Vulnerability analysis and attack detection for cyber-physical systems: a zonotopic approach

Presented by: **Jitao Li**

Joint work with: **Zhenhua Wang**[1] and **Lihua Xie**[2]

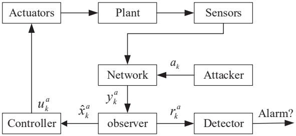
[1]School of Astronautics, Harbin Institute of Technology.

[2] School of Electrical and Electronic Engineering, Nanyang Technological University

January, 29th, 2021

Cyber physical systems

- **Cyber physical systems** refer to a new generation of systems consists of computation, communication and physical process.



Application fields

Security is important for CPS

Security

Attack

- Malicious adversaries may inject attacks to CPS.
- The motivation may be finance or terrorism.

Example

Stuxnet

- 2010 on Iranian Nuclear Facilities.
- 984 uranium enriching centrifuges are destroyed.

To avoid such catastrophes, system vulnerability analysis and attack detection are important.

System vulnerability

The system tolerance for potential stealthy attacks.

Vulnerability analysis

1. Is it possible for a CPS to be destroyed by potential stealthy attacks?
2. How to verify the safety of a CPS for potential stealthy attacks?
3. If a CPS is not safe, how to evaluate the degree of the threat of potential stealthy attacks?

Attack detection

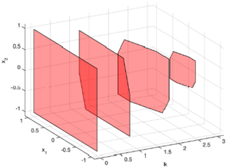
Attack detection: The strategy to detect specific attack.

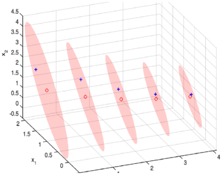
Types of attacks: denial-of-service attack (DoS), replay attack and false-data-injection attack (FDI)

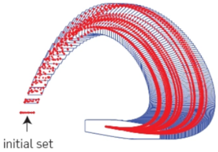
Reachability analysis

Reachability analysis

Compute a set which includes all possible state values based on available information







Geometrical sets

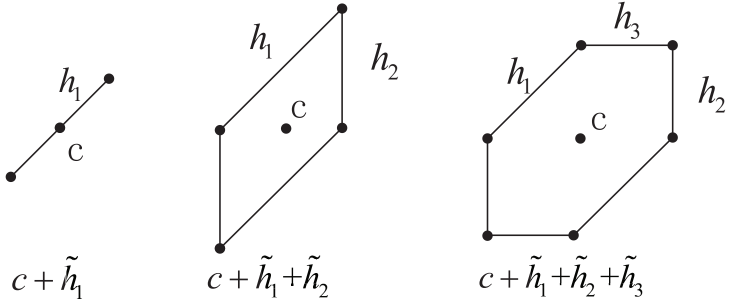
- Polytope: accuracy but complex
- Zonotope: good accuracy and computional efficiency
- Ellipsoid: simple but less accuracy

Zonotope

A zonotope is defined as

$$\mathcal{Z} = \left\{ x \in \mathbb{R}^n : x = c + \sum_{i=1}^{m} h_i b_i, b_i \in \mathbb{B} \right\} = \langle c, H \rangle,$$

where c determines center and h_i determines the shape. The geometrical interpretation is the Minkowski sum of m line segments $\tilde{h}_i = h_i \mathbb{B}$.

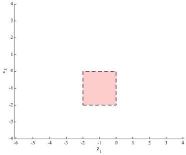


Property

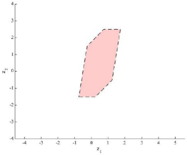
Minkowski sum

The Minkowski sum of two zonotopes $\mathcal{Z}_1 = \langle c_1, H_1 \rangle$ and $\mathcal{Z}_2 = \langle c_2, H_2 \rangle$ is also a zonotope, and the following equality holds

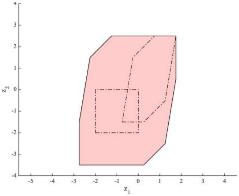
$$\langle c_1, H_1 \rangle \oplus \langle c_2, H_2 \rangle = \langle c_1 + c_2, [H_1, H_2] \rangle.$$



$$\mathcal{Z}_1 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbb{R}^2$$



$$\mathcal{Z}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.5 & 0.25 \\ 0 & 0.5 & 1.5 \end{bmatrix} \mathbb{R}^3$$



