety and stability of a desired goal set under faults.
- Evaluate our approach via a numerical study. The proposed control policy ensured convergence to a desired goal set without violating safety in the presence of a sensor attack.

Preliminaries: System Model

Consider a nonlinear control system with state $x_t \in \mathbb{R}^n$, input $u_t \in \mathbb{R}^p$ and the observation $y_t \in \mathbb{R}^q$. The impact of the fault is denoted by a_t.

$$dx_t = (f(x_t) + g(x_t)u_t)\, dt + \sigma_t\, dW_t \tag{1}$$
$$dy_t = (cx_t + a_t)\, dt + \nu_t\, dV_t \tag{2}$$

$f:$	$\mathbb{R}^n \to \mathbb{R}^n$	Locally Lipschitz
$g:$	$\mathbb{R}^n \to \mathbb{R}^{n\times p}$	Locally Lipschitz
W_t	$\in \mathbb{R}^n$	Brownian motion
V_t	$\in \mathbb{R}^q$	Brownian motion
c	$\in \mathbb{R}^{q\times n}$	Observation Matrix
a_t	$\in \mathbb{R}^q$	Impact of the fault
σ_t	$\in \mathbb{R}^{n\times n}$	Standard deviation
ν_t	$\in \mathbb{R}^{q\times q}$	Standard deviation

Table: Notation

Preliminaries: Safety and Fault Model

Safety Model:

- The safe region of the system $\mathcal{C} \subseteq \mathbb{R}^n$ defined by

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

  where $h \in C^2(\mathcal{C}) : \mathbb{R}^n \to \mathbb{R}$. Assume that $x_0 \in \mathrm{int}(\mathcal{C})$.

Fault Model:

- $\{r_1, \ldots, r_m\}$: the set of possible faults.
- $r \in \{r_1, \ldots, r_m\}$: the index of the fault.
- $\mathcal{F}(r_i) \subseteq \{1, \ldots, q\}$: affected observations.
- Assume that $\mathcal{F}(r_i) \cap \mathcal{F}(r_j) = \emptyset$ for $i \neq j$.

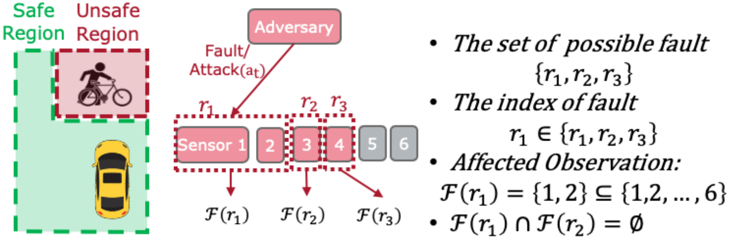


Figure: Illustration of safety model and fault model

Preliminaries: Problem Formulation

- **Problem Formulation:**
  Given a set $\mathcal{C}$ and a parameter $\epsilon \in (0, 1)$,
  construct a control policy: $\{y_{t'} : t' \in [0, t)\} \to u_t, \forall t$, s.t.

$$Pr(x_t \in \mathcal{C} \ \forall t) \geq (1 - \epsilon),$$

  for any fault $r \in \{r_1, \ldots, r_m\}$.

Preliminaries: Assumptions

Define $\bar{f}(x, u) = f(x) + g(x)u$.
We assume that the system (1)(2) satisfy the conditions [Reif et al‘2000]:

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 $[\frac{\partial \bar{f}}{\partial x}(x, u), c]$ is uniformly detectable.
3. Let ϕ be defined by

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

   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 $||x - \hat{x}||_2 \leq \epsilon_\phi$.

Preliminaries: EKF

The Extended Kalman Filter (EKF) for the system (1)(2) 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 $A_t = \frac{\partial \bar{f}}{\partial x}(\hat{x}_t, u_t)$.

