Online Stealthy Attack for Cyber-Physical Systems

with Applications in Autonomous Driving

Xingzhou CHEN
Supervisor: Prof. Ling SHI

Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology

Nov 01, 2022

Self-introduction

Education history

- Sept 2017 – July 2021,
  College of Control Science and Engineering, Zhejiang University
- Sept 2021 – now,
  Department of Electronic and Computer Engineering, HKUST

Key classes taken in HKUST

- ELEC 5600: Linear-system Theory
- ELEC 5650: Introduction to Networked Sensing, Estimation and Control
- MATH 5411/5412: Advanced Probability Theory I and II
- MATH 5311/5312: Advanced Numerical Method I and II

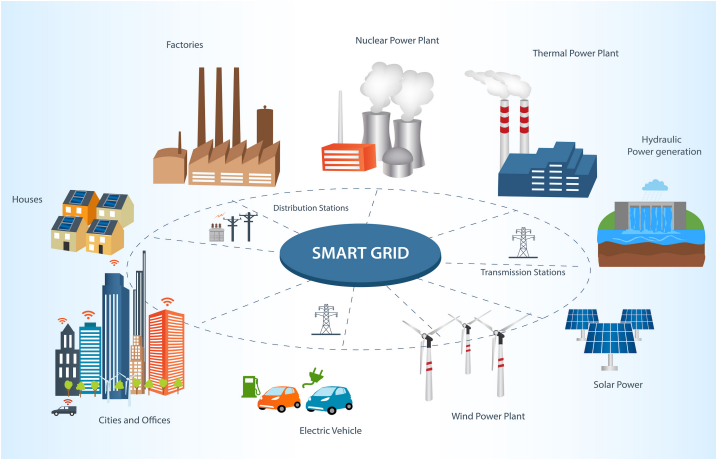
Current research interests

- Security in cyber physical system, and Trajectory optimization

Table of Contents

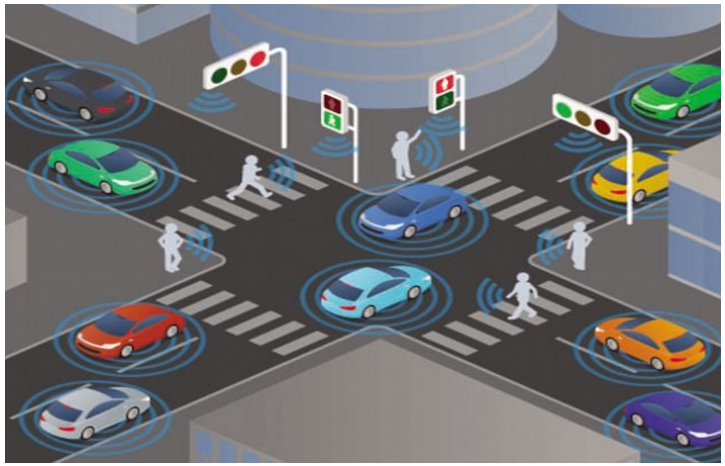
Background and Motivation

Our world is becoming more automatic, efficient and connected.