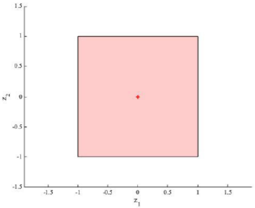
$$\mathcal{Z}_1 + \mathcal{Z}_2 = \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0.5 & 0.5 & 0.25 \\ 0 & 1 & 0 & 0.5 & 1.5 \end{bmatrix} \mathbb{R}^5$$

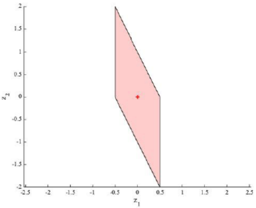
Property

Linear transformation

Given a zonotope $\mathcal{Z} = \langle p, H \rangle$, its linear transformation associated with a matrix K is

$$K\langle p, H \rangle = \langle Kp, KH \rangle.$$





$$\mathcal{Z} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\mathbb{B}^2 \xRightarrow{K = \begin{bmatrix}0 & -0.5\\ 1 & 1\end{bmatrix}} K\mathcal{Z}$$

Property

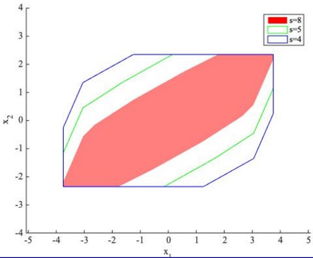
Order reduction

$\mathcal{Z}\langle c, H\rangle \subseteq \langle c, \downarrow_q(H)\rangle$: maintain large generator, over-approximate small one

1. Reordering the columns of the matrix H in decreasing Euclidean norm:

$$H = [h_1, h_2, \cdots, h_m], ||h_j|| \geq ||h_{j+1}||, j = 1, \cdots m-1.$$

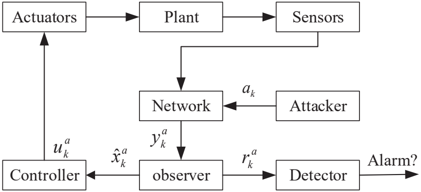
2. Replacing the last $m - q + n$ smallest columns by a diagonal matrix.



Cyber-Physical system

CPS

- State dynamics of physical plant.
- Remote estimator and controller.



Cyber-Physical system

Consider the plant is modeled as a discrete-time linear time invariant system

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + w_k, \\ y_k = Cx_k + v_k, \end{cases} \tag{1}$$

where $x_k \in \mathbb{R}^{n_x}$ denotes the state, $y_k \in \mathbb{R}^{n_y}$ denotes the measurement, $u_k \in \mathbb{R}^{n_u}$ denotes the control input, $w_k \in \mathbb{R}^{n_w}$ is the process disturbance, and $v_k \in \mathbb{R}^{n_v}$ is the measurement noise.

Assumptions

Assumption 1

(A, C) is detectable and (A, B) is stabilizable.

Assumption 2

w_k, v_k and the initial state x_0 are unknown but peak bounded, i.e.

$$x_0 \in \overline{\mathcal{X}}_0 = \langle p_0, H_0 \rangle,$$

$$w_k \in \mathcal{Z}_w = \langle 0, H_w \rangle, v_k \in \mathcal{Z}_v = \langle 0, H_v \rangle.$$

Assumption 3

Estimator: $\hat{x}_{k+1} = A\hat{x}_k + Bu_k + L(y_{k+1} - CA\hat{x}_k - CBu_k)$,

Controller $u_k = K\hat{x}_k$

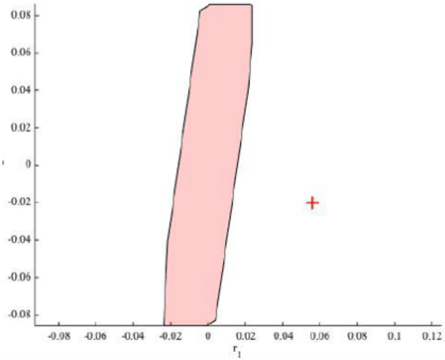
Any choice of L and K such that $A - LCA$ and $A + BK$ are stable is acceptable.