Theorem 1 [Reif et al‘00]

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

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

Preliminaries: Control Barrier Function (CBF)

- CBF is used to guarantee safety constraints $h(x) \geq 0$.
- Impose an affine constraint on the control at each time step.
- Ensure that when h approaches the boundary, the the rate of increase $\frac{dh}{dx}$ decreases to zero.
- Hence, If the system is initially in the safe set and satisfies the CBF for all time t, then safety condition will be satisfied for all time

Preliminaries: SCBF

Theorem 2 [Clark'20]

For a system (1)–(2) 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 $\hat{h}_\gamma(x) = h(x) - \overline{h}_\gamma$. 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

$$\frac{\partial h}{\partial x}(\hat{x}_t)\overline{f}(\hat{x}_t, u_t) - \gamma||\frac{\partial h}{\partial x}(\hat{x}_t)K_t c||_2 + \frac{1}{2}\mathbf{tr}\left(\nu_t^T K_t^T \frac{\partial^2 h}{\partial x^2}(\hat{x}_t)K_t\nu_t\right) \geq -\hat{h}(\hat{x}_t). \quad (4)$$

Then $Pr(x_t \in \mathcal{C}\ \forall t|\ ||x_t - \hat{x}_t||_2 \leq \gamma\ \forall t) = 1$. We call a function h satisfying (4) a **Stochastic Control Barrier Function (SCBF)**.

Proposed CBF Construction: Strategy

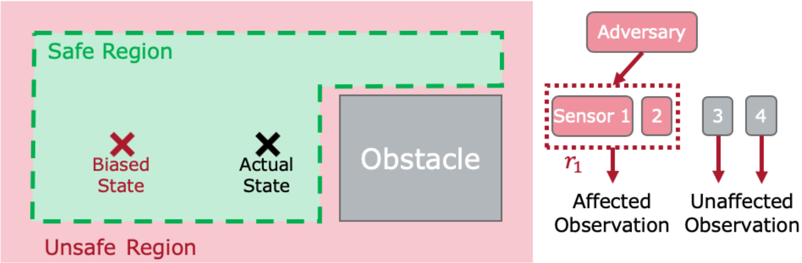


Figure: The case of one-fault-pattern observation

- If there is one fault pattern, then we can just exclude the affected sensors from the estimation process and use a CBF constraint.

Proposed CBF Construction: Strategy

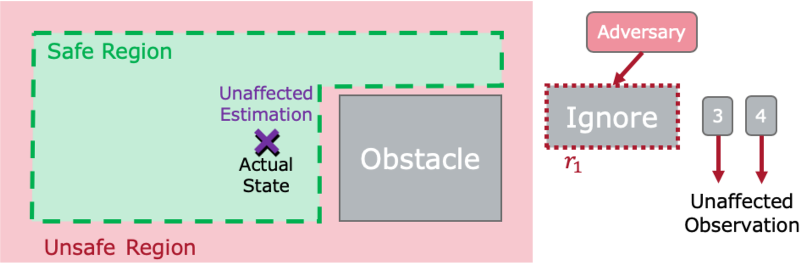


Figure: The case of one-fault-pattern observation

- If there is one fault pattern, then we can just exclude the affected sensors from the estimation process and use a CBF constraint.

Proposed CBF Construction: Strategy

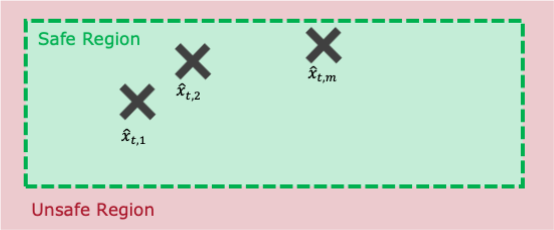


Figure: The case of m-fault-pattern observation

- If there are m fault patterns, then we can maintain m state estimates $\hat{x}_{t,i} : i = 1, ..., m$, each excluding sensors affected by one fault $\{1, ..., m\} \setminus F(r_i)$, and have a corresponding CBF constraint for each.
- The problem that arises is: what if there is no control input that simultaneously satisfies the CBF constraints? This can be viewed as a conflict between the estimates.