(a) smart grid

(b) smart industry



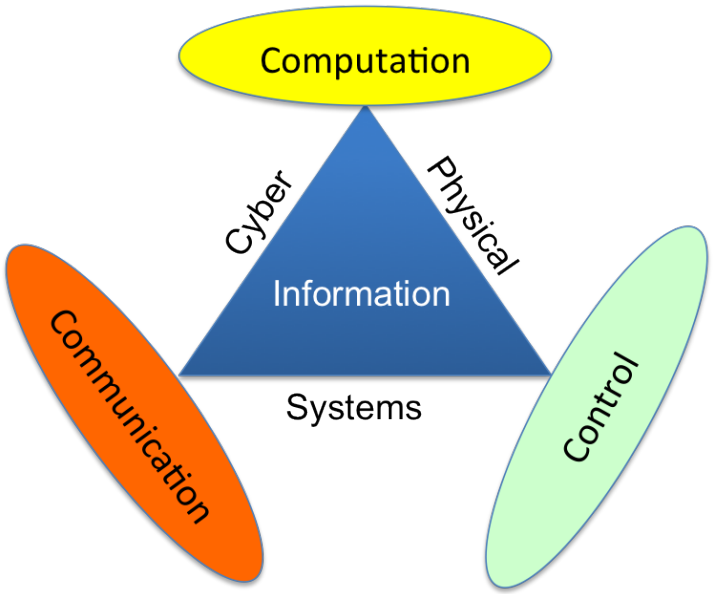
(c) smart traffic

(d) smart house

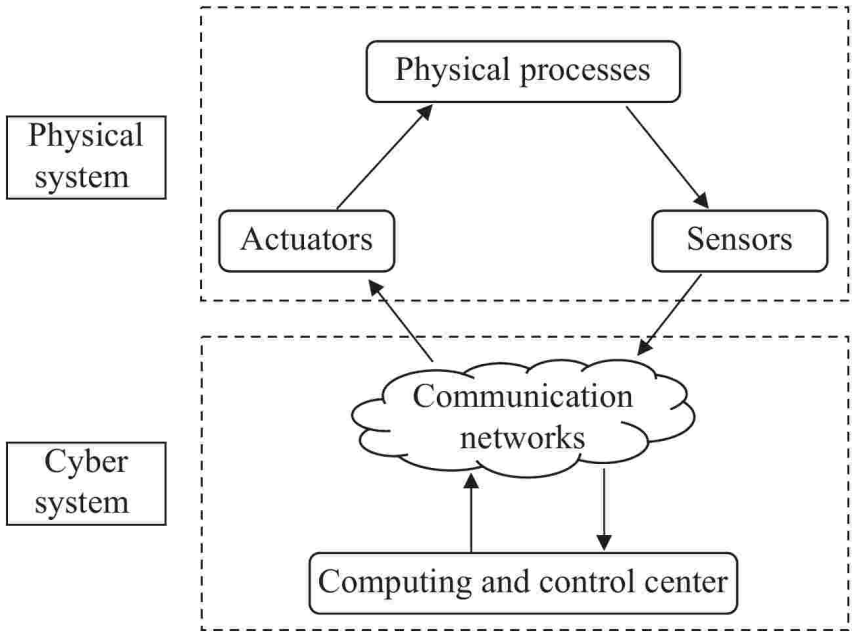
Background and Motivation

Cyber Physical Systems (CPSs)

- Integrate the ability of computing and communications
- Combine cyber and physical components
- Network at multiple scales / high degrees of automation



(a) CPS components



(b) CPS architecture

Background and Motivation

Advantages of Cyber Physical Systems

- Interconnected subsystems form a bigger system
- Communication networks transmit large amounts of data
- Ample computational resources complete real-time analysis and decision-making

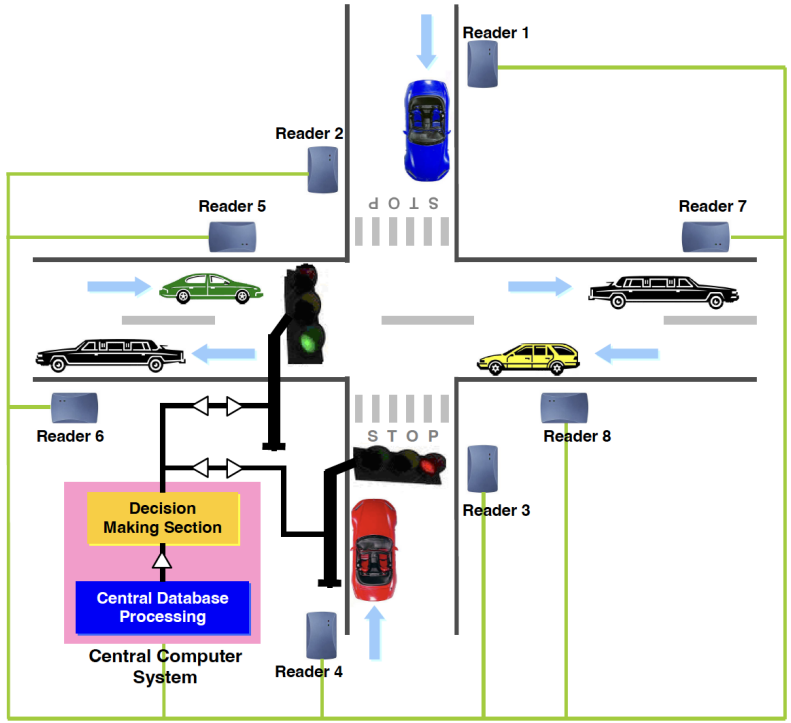
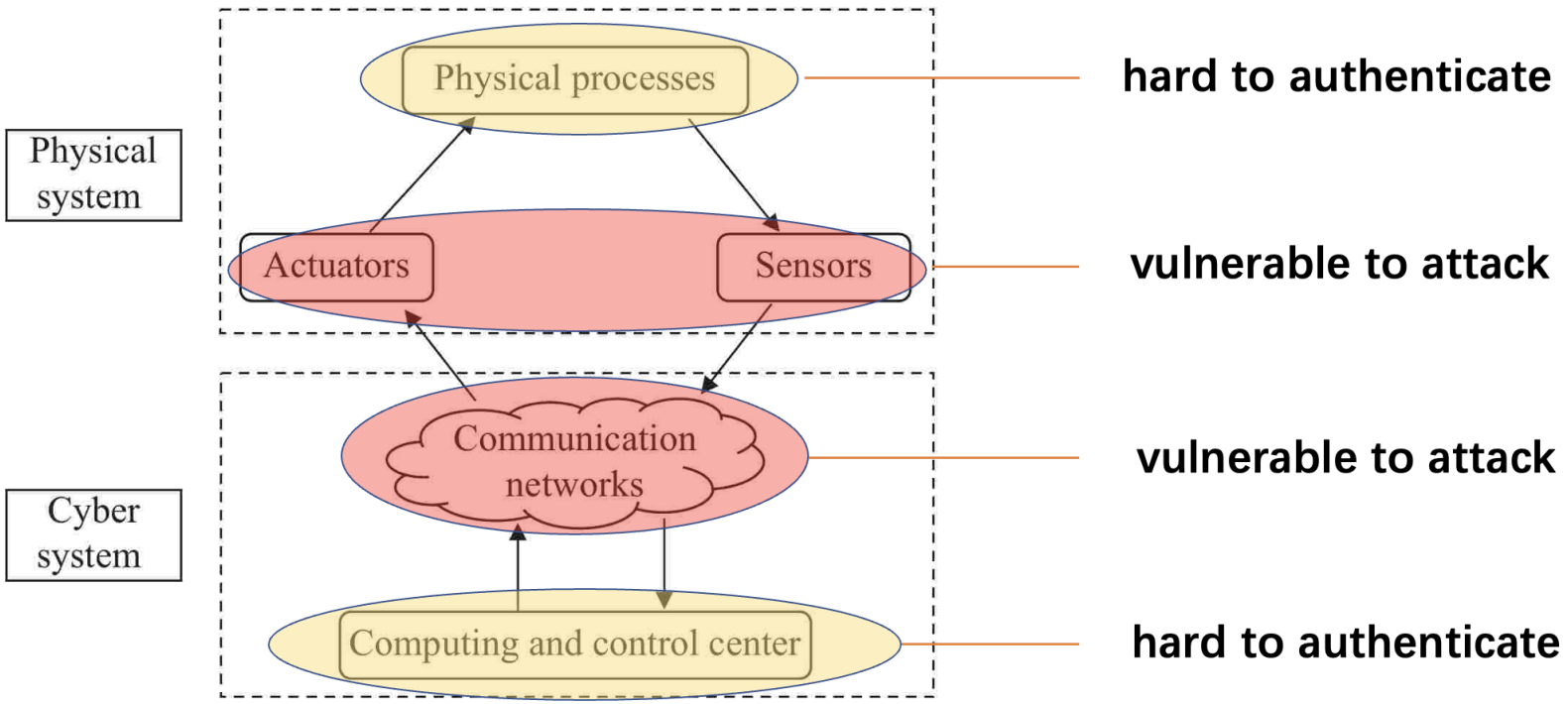