Detector

Estimation error: $e_k = x_k - \hat{x}_k$, Residual signal: $r_{k+1} = y_{k+1} - CA\hat{x}_k - CBu_k$

$$\begin{cases} e_{k+1} = (A - LCA)e_k + (I - LC)w_k - Lv_{k+1}, \\ r_{k+1} = CAe_k + Cw_k + v_{k+1}. \end{cases}$$

$r_k \notin \mathcal{R}_k$ – Attack or fault is detected.



Attack model

To capture the attacks' impact on the system, we rewrite system (1) as

$$\begin{cases} x_{k+1}^a = Ax_k^a + Bu_k^a + w_k, \\ y_k^a = Cx_k^a + v_k + Fa_k, \end{cases} \tag{2}$$

where $(*)^a$ denotes the variable $*$ in the presence of attacks, $a_k \in \mathbb{R}^{n_a}$ is the malicious attack signal, $F \subseteq \{\gamma_1, \cdots, \gamma_{n_y}\}$ is the attacker's sensor selection matrix and is unknown to system designers, where γ_i is the ith vector of the canonical basis of $\mathbb{R}^{n_y}$.

Assumption 3

The adversary knows the system parameters, i.e., A, B, C, L, K.

Assumption 4

The adversary has the required resources to launch any suitable attack signals.

Problem formulation

Safe region

$\mathcal{X} = \langle x_s, H_s \rangle$, where x_s and H_s are known. Such a region may represent, for example, the constraint of the acceleration and the velocity of a vehicle.

Stealthy attack

An attack sequence is said to be stealthy if $r_k^a \in \mathcal{R}_k, \forall k \in \mathbb{N}_+$ holds.

Vulnerability analysis

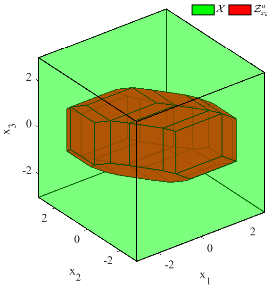
1. Is it possible for a CPS to be destroyed by potential stealthy attacks?
2. How to verify the safety of a CPS for potential stealthy attacks?
3. If a CPS is not safe, how to evaluate the degree of the threat of potential stealthy attacks?

Main idea

Vulnerability

Denote $\mathcal{Z}^a_{x_k}$ as the out-approximation of the reachable set of x^a_k.

1. The CPS is strictly vulnerable if $\exists a_k$ such that $\mathcal{Z}^a_{x_k}$ is unbounded.
2. The CPS is vulnerable if there exists a stealthy attack a_k such that $\mathcal{Z}^a_{x_k} \nsubseteq \mathcal{X}$.
3. The CPS is safe if $\mathcal{Z}^a_{x_k} \subseteq \mathcal{X}$ holds for any stealthy attack.



Strictly vulnerable

Strict vulnerability

$\exists a_k$ such that Δr_k is bounded while Δe_k is unbounded.

$$\Delta e_{k+1} = (A - LCA)\Delta e_k - LFa_{k+1},$$
$$\Delta r_{k+1} = CA\Delta e_k + Fa_{k+1}.$$

Conditions for strict vulnerability

Δe_k is unbounded iff $\exists$ an eigenvector ζ corresponding to the unstable mode of A satisfies $\zeta \in \text{range}(Q_{es}), -C\zeta \in \text{range}(F)$, where Q_{es} is the controllability matrix of the pair $(A - LCA, -LF)$[a].

[a] Mo Y, Sinopoli B. False data injection attacks in control systems[C]//Preprints of the 1st workshop on Secure Control Systems. 2010: 1-6.

Vulnerable

Vulnerability

If there exists a stealthy attack a_k such that $\mathcal{Z}^a_{x_k} \nsubseteq \mathcal{X}$.