Proposed CBF Construction: Strategy

The state estimates may appear in three forms

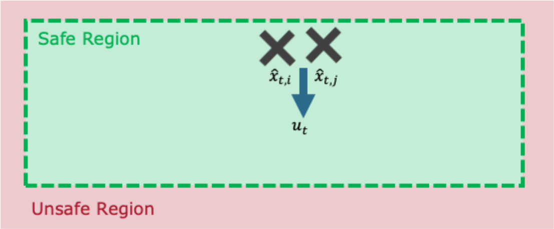


Figure: The state estimates have a single control input that ensures safety for both

Proposed CBF Construction: Strategy

The state estimates may appear in three forms

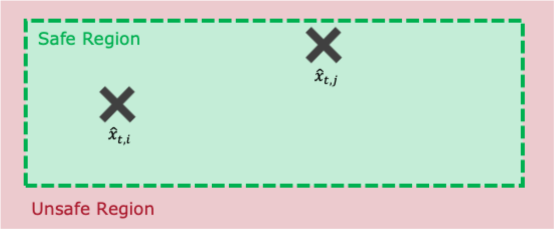


Figure: The state estimates are far enough from the boundary that the system can prioritize the more ”critical” one

Proposed CBF Construction: Strategy

The state estimates may appear in three forms

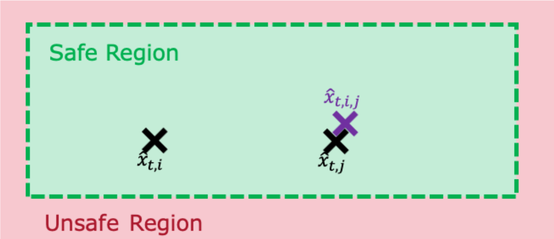


Figure: The state estimates are far enough apart that the erroneous estimate (estimate from faulty sensors) can be detected

- To enable detection, we may use the estimators of unaffected or pruned sensor sets $\hat{x}_{t,i,j} : i < j$, 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.

Proposed CBF Construction: Safe Control Policy

1. Attempt to select a control input that guarantees safety for the fault pattern that estimates are close to each other. If no such control input exists, then go to Step 2.
2. Prune constraints corresponding to $\hat{x}_{t,i}$ if $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2$ exceed a certain threshold value. If u_t still cannot be found, then go to Step 3.
3. Prune the corresponding constraints with the largest EKF residue.

Proposed CBF Construction: Safe Control Policy

We compute the control input by following three steps:

1. Select u_t satisfying all the constraints.
   - Define $X_t(\delta) = \{i : \hat{h}_i(\hat{x}_{t,i}) < \delta\}, \delta > 0$. Let $Z_t = X_t(\delta)$.
   - Define a collection of sets Ω_i, $i \in Z_t$, by

$$\Omega_i \triangleq \Big\{ u : \frac{\partial h_i}{\partial x}(\hat{x}_{t,i})\bar{f}(\hat{x}_{t,i}, u_t) - \gamma_i || \frac{\partial h}{\partial x}(\hat{x}_{t,i}) K_{t,i} c ||_2 + \frac{1}{2}\mathbf{tr}(\overline{\nu}_{t,i}^T K_{t,i}^T \frac{\partial^2 h_i}{\partial x^2}(\hat{x}_{t,i}) K_{t,i} \overline{\nu}_{t,i}) \geq -\hat{h}_i(\hat{x}_{t,i}) \Big\}. \quad (5)$$

   - 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.

Proposed CBF Construction: Safe Control Policy

2. If $\hat{x}_{t,i}$ deviates from $\hat{x}_{t,i,j}$ by more than a threshold value, the corresponding constraints Ω_i need to be removed from the set of constraints, since such deviations are likely to be due to faults. If u_t cannot be select, then go to Step 3.

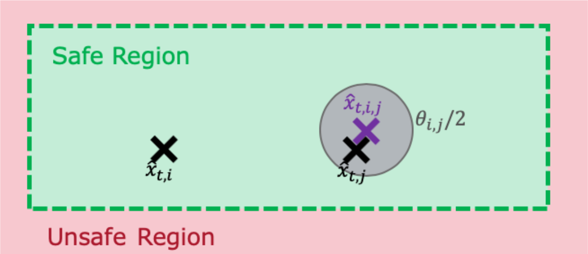


Figure: An example to help illustrate step 2

3. Remove the indices i from Z_t corresponding to the estimators with the largest residue values until there exists $u_t \in \bigcap_{i \in Z_t} \Omega_i$.