Figure: traffic light scenario

Background and Motivation

The **security** issue is one of the main **challenges** in CPSs.

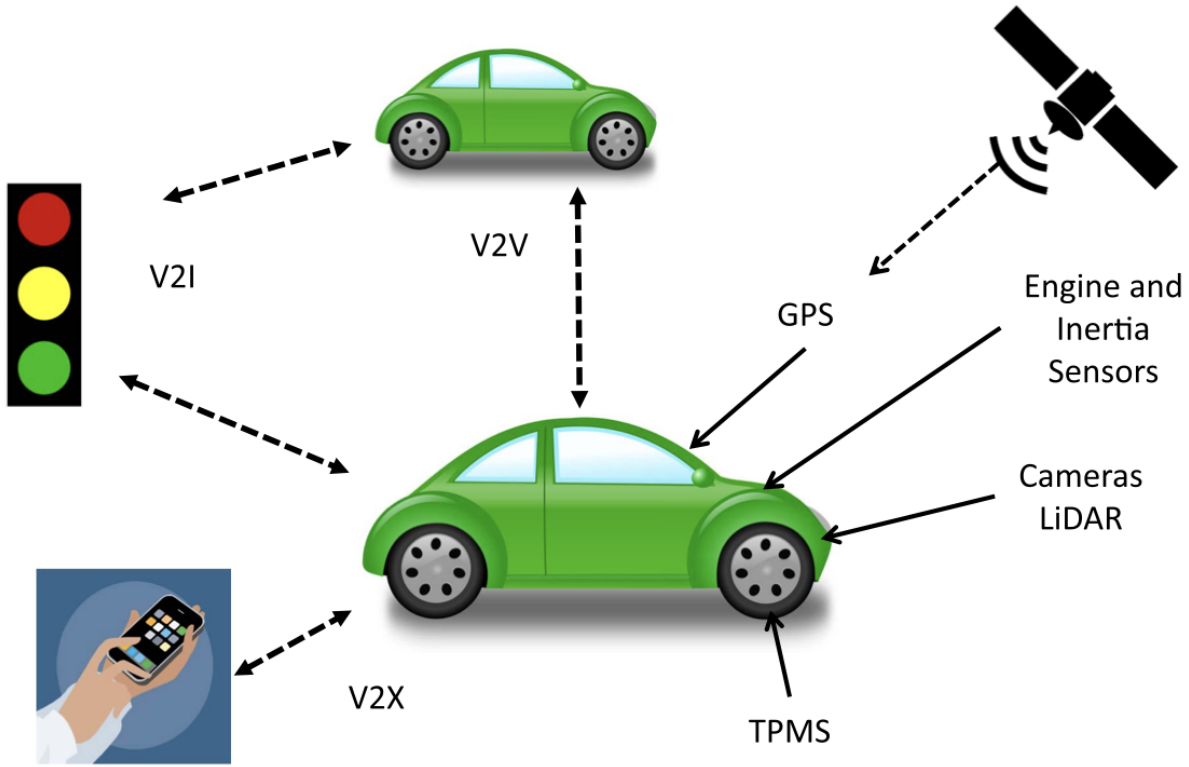
- Massive sensors and actuators work in open environments
- Sensor data and actuation commands are transmitted using relatively open networks
- Controller and physical entities are hard to authenticate the data



Background and Motivation

An example:

- Cyber and physical threats in autonomous driving



Background and Motivation

How to ensure the safety and security of CPSs?

countermeasures in some areas

- protect the components individually
- communication networks: protocol / cryptography
- actuators and sensors: mechanical design /physical protection

countermeasures based on system theory

- a closed-loop control system
- any attack will affects the system state
- detect attacks by analyzing the system state

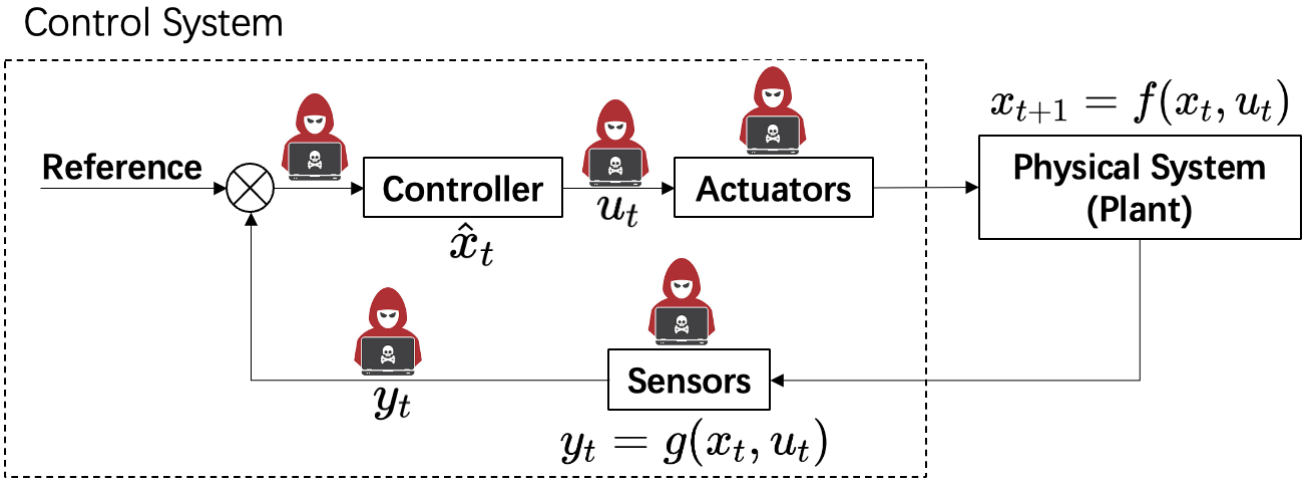


Table of Contents

System Model

Linear CPS model

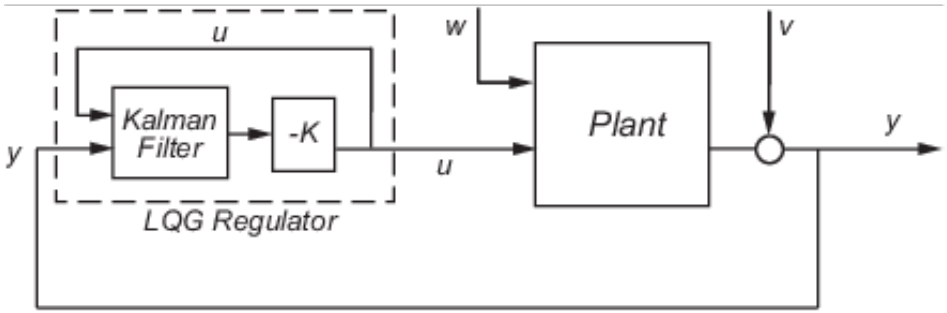
In most literature, the Cyber Physical System is often modeled as a linear discrete-time stochastic system in state-space form:

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + w_t \\ y_t &= Cx_t + v_t \end{aligned} \tag{1}$$

where

- system state $x_t \in \mathbb{R}^n$, system input $u_t \in \mathbb{R}^m$, sensor output $y_t \in \mathbb{R}^p$
- the pair (A, B) is controllable, and the pair (A, C) is observable
- the process noise $\{w_t \in \mathbb{R}^n\}$ has (i.i.d) distribution $\mathbb{N}(0, \Sigma_w)$
- the sensor noise $\{v_t \in \mathbb{R}^p\}$ has (i.i.d) distribution $\mathbb{N}(0, \Sigma_v)$
- $\{w_t\}$ is independent of $\{v_t\}$

As a linear time-invariant model with Gaussian noise, the most effective control approach is to design proper **Kalman filter** and **LQG controller**.



Kalman filter

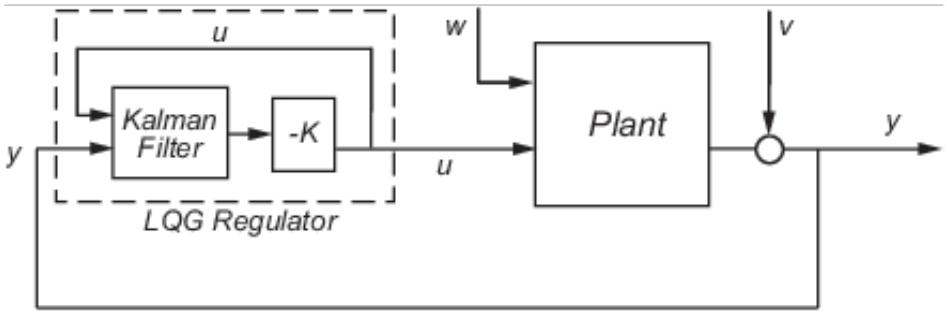
- MMSE (minimum mean-square error) estimate of the states

$$\widehat{x}_t^{MMSE} = E[x_t|x_0, y_1, ..., y_t] \tag{2}$$

- If (A, C) is observable and $(A, \sqrt{\Sigma_w})$ is controllable, the Kalman filter converges to a fixed gain estimator K

$$\begin{aligned} \widehat{x}_t &= \widehat{x}_{t|t-1} + K(y_t - C\widehat{x}_{t|t-1}) \\ \widehat{x}_{t+1|t} &= A\widehat{x}_t + Bu_t \end{aligned} \tag{3}$$

Following the separation principle, the optimal Kalman gain and controller gain can be obtained by solving algebraic Riccati equations, respectively.



LQG (linear–quadratic–Gaussian) controller

- minimize the cost function $\mathbb{J}$

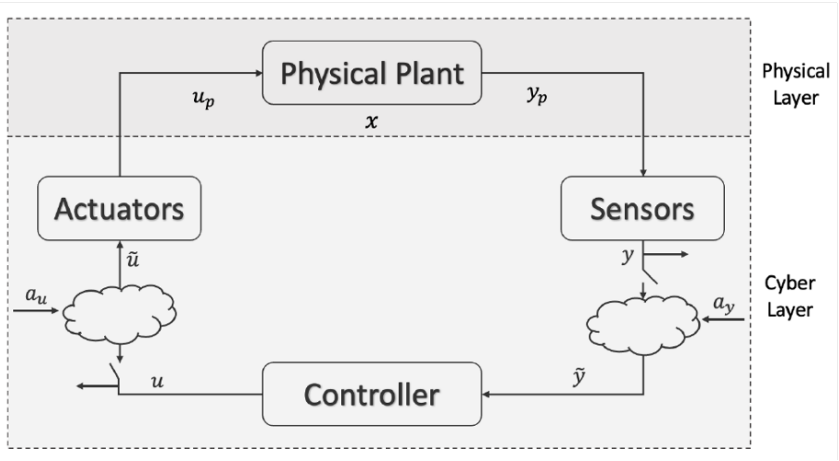
$$\mathbb{J} = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1}(x_t^T Q x_t + u_t^R U u_t)\right] \tag{4}$$

- When (A, B) is controllable and $(A, \sqrt{Q})$ is observable, the optimal linear feedback gain converges to a time-invariant matrix L:

$$u_t = L\widehat{x}_t \tag{5}$$

System Model

Type of CPS attacks



Disclosure attacks

- observe control and sensory data over the communication channels
- disclose the information of system state $\{x_t\}$, which is private

Deception attacks

- happen to sensory data and control input
- inject the false data to sensory data $\tilde{y}_t = y_t + a_t$, like spoofing attacks

Disruption attacks

- aim to compromise availability of critical information
- cause sensory data $y_t = 0$ or control inputs $u_t = 0$ at some point

Lemma (Kalman innovation)

$\{r_t | r_t = y_t - C\widehat{x}_{t|t-1}\}$ is (i,i,d) distribution $\mathbb{N}(0, \Sigma_r)$,
where $\Sigma_r = CPC^T + \Sigma_v$ and $P = APA^T + \Sigma_w - APC^T(CPC^T + \Sigma_v)^{-1}CPA^T$.

Proof

Let's denote $e_t = x_t - \widehat{x}_{t|t-1}$ and $\Gamma_k = \mathbb{E}[r_t r_{t-k}^T]$. When it converges to a fixed gain estimator, $K = PC^T(CPC^T + \Sigma_v)^{-1}$ and $\mathbb{E}[e_t e_t^T] = P$.
Derive the dynamic system of e_t and iterate k times,

$$\begin{aligned} e_t &= A(I - KC)e_{t-1} - AKv_{t-1} + w_{t-1} \\ &= A(I - KC)^k e_{t-k} - \textstyle\sum_{j=1}^{k}[A(I - KC)]^{j-1}AKv_{t-j} + \sum_{j=1}^{k}[A(I - KC)]^{j-1}w_{t-j} \end{aligned} \tag{6}$$

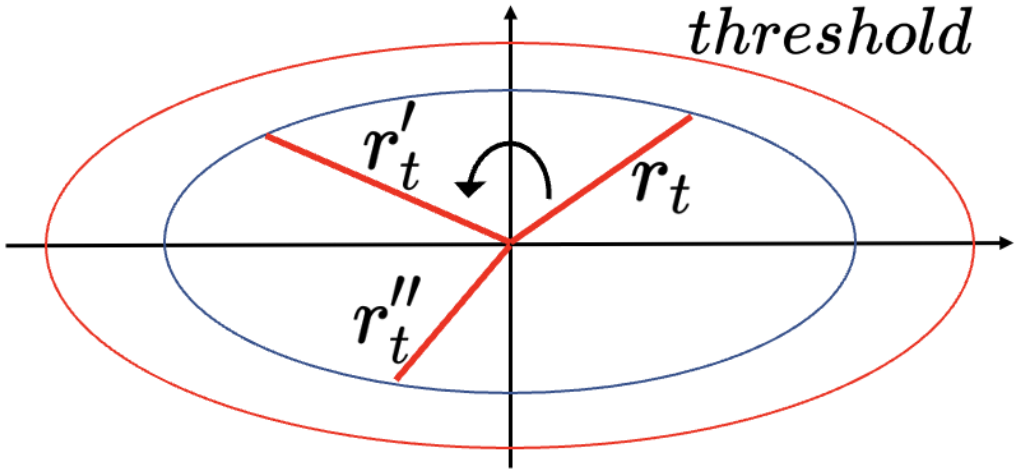
When $k = 0$, $\Gamma_0 = CPC^T + \Sigma_v$. For $k > 0$, use (6) and calculate the expected value:

$$\begin{aligned} \Gamma_k &= \mathbb{E}[(Ce_t + v_t)(Ce_{t-k} + v_{t-k})^T] = C\mathbb{E}[e_t e_{t-k}^T]C^T + C\mathbb{E}[e_t v_{t-k}^T] \\ &= C[A(I - KC)]^k PC^T - C[A(I - KC)]^{k-1}AK\Sigma_v \\ &= C[A(I - KC)]^{k-1}A[PC - K(CPC + \Sigma_v)] = 0 \end{aligned} \tag{7}$$

χ^2 detector[1]

$$g_t = \sum_{i=t-\tau+1}^{t} r_i^T \Sigma_r^{-1} r_i \gtrless threshold$$

- The higher the value of g_t, the lower the probability of the occurrence of the $\{r_t\}$ in normal case, the more likely the system to be attacked.
- The disadvantage is that only consider the length of r_i.

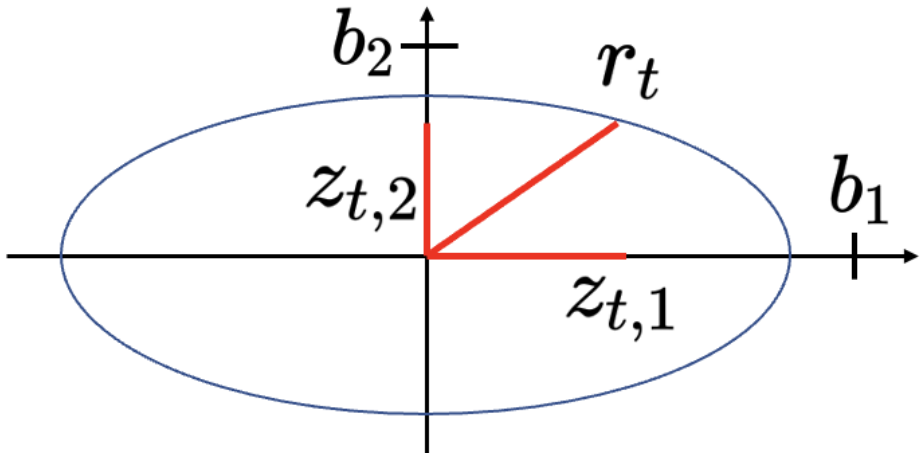


[1] Yilin Mo, Rohan Chabukswar, and Bruno Sinopoli. "Detecting integrity attacks on SCADA systems". In: *IEEE Transactions on Control Systems Technology* 22.4 (2013), pp. 1396–1407.

CUSUM detector[2]

$$S_{t,i} = max(0, S_{t-1,i} + z_{t,i} - b_i) \gtrless threshold$$

- denote $r_{t,i}$ as the i_{th} entry of r_t, which has the distribution $\mathcal{N}(0, \sigma_i^2)$
- denote $z_{t,i} := |r_{t,i}|$ as the absolute value, which follows the distribution $\mathcal{N}(\frac{\sqrt{2}}{\sqrt{\pi}}\sigma_i, \sigma_i^2(1 - \frac{2}{\pi}))$
- design the bias $b_i > 0$ and start with $S_{1,i} = 0$



[2] Carlos Murguia and Justin Ruths. "Characterization of a cusum model-based sensor attack detector". In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 2016, pp. 1303–1309.

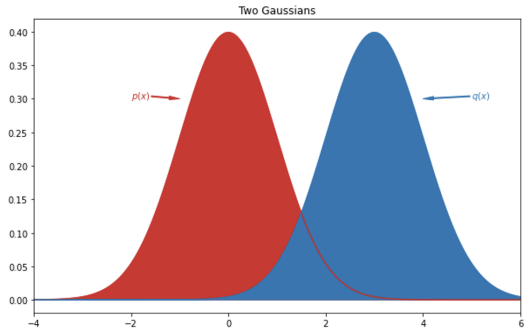
KLD detector[3]

$$\mathbb{D}_{KL}(r_t^a \parallel r_t^0) \gtrless \textit{threshold}$$

- Kullback–Leibler divergence mearsure how one probability distribution P is different from a reference probability distribution Q:

$$\mathbb{D}_{KL}(P \parallel Q) = \int_{\chi} p(x) log\left(\frac{p(x)}{q(x)}\right) \mu(x)$$

- In general, the lower the KL divergence value, the closer the two distributions are to one another.



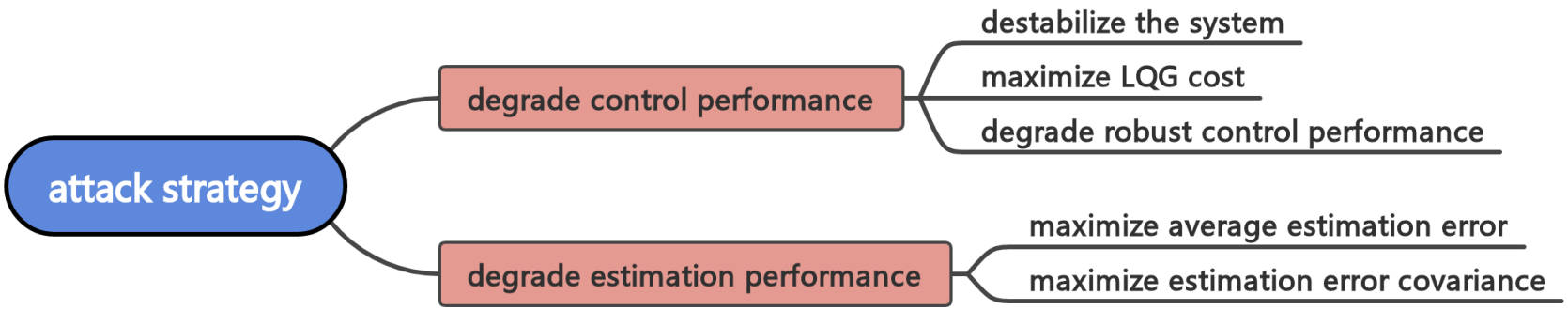
[3] Cheng-Zong Bai, Fabio Pasqualetti, and Vijay Gupta. "Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs". In: *Automatica* 82 (2017), pp. 251–260.

Table of Contents

How to design the attack strategy?

Attacks in Cyber Physical Systems

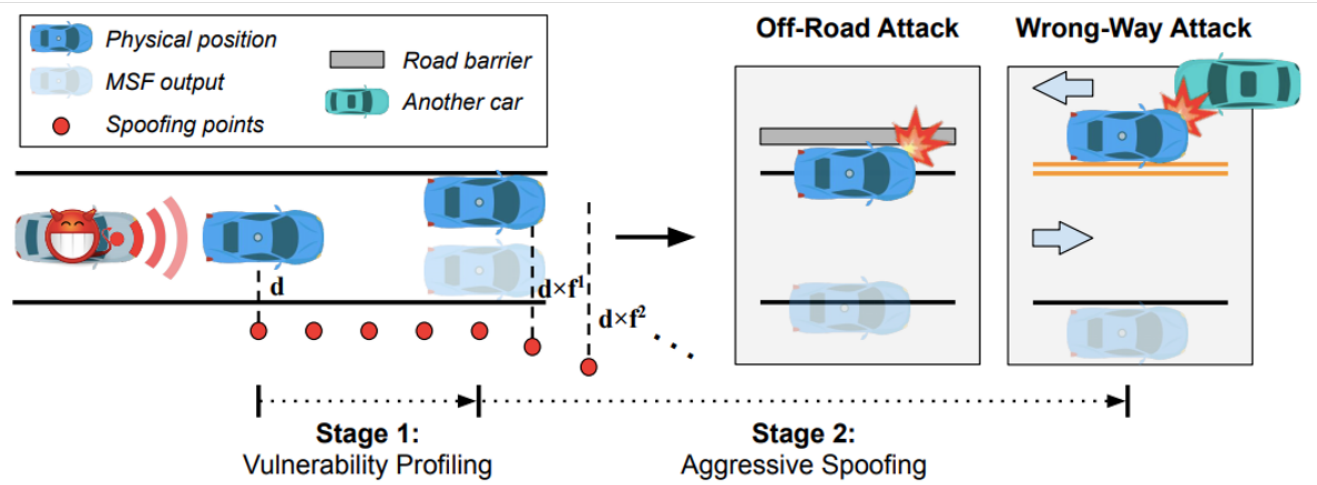
- goal
  - degrade the control performance / estimation performance
- strength
  - have sophisticated analysis tools / modeling approaches
- weakness
  - focus on the impact on system performance rather than system state
  - ignore the time limits and sacrifice efficiency for stealthness



How to design the attack strategy?

Attacks in Autonomous Driving

- goal
  - induce dramatic changes in system state in a short period of time
- strength
  - focus on feasibility / practical results in real-world scenarios
- weakness
  - have difficulties to quantify the impact of the attack on system state
  - Countermeasures are not seriously considered when designing

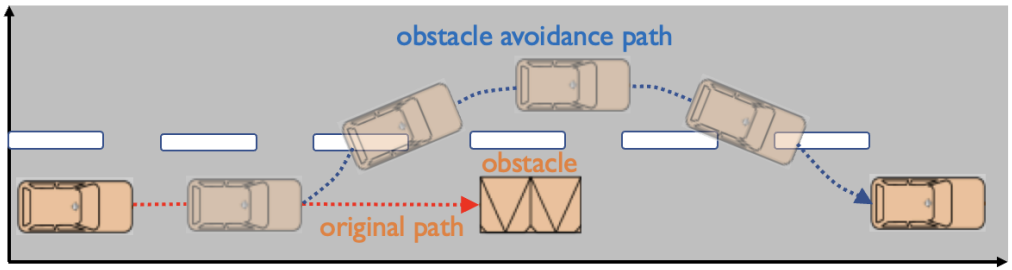


My Research Topic

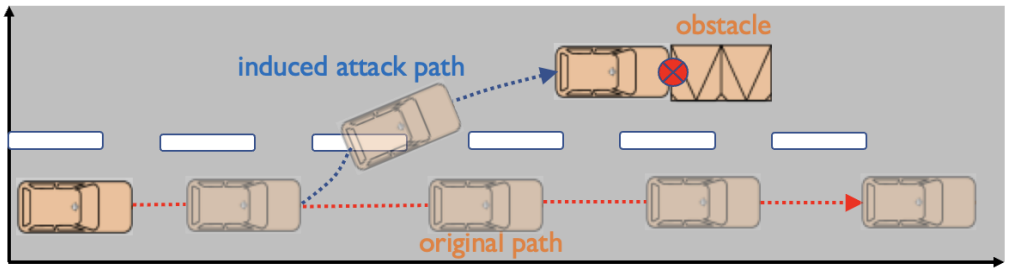
My research goal

- design an online sensor attack strategy
- move the system state to a target state
- keep stealthy under KLD detection
- limit attacks to occur within a fixed short period of time

Compare with obstacle avoidance planning



(a) obstacle avoidance planning



(b) induced attack planning

The system dynamics under attack

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + w_t \\ y_t &= Cx_t + E\xi_{t-1} + v_t \end{aligned} \tag{8}$$

where ξ_t is the attack in the sensors.

- CPS begins running at $t = -\infty$, already has fixed filter gain K and feedback gain L
- attacker monitors from $t = -\infty$, and deploys attack at $t = 0$ lasting until $t = N$
- attacker knows the full information of the system model, including:
  - $A, B, C, E, L, K, \Sigma_w, \Sigma_v$
  - system's state estimate $\hat{x}_t$ and the control input $u_t = L\hat{x}_t$
- the attacker has information set

$$\mathcal{I}_0 = \{y_0\} \ , \ \mathcal{I}_{t+1} = \{\mathcal{I}_t, y_{t+1}, \xi_t\}$$

- the attacker's aim is to find an attack policy

$$\xi_t = \mathcal{F}(\mathcal{I}_t)$$

Problem 1.

$$\begin{aligned} \min_{\mathcal{F}(\mathcal{I}_0),..,\mathcal{F}(\mathcal{I}_N)} \quad & J = \mathbb{E}\left\{\sum_{t=1}^{N+1} \|x_t - x^*\|_{Q_t}^2\right\} \\ \text{s.t.} \quad & \mathbb{D}_{KL}(r_t^a \parallel r_t^0) \leq \delta, t = 1, ..., N+1 \end{aligned} \tag{9}$$

Variable table

	attacker	CPS		
		without attack	under attack	bias
state estimate	$\widetilde{x}_t$	$\hat{x}_t^0$	$\hat{x}_t^a$	
estimation error	$\widetilde{e}_t$	$\hat{e}_t^0 = \widetilde{e}_t$	$\hat{e}_t^a$	Δe_t
Kalman innovation	$\widetilde{r}_t$	$\hat{r}_t^0 = \widetilde{r}_t$	$\hat{r}_t^a$	

An attack sequence $\{\xi_0, ..., \xi_N\}$ will induce a bias to two critical variables

- system's real state $\{x_1, ..., x_N\}$
- CPS's Kalman innovation $\{\hat{r}_1^a, ..., \hat{r}_N^a\}$

Derive attacker's state estimate $\widetilde{x}_t$ under attack

$$\begin{aligned} r^a_{t+1} &= y_{t+1} - C\widehat{x}^a_{t+1} \\ &= CAe_t + E\xi_t + Cw_k + v_{t+1} \\ &= CA\Delta e_t + E\xi_t + \widetilde{r}_{t+1} \\ \Delta e_{t+1} &= (A - KCA)\Delta e_t - KE\xi_t \end{aligned}$$

Define the virtual state $\theta_t = [\widetilde{x}_t^T \ \widehat{x}_t^{aT} \ \Delta e_t^T]^T$, and obtain the dynamic system

$$\theta_{t+1} = \mathcal{A}\theta_t + \mathcal{B}\xi_t + \mathcal{H}\widetilde{r}_{t+1}, \tag{10}$$

where

$$\mathcal{A} = \begin{bmatrix} A & BL & 0 \\ 0 & A+BL & KCA \\ 0 & 0 & A-KCA \end{bmatrix}, \mathcal{B} = \begin{bmatrix} 0 \\ KE \\ -KE \end{bmatrix}, \mathcal{H} = \begin{bmatrix} K \\ K \\ 0 \end{bmatrix}$$

and $\Delta e_0 = \widetilde{x}_0 - \widehat{x}^a_0 = 0$.

Detection constraint

From previous knowledge, we know $\{r_t^0\}$ has (i.i.d) distribution $\mathbb{N}(0, \Sigma_r)$. Given the condition $\mathcal{I}_k$, we can derive the distribution of $\{r_t^a | t > k\}$:

$$r_t^a | \mathcal{I