Vulnerability

Calculate $\mathcal{Z}^a_{x_k}$ includes all possible x^a_k such that $r^a_k \in \mathcal{R}_k$.

$$\begin{cases} x^a_{k+1} = Ax^a_k + Bu^a_k + w_k, \\ y^a_k = Cx^a_k + v_k + Fa_k, \\ u^a_k = K\hat{x}^a_k, \\ r^a_{k+1} = y^a_{k+1} - CA\hat{x}^a_k - CBu^a_k. \end{cases} \tag{3}$$

Verification of safety

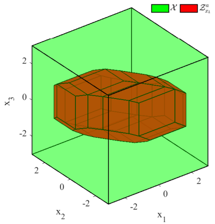
Zonotope containment

Consider two zonotopes $\langle c_1, H_1 \rangle$ and $\langle c_2, H_2 \rangle$, the relation $\langle c_1, H_1 \rangle \subseteq \langle c_2, H_2 \rangle$ holds if there exists a matrix Γ and a vector β such that [a]

$$H_1 = H_2 \Gamma, c_2 - c_1 = H_2 \beta, \left\| \begin{bmatrix} \Gamma & \beta \end{bmatrix} \right\|_\infty \leq 1.$$

[a] S. Sadraddini, R. Tedrake, Linear encodings for polytope containment problems, IEEE 58th CDC, 2019, pp. 4367–4372.

- $\mathcal{Z}^a_{x_k} \nsubseteq \mathcal{X}$–Vulnerable
- $\mathcal{Z}^a_{x_k} \subseteq \mathcal{X}$–Safe



Metric of vulnerability

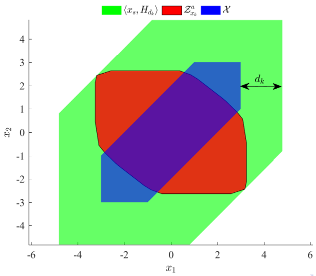
Metric

For a vulnerable CPS, evaluate the threat degree of potential stealthy attacks.

One-sided Hausdorff distance

The one-sided Hausdorff distance from set $\mathcal{Z}^a_{x_k}$ to set $\mathcal{X}$ is defined as

$$d(\mathcal{Z}^a_{x_k}, \mathcal{X}) = \max_{a \in \mathcal{Z}^a_{x_k}} \min_{b \in \mathcal{X}} ||a - b||.$$



Simulatioin

LTI system

$$\begin{cases} x_{k+1}^a = Ax_k^a + Bu_k^a + w_k, \\ y_k^a = Cx_k^a + v_k + Fa_k, \end{cases}$$

with

$$A = \begin{bmatrix} 0.62 & 0.21 & 0.03 \\ 0.08 & 0.72 & 0.54 \\ 0.02 & 0.02 & 0.65 \end{bmatrix}, B = \begin{bmatrix} 0.07 & 1 \\ 0.23 & 0.5 \\ 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

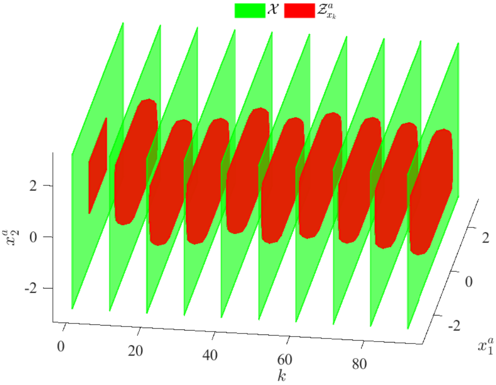
Uncertainties: $w_k \in \langle 0, 0.2I \rangle$ and $v_k \in \langle 0, 0.1I \rangle$

Safe region: $\mathcal{X} = \langle 0, 3I \rangle$.

Eigenvalue of A: $\begin{bmatrix} 0.8755 + 0.0000i \\ 0.5573 + 0.0540i \\ 0.5573 - 0.0540i \end{bmatrix}$, Thus $\mathcal{Z}_{x_k}^a$ is bounded.

