Provable Security of CPS using Control Barrier Functions

Workshop: Cyber-security in control of CPS: Recent developments and open challenges

61st IEEE Conference on Decision and Control, 2022

Kunal Garg
Postdoctoral Associate
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Motivation

Figure: Autonomous connected system*

Figure: Cyber-physical energy system*

Figure: Unmanned multi-agent system.

*Figure courtesy: internet.
Attacks on Cyber-physical systems

- Denial-of-Service (DoS) attack
- Integrity attack
  - False-data injection in sensor output
  - False-data injection in actuator input
- Remote access
Motivation

- CPS are vulnerable to cyber-attacks
  - Several attacks in past few years
    - E.g.: Stuxnet (Iran's nuclear facilities), Triton malware (Saudi Arabian petrochemical plant), Industroyer malware (Ukraine's power grid)
  - Adversarial attacks can lead to system failure, loss of money and human-life
  - Fast and effective response crucial for safe operations
  - Provable guarantees: necessary for safety-critical missions
  - Prior work: no formal guarantees on safety
  - Prior work: forward reachability-based (computationally expensive)
Fault model

Consider the dynamical control system:

\[ \dot{x} = f(x) + g(x)((u_v, u_s)), \quad x(0) \in S, u_v \in \mathcal{U}_v, u_s \in \mathcal{U}_s, \]

Vulnerable input  Secure input

Examples:
- Faulty actuator/propellor(s) in a quadrotor
- Loss of partial/full actuation in control surfaces
- Cyber-attack on input signal(s)
Problem formulation

Consider the dynamical control system:

\[ \dot{x} = f(x) + g(x)((u_v, u_s)) + d(t, x), \quad x(0) \in X_0 \subset S, u \in \mathcal{U}. \]

Problem

Given the system above, a safe set $S$ and the attack model described before, design an attack-detection mechanism and a safe input assignment policy such that for a set of initial conditions $x(0) \in X_0 \subset S$, the resulting closed-loop trajectories $x(\cdot)$ satisfy $x(t) \in S$ for all $t \geq 0$. 
Safety under Cyber-attack
Review: Control barrier functions

Consider the dynamical control system:

\[ \dot{x} = f(x) + g(x)u + d(t,x) \]

\[ x(0) \in S = \{ x \mid B(x) \leq 0 \}, \quad u \in \mathcal{U}, \quad |d(t,x)| \leq \delta \]

For the set \( S \), \( B: \mathbb{R}^n \rightarrow \mathbb{R} \) is called control barrier function (CBF) if \( \forall x \in \partial S \)

\[ H(x) := \inf_{u \in \mathcal{U}} \{ L_f B(x) + L_g B(x) u \} \leq -l_B \delta, \]

where \( L_f B(x) = \nabla B(x) \cdot f(x) \).

If \( B \) is a CBF then \( S \) can be rendered forward-invariant i.e., \( x(0) \in S \Rightarrow x(t) \in S \ \forall t \geq 0 \).

Safety condition under attacks

Consider the dynamical control system:

$$\dot{x} = f(x) + g(x)(u_s, u_v) + d(t, x)$$

$$x(0) \in S, \quad u_s \in \mathcal{U}_s, u_v \in \mathcal{U}_v, \quad |d(t, x)| \leq \delta$$

Define $$H(x, u_v) := \inf_{u_s \in \mathcal{U}_s} L_f B(x) + L_g B(x)(u_v, u_s)$$

Barrier function condition under attacks:

$$\sup_{u_v \in \mathcal{U}_v} H(x, u_v) \leq -l_B \delta \quad \forall \ x \in \partial S$$
Sampling-based computation of viability domain

Sampling of $\partial S$

Question: is it possible to evaluate $H$ only at a finite number of points and still guarantee safety?