Proposed CBF Construction: FT-CBF

Theorem 3

Define $\overline{h}_{\gamma_i} = \sup\left\{h(x) : ||x - x^0||_2 \leq \gamma_i \text{ for some } x^0 \in h^{-1}(\{0\})\right\}$ 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
$$\Lambda_i(\hat{x}_{t,i})u > 0 \tag{6}$$
for all $i \in X_t'$.
2. For each i, when $r = r_i$,
$$Pr(||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \leq \theta_{ij}/2 \ \forall j, ||\hat{x}_{t,i} - x_t||_2 \leq \gamma_i \ \forall t) \geq 1 - \epsilon. \tag{7}$$

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

Proposed CBF Construction: 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.
- We analyze for the following special cases for LTI systems with dynamics

$$dx_t = (Fx_t + Gu_t)\,dt + \sigma dW_t. \tag{8}$$

- Half-plane Constraint with LTI System
  - 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 $\Lambda_i(\hat{x}_{t,i}) = a^T G$.
- Ellipsoid Constraints with LTI System is described in the thesis.

Proposed CBF Construction: Half-plane Constraint with LTI System

- Suppose $a^T G \neq 0$, we can choose an index $l \in \{1, \ldots, p\}$ such that $[a^T G]_l \neq 0$.
  - $[u]_s = 0$ for $s \neq l$
  - $[u]_l > 0$ if $[a^T G]_l > 0$
  - $[u]_l < 0$ if $[a^T G]_l < 0$.
  - Hence, we can choose u satisfying $a^T G u > 0$, the first condition of FT-CBF.
- Suppose $a^T G = 0$ and the system is controllable
  - There exists a minimum i such that $a^T F^i G \neq 0$, since the LTI system is controllable.
  - We can choose an index $l \in \{1, \ldots, p\}$ such that $[a^T F^i G]_l \neq 0$
  - $[u]_s = 0$ for $s \neq l$
  - $[u]_l > 0$ if $[a^T G]_l > 0$
  - $[u]_l < 0$ if $[a^T G]_l < 0$.
  - A high relative degree half-plane constraint can be satisfied with the desired probability.

Joint Safety and Stability

- Stability Problem Statement:
  Define the goal set $\mathcal{G}$ by $\mathcal{G} = \{\mathbf{x} : w(\mathbf{x}) \geq 0\}$ for some function w. The goal of the system is to asymptotically approach the set $\mathcal{G}$ with some desired probability.
- Stochastic Control Lyapunov Functions: [Florchinger'97]

Joint Safety and Stability: CBF-CLF

- The policy is similar to the CBF-based approach, with additional constraints to satisfy the stability condition. This leads to another *m* linear inequalities.
- 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 the optimization problem

$$\begin{array}{lll} \text{minimize} & u_t^T R u_t & \\ \text{s.t.} & \Lambda_i(\hat{x}_{t,j}) u_t \leq \overline{\omega}_j \; \forall j \in X_t(\delta) & \text{(CBF)} \\ & \Gamma_i(\hat{x}_{t,i}) u_t \leq \overline{\tau}_i \; \forall i \in Y_t(\overline{V}) & \text{(CLF)} \end{array} \tag{9}$$

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

Case Study: System Model

For a reach and avoid task, consider a differential drive wheeled mobile robot (WMR), with dynamics

$$\begin{pmatrix} [\dot{x}_t]_1 \\ [\dot{x}_t]_2 \\ \dot{\theta}_t \end{pmatrix} = \begin{pmatrix} \cos\theta_t & 0 \\ \sin\theta_t & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} [\omega_t]_1 \\ [\omega_t]_2 \end{pmatrix} + \mathbf{w}_t \tag{10}$$

- $([x_t]_1, [x_t]_2, \theta_t)^T$: the vector of the horizontal, vertical, and orientation coordinates for the WMR
- $([\omega_t]_1, [\omega_t]_2)^T$ (the linear velocity and the angular velocity around the vertical axis): the control input
- $\mathbf{w}_t$: the process noise.