Simulation

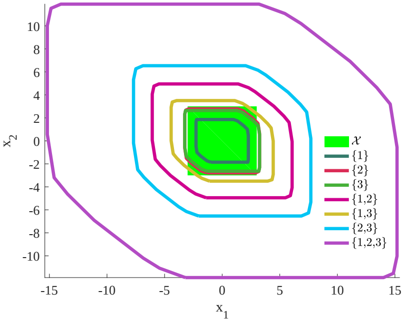
$F = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$, safe



Simulation

Table 1
Vulnerability under different cases

Attacked sensors	Secure	Degree of vulnerability
{1}	Yes	–
{2}	No	0.2439
{3}	No	0.2492
{1,2}	No	3.0431
{1,3}	No	1.4312
{2,3}	No	4.6958
{1,2,3}	No	12.1816



If we can only protect one sensor from adversaries due to resource limitation, sensor 2 should be protected.

System dynamics

System description

Consider linear time invariant system

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + w_k, \\ y_k = Cx_k + v_k + a_k. \end{cases},$$

where a_k indicates the injected bias on sensor data.

Assumptions

- (A, C) is detectable and (A, B) is stabilizable.
- w_k, v_k and the initial state x_0 are unknown but peak bounded, i.e.

$$x_0 \in \overline{\mathcal{X}}_0 = \langle p_0, H_0 \rangle,$$

$$w_k \in \mathcal{Z}_w = \langle 0, H_w \rangle, v_k \in \mathcal{Z}_v = \langle 0, H_v \rangle.$$

Problem formulation

Predicted state set

Take the state x_{k-1} is bounded in the zonotope $\overline{\mathcal{X}}_{k-1} = \langle p_{k-1}, H_{k-1} \rangle \subseteq \mathbb{R}^{n_x}$ as a prior, given the system dynamic, the predicted state set is defined as the set of all possible solutions of x_k, i.e,

$$\mathcal{X}_{k/k-1} = \left\{ x_k \in \mathbb{R}^{n_x} : (x_k - Bu_{k-1}) \in (A\overline{\mathcal{X}}_{k-1} \oplus \mathcal{Z}_w) \right\}.$$

Measurement state set

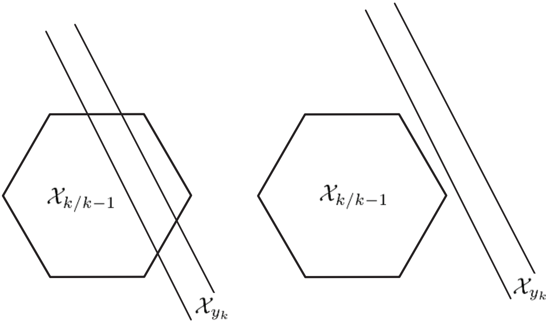
Given the observervation equation and Assumption 2 holds, the measurement state set is defined as the set of all possible solutions x_k, which can be reached by y_k and v_k, i.e,

$$\mathcal{X}_{y_k} = \left\{ x_k \in \mathbb{R}^{n_x} : (y_k - Cx_k) \in \mathcal{Z}_v \right\}.$$

Problem formulation

Objective

- How to checking the existence of the intersection of $\mathcal{X}_{k/k-1}$ and $\mathcal{X}_{y_k}$?
- What is the detection performance that the proposed method?

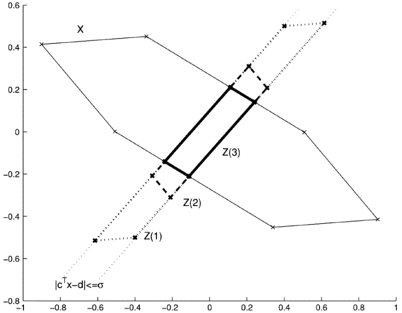


Set computation

Parameterized $\overline{\mathcal{X}}_k$

$\exists L_k \in \mathbb{R}^{n_x \times n_y}$ such that $\mathcal{X}_{k/k-1} \cap \mathcal{X}_{y_k} \subseteq \overline{\mathcal{X}}_k = \langle p_k, H_k \rangle$, where

$$
\begin{aligned}
p_k &= Ap_{k-1} + Bu_{k-1} + L_k(y_k - CAp_{k-1} - CBu_{k-1}),\\
H_k &= \Big[(A - L_kCA) \downarrow_q (H_{k-1}) \quad (I - L_kC)H_w \quad -L_kH_v \Big].
\end{aligned}
$$



Set computation

Size criterion

The size of the segments of the zonotope: $J_k = ||H_k||_F^2 = \text{tr}(H_k^T H_k)$

optimal correction matrix

By minimizing the size criterion J_k.

$$L_k = (AP_{k-1}A^T + Q_w)C^T Y_{k-1}^{-1},$$

where

$$P_{k-1} = \downarrow_q (H_{k-1}) \downarrow_q (H_{k-1})^T, Q_w = H_w H_w^T,$$
$$Q_v = H_v H_v^T, Y_{k-1} = C(AP_{k-1}A^T + Q_w)C^T + Q_v.$$

Set computation

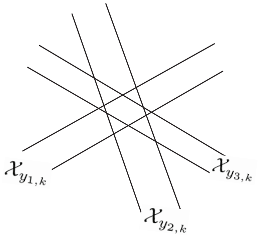
Predicted set

$$\mathcal{X}_{k/k-1} = \langle p^p_{k/k-1}, H^p_{k/k-1} \rangle = A\overline{\mathcal{X}}_{k-1} \oplus \{Bu_{k-1}\} \oplus \mathcal{Z}_w.$$

with $p^p_{k/k-1} = Ap_{k-1} + Bu_{k-1}, H^p_{k/k-1} = \begin{bmatrix} AH_{k-1} & H_w \end{bmatrix}.$

Measurement state set

- For the i-th subequation $y_{i,k} = C_i x_k + v_{i,k}$, the corresponding state set is $\mathcal{X}_{y_{i,k}} = \{x_k \in \mathbb{R}^{n_x} : y_{i,k} - \overline{v}_i \leq C_i x_k \leq y_{i,k} + \overline{v}_i\}$,
- $\mathcal{X}_{y_k} = \bigcap_{i=1}^{n_y} \mathcal{X}_{y_{i,k}}$.



Intersection checking based on projection

If $\exists \mathcal{X}_{y_{i,k}}$ such that $\mathcal{X}_{k/k-1} \cap \mathcal{X}_{y_{i,k}} = \emptyset$, we have $\mathcal{X}_{k/k-1} \cap \mathcal{X}_{y_k} = \emptyset$.

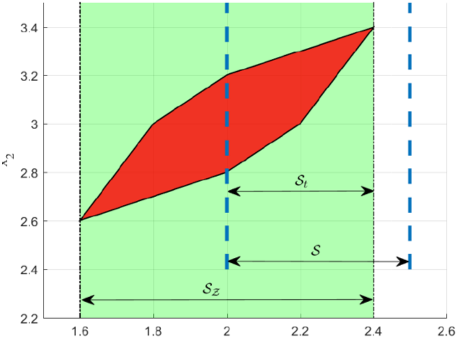
Projection method

An attack is detected if

$$y_{i,k} + \overline{v}_i < q_i^{\text{min}},$$
$$y_{i,k} - \overline{v}_i > q_i^{\text{max}},$$

where

$$q_i^{\text{min}} = C_i p^p_{k/k-1} - \|(H^p_{k/k-1})^T C_i^T\|_1,$$
$$q_i^{\text{max}} = C_i p^p_{k/k-1} + \|(H^p_{k/k-1})^T C_i^T\|_1.$$



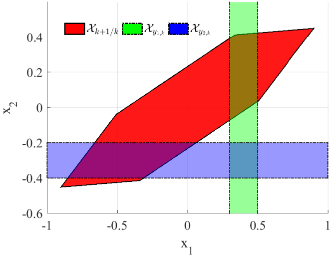
Example

$$\mathcal{X}_{k/k-1} = \left\langle \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.2812 & 0.1968 & 0.4235 \\ 0.0186 & 0.2063 & 0.2267 \end{bmatrix} \right\rangle$$

$$\mathcal{X}_{y_{1,k}} = \{x_k \in \mathbb{R}^2 : 0.3 \leq \begin{bmatrix} 1 & 0 \end{bmatrix} x_k \leq 0.5\},$$
$$\mathcal{X}_{y_{2,k}} = \{x_k \in \mathbb{R}^2 : -0.4 \leq \begin{bmatrix} 0 & 1 \end{bmatrix} x_k \leq -0.2\}.$$

Using projection method, we have

$$q_1^{\text{min}} = -0.9015, q_1^{\text{max}} = 0.9015,$$
$$q_2^{\text{min}} = -0.4516, q_2^{\text{max}} = 0.4516,$$



The projection method fails, but $\mathcal{X}_{k/k-1} \cap \mathcal{X}_{y_k} = \emptyset$ holds

Intersection checking based on polytopic conversion

Polytopic conversion

$\mathcal{X}_{k/k-1} \Longrightarrow \mathcal{P}_{k/k-1} = \{x_k \in \mathbb{R}^{n_x} : Qx_k \leq q\}$.

$\mathcal{X}_{y_k} \Longrightarrow \mathcal{P}_{y_k} = \{x_k \in \mathbb{R}^{n_x} : Q_y x_k \leq q_y\}$.

An attack is detected if not $\exists x_k$ such that

$$\mathcal{P}_{k/k-1} \cap \mathcal{P}_{y_k} = \left\{ x_k \in \mathbb{R}^{n_x} : \begin{bmatrix} Q \\ Q_y \end{bmatrix} x_k \leq \begin{bmatrix} q \\ q_y \end{bmatrix} \right\}.$$

Features

- Projection: simple but less satisfactory
- Polytopic conversion: satisfactory but heavy computation

Stealthy attack set

Stealthy attack set $\mathcal{A}_k$

For given detection method, any attack outside $\mathcal{A}_k$ will be detected.

$\mathcal{A}_k$ by projection

$$|a_{i,k}| \leq 2\|(H^p_{k/k-1})^T C_i^T\|_1 + 2\overline{v}_i,$$

where $a_{i,k}, i = 1, 2, \cdots, n_y$ is the i-th component of a_k.

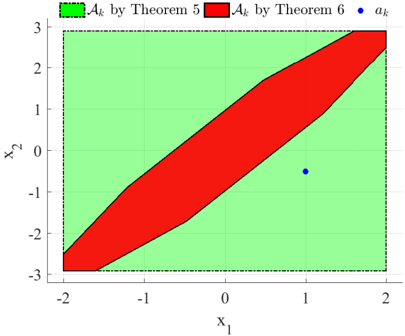
$\mathcal{A}_k$ by polytopic conversion

$\mathcal{A}_k = \langle 0, H^a_k \rangle$, with

$$H^a_k = \begin{bmatrix} CH^p_{k/k-1} & H_v & CH^p_{k/k-1} & H_v \end{bmatrix}.$$

Example

- Predicted state set $\mathcal{X}_{k/k-1} = \left\langle \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.2812 & 0.1968 & 0.4235 \\ 0.0186 & 0.2063 & 0.2267 \end{bmatrix} \right\rangle$
- Noise $v_k \in \mathcal{Z}_v = \langle 0, \mathrm{diag}(0.1, 0.1) \rangle$
- Output matrix $C = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$



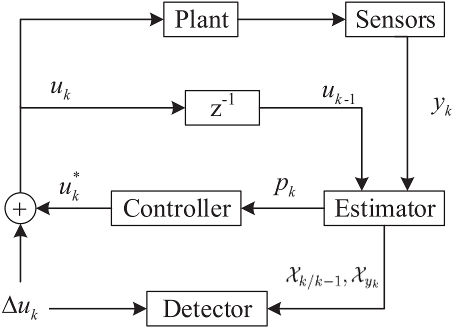
Features:

- $\mathcal{A}_k$ by polytopic conversion $\subset$ $\mathcal{A}_k$ by projection.
- a_k may not be detected if $a_k \in \mathcal{A}_k$.
- Particularly, replay attack cannot be detected.

Replay attack

Replay attack

Record a period of y_k, then replay it when implementing attack: $y_k = y_{k-\Delta k}$



- $\Delta u_k \in \langle 0, H_u \rangle$ is randomly generated with a uniform distribution
- Δu_k is chosen independent of u_k^*

Analysis

Control performance loss: increased size of $\mathcal{X}_k$

$\Delta J = \text{tr}(\mathcal{L}(Q_u))$, where $Q_u = H_u H_u^T$ and $\mathcal{L}(Q_u)$ is a linear operator defined as

$$\mathcal{L}(Q_u) = \sum_{i=0}^{\infty} (A+BK)^i B Q_u B^T ((A+BK)^i)^T$$

Detection performance: the deviation size of $\mathcal{Y}_k$ under replay attacks.

$\Lambda = 2\text{tr}(\mathcal{D}(Q_u))$, where

$\mathcal{D}(Q_u) = C(A+BK)\mathcal{D}(Q_u)(A+BK)^T C^T + CBQ_u B^T C^T.$

Design of Δu_k

$$Q_u = \arg\max_{Q_u \succeq 0} \ 2\text{tr}(\mathcal{D}(Q_u)), \qquad \text{subject to} \quad \text{tr}(\mathcal{L}(Q_u)) \le \delta,$$

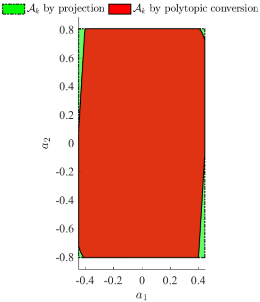
where δ is the tolerable control performance loss.

Simulation Results

System parameters:

$$A = \begin{bmatrix} 0.1046 & -0.0725 \\ 1.7287 & 0.0974 \end{bmatrix}, B = \begin{bmatrix} 0.4198 \\ 2.6429 \end{bmatrix},$$

$$C = I, |w_k| \leq \begin{bmatrix} 0.1 & 0.1 \end{bmatrix}^T, |v_k| \leq \begin{bmatrix} 0.1 & 0.1 \end{bmatrix}^T$$



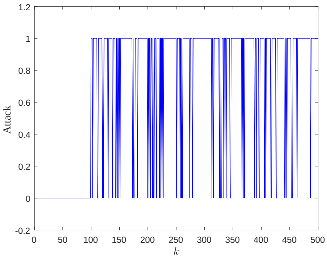
- Polytopic conversion better than projection
- $\mathcal{A}_k$ includes all possible stealthy attacks

Simulation Results

Table: Computational time (CPU: Intel(R) Core(TM) i7-9750H @2.6GHz)

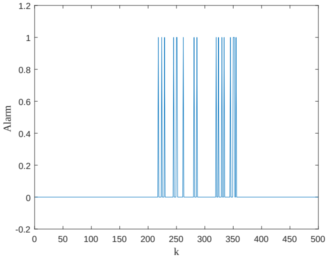
Method	Time
Projection	0.088760s
Polytopic conversion	12.291986s

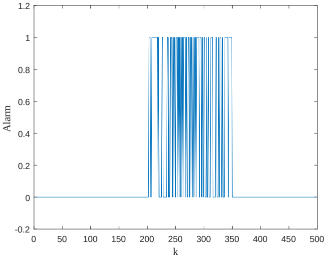
$a_k = [0.3 \quad 0.5]^T, k \geq 100$



Simulation Results

- Record $y_{50} - y_{200}$, replay at $k = 200 - 350$
- From left to right: $|\Delta u(k)| \leq 0.1$, $|\Delta u(k)| \leq 0.3$





Simulation Results

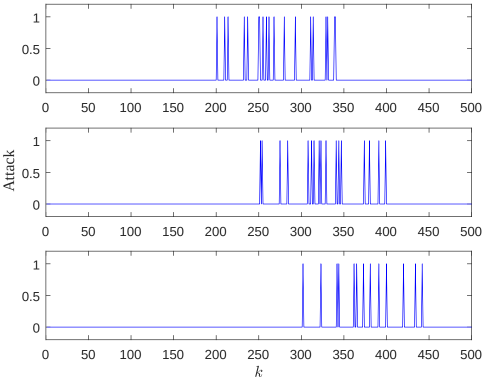


Fig: From top to bottom, detection results when the recorded sensor data $y_{50} \sim y_{200}$ is replayed at $k = 200$, $k = 250$ and $k = 300$, respectively

Conclusion

- A CPS vulnerability analysis and attack detection framework based on zonotopic reachability analysis is established.
- The capability of stealthy attacks is analyzed. This is helpful for establishing defence strategy.
- The detection of false data injection attack and replay attak is considered.

Thank You. Questions?