Can we find $\gamma > 0$ and collection of points $\{x_i\}$ on $\partial S$ such that

$$\sup_{u_v \in \mathcal{U}_v} H(x, u_v) \leq -l_B \delta - \gamma$$

$\forall \ x \in \{x_i\}$

$$\sup_{u_v \in \mathcal{U}_v} H(x, u_v) \leq -l \delta$$

$\forall \ x \in \partial S$
Sampling of $\partial S$ for a unit sphere

Triangulate the boundary with $\{x_i\}_I$:

- For each $\bar{x} \in \partial S$, there is a triangular face $T_j$ with vertices $x_{j_1}, x_{j_2}, x_{j_3}$ such that $\bar{x} \in T_j$
- $\sup_{x_i \neq x_j} d_S(x_i, x_j) \leq d_a$

where $d_a \in (0, d_M]$ with $d_M = 2\sin^{-1}\left(\frac{3}{4}\right)$ and $d_S$ denotes the arc-length between $x_i, x_j \in \partial S$
Sampling of $\partial S$ for a unit sphere

$$\sup_{u \in \mathcal{U}} H(x, u) \leq -l_B \delta - l_H d_a$$

\[ \forall x \in \{x_i\} \]

Extendible to general sets $S$ that are diffeomorphic to a unit sphere

$$\sup_{u \in \mathcal{U}} H(x, u) \leq -l_B \delta$$

\[ \forall x \in \partial S \]

Extendible to higher dimensions using ‘simplex’
Sampling of $\partial S$ for a unit sphere

- For each $\bar{x} \in \partial S$, there exists $i \in I = \{1, 2, \ldots, N\}$ such that $|\bar{x} - x_i| \leq d_S(\bar{x}, x_i) \leq d_a$
- Minimum number of sampling points $N = (n + 1)$
- Larger $d_a \Rightarrow$ smaller $N \Rightarrow$ less points to check
- Smaller $d_a \Rightarrow$ smaller ($\gamma = -l_H d_a$) $\Rightarrow$ easier to satisfy
  \[
  \sup_{u_v \in \tilde{U}_v} H(x, u_v) \leq -l_B \delta - l_H d_a \ \forall \ x \in \{x_i\}
  \]
- Smaller $\tilde{U}_v \Rightarrow$ easier to satisfy the inequality
Computation of viability domain

Algorithm (data: $F, B, \mathcal{U}_v, \mathcal{U}_s, d_a, \epsilon_1, \epsilon_2, N_M, N_0, \delta$)

1. Initialize $c = 0, N = N_0, \tilde{\mathcal{U}}_v = \mathcal{U}_v$
2. While $c \leq c_M$
   3. While $N < N_M$
      4. Sample $\{x_i\}_1^N$ from $S_c = \{x \mid B(x) \leq -c\}$
      6. If $\{ \sup_{u_v \in \tilde{\mathcal{U}}_v} H(x_i, u_v) > -l_H d_a - l_B \} \neq \emptyset$
         7. $\tilde{\mathcal{U}}_v = \tilde{\mathcal{U}}_v \ominus \epsilon_1$
         8. $N = 2N$
         9. $\tilde{\mathcal{U}}_v = \mathcal{U}_v$
     10. $c = c + \epsilon_2$
     11. update $d_a$
     12. $N = N_0$
13. Return $c, \tilde{\mathcal{U}}_v$
Results
Attack detection using CBFs

Garg et al., "Control barrier function based attack-recovery with provable guarantees", CDC ‘22. (ThT08 session)
Approach overview

• Design nominal feedback law $\lambda(x) = (\lambda_v(x), \lambda_s(x))$
  • To be used when there is no attack
• Design safe feedback law $k(x) = (\cdot, k_s(x))$
  • To be used under attack
• Design an attack-detection mechanism
  • To be used for switching between $\lambda$ and $k$
CBF-based attack detection

- Under attack, actual input to the system is unknown
  - Not possible to evaluate $\dot{B}(x, u)$
- Use first-order approximation:

$$\exists t^* \in [0, \tau] : B(x(t - \tau)) = B(x(t)) - \dot{B}(x(t)) \tau + \frac{\ddot{B}(x(t^*)) \tau^2}{2}$$

- Thus:

$$\left| \frac{B(x(t)) - B(x(t - \tau))}{\tau} - \dot{B}(x(t)) \right| \leq \frac{\ddot{B}(x(t^*)) \tau}{2} \frac{\eta \tau}{2}$$
CBF-based attack detection

- Under nominal conditions:
  \[ \hat{B}(x(t), \tau) + \frac{\eta \tau}{2} \leq 0 \]

- Consider a set \( S_c = \{ x \mid B(x) \leq -c \} \) (computed using sampling-based algorithm).