Feedback Linearization [Chen et al‘20]:

- The controllable linearized model and the obsevation model with $\mathbf{w}_t'$: the process noise, $\mathbf{a}_t$: the impact of the attack and $\mathbf{v}_t$: the measurement noise.

$$\begin{pmatrix} [\dot{x}_t]_1 \\ [\dot{x}_t]_2 \\ [\ddot{x}_t]_1 \\ [\ddot{x}_t]_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} [x_t]_1 \\ [x_t]_2 \\ [\dot{x}_t]_1 \\ [\dot{x}_t]_2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} [u_t]_1 \\ [u_t]_2 \end{pmatrix} + \mathbf{w}_t' \tag{11}$$

$$\begin{pmatrix} [y_t]_1 \\ [y_t]_2 \\ [y_t]_3 \\ [y_t]_4 \\ [y_t]_5 \\ [y_t]_6 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} [x_t]_1 \\ [x_t]_2 \\ [\dot{x}_t]_1 \\ [\dot{x}_t]_2 \end{pmatrix} + \mathbf{a}_t + \mathbf{v}_t \tag{12}$$

Case Study: Settings

Settings:

- There is one redundant sensor for the horizontal coordinate and one for the vertical coordinate.
- The observation for the orientation coordinate θ_t is attack-free and noise-free, which enables feedback linearization based on the variable θ_t.
- The WMR is initially in the safe region. It aims to reach the goal area without entering unsafe region (e.g. black line).
- The adversary aims to drive the robot into unsafe region by spoofing its sensor measurement with consistent bias (e.g. red line).
- The CLF: $V(x) = (x_t - x_g)^T P_d (x_t - x_g)$, parameter settings are in the appendix

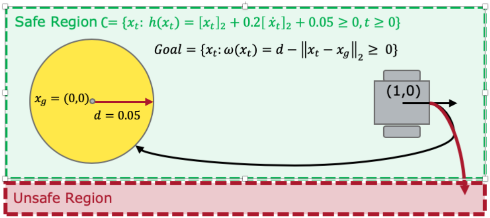
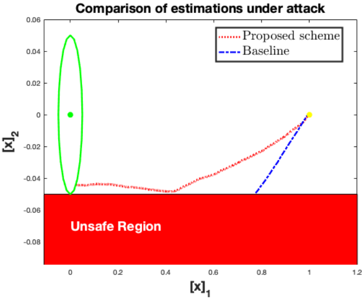


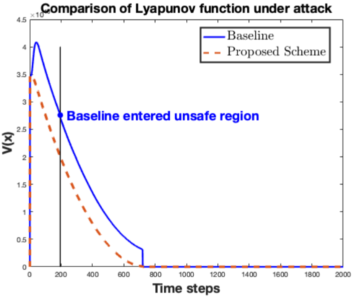
Figure: Visualization of settings in case study

Case Study: Numerical Results

The proposed algorithm was performed using Matlab.



(a)



(b)

Figure: Evaluation of our proposed approach on a linearized wheeled mobile robot model.

Conclusion and Future Work: Conclusion

- Proposed a new class of CBFs for safety and stability of stochastic systems under sensor faults and attacks.
- Constructed a CBF for each state estimator
- Proposed a scheme for using additional state estimators to resolve conflicts between constraints
- Derived sufficient conditions for ensuring safety with a desired probability
- Showed how to compose our proposed CBFs with CLFs to achieve joint safety and stability under faults and attacks.
- Our approach was validated using MATLAB-based numerical study.

Conclusion and Future Work: Future Work

- Attacks that jointly affect sensors and actuators.
- Analysis under arbitrary geometries and nonlinear dynamics.

Thank you for your time and attention

Advisor: Prof. Clark
Committee: Prof. Fu and Prof. Zhang

NSF grant CNS-1941670

Proposed CBF Construction: Safe Control Policy

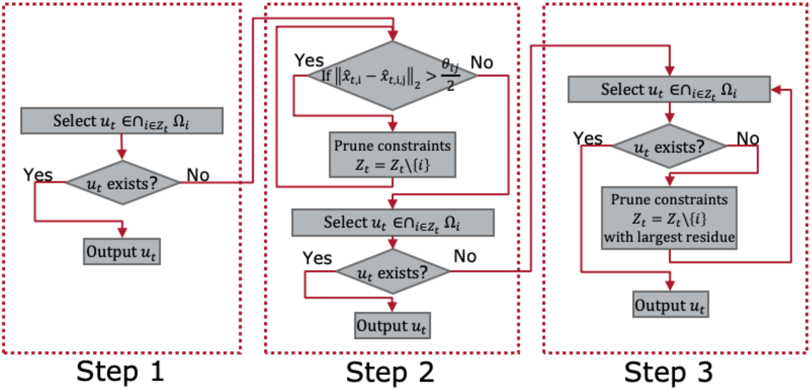


Figure: Safe control strategy flow chart