- Attack flagging at time instant (with \( \hat{t}_d^0 = 0 \)):
  \[ \hat{t}_d^j = \inf \{ t \geq \max\{ \hat{t}_d^{j-1}, \bar{t} \} \mid \hat{B}(x(t), \tau) + \frac{\eta \tau}{2} > \gamma(t), \ x(t) \in S \setminus \text{int}(S_c) \} \]
  where \( \bar{t} = \inf \{ t \mid x(t) \in \partial S_c \} \) and \( \gamma(t) = \tilde{\delta} \tilde{c} e^{\delta(t-\bar{t})} \).
Detection-based switching policy

Closed-loop dynamical system:

$$\dot{x} = f(x) + g(x)(u_v(t, x), u_s(t, x)), \quad x(0) \in X_0 \subset S.$$ 

Let $\mathcal{T}_a = \bigcup_i [t_1^i, t_2^i]$ be the time intervals when attacks occur

$$u_v(t, x) = \begin{cases} 
\lambda_v(x) & \text{if } t \notin \mathcal{T}_a \\
 u_a(t) & \text{if } t \in \mathcal{T}_a
\end{cases}$$

$$u_s(t, x) = \begin{cases} 
\lambda_s(x) & \text{if } t \in [\hat{t}_d^{j-1} + T, \hat{t}_d^j) \\
 k_s(x) & \text{if } t \in [\hat{t}_d^j, \hat{t}_d^j + T)
\end{cases}$$
Safe quadrotor control under cyber-attack

Consider the 6-DOF quadrotor dynamics:

\[
\begin{align*}
\ddot{x} &= \frac{1}{m} \left( (c(\phi)c(\psi)s(\theta) + s(\phi)s(\psi)) u_f - k_t \dot{x} \right) \\
\ddot{y} &= \frac{1}{m} \left( (c(\psi)s(\theta) + s(\phi)c(\psi)) u_f - k_t \dot{y} \right) \\
\ddot{z} &= \frac{1}{m} \left( c(\theta)c(\phi)u_f - mg - k_t \dot{z} \right) \\
\dot{\phi} &= p + qs(\phi)t(\theta) + rc(\phi)t(\theta) \\
\dot{\theta} &= qc(\phi) - rs(\phi) \\
\dot{\psi} &= \frac{1}{c(\theta)} (qs(\phi) + rc(\phi)) \\
\dot{p} &= \frac{1}{I_{xx}} \left( - k_r p - qr(I_{zz} - I_{yy}) + \tau_p \right) \\
\dot{q} &= \frac{1}{I_{yy}} \left( - k_r q - pr(I_{xx} - I_{zz}) + \tau_q \right) \\
\dot{r} &= \frac{1}{I_{zz}} \left( - k_r r - pq(I_{yy} - I_{zz}) + \tau_r \right),
\end{align*}
\]

- Motor #4 is attacked
- Objective: Keep quadrotor hovering at 5m
- Safety: Quadrotor should not crash on ground
Quadrotor safety under failure/attacks

$\text{Attack} = 0 \quad \text{Detect} = 0$
Attack on system sensors

Attacks on system sensors

Consider the dynamical control system:

\[
\dot{x} = Ax + Bu
\]

\[
y = Cx = [\tilde{C}] \bar{C} x
\]

• Some of the output components \((y_a = \bar{C}x)\) can be attacked

• The pair \((\tilde{C}, A)\) might not be observable: state-reconstruction not possible

• Approach:
  • Design a switching observer to keep estimation error bounded
  • Compute set of initial conditions to account for estimation error
Observer design:

\[
\dot{x} = \begin{cases} 
A\dot{x} + L(Cx - C\dot{x}) + Bu & \text{if } t \notin T_a, \\
A\dot{x} + \tilde{L}(\tilde{C}x - \tilde{C}\dot{x}) + Bu & \text{if } t \in T_a, 
\end{cases}
\]
More work needs to be done!
Future work

• More general attacks on CPS
  • Simultaneous attacks on sensors and actuators
• Methods for detecting end of an attack
  • Decrease conservatism of attack detection
• Hybrid control strategy for improved performance
  • Guarantees on convergence (stability) along with safety
Acknowledgement

Collaborators at UC Santa Cruz:
• Prof. Ricardo G. Sanfelice
• Prof. Alvaro A. Cardenas
• Santiago J. Leudo

Sponsors:
• NSF: Grant no. ECS-1710621, CNS-2039054, and CNS-2111688
• AFOSR: Grant no. FA9550-19-1-0053, FA9550-19-1-0169, and FA9550-20-1-0238
• ARO: Grant no. W911NF-20-1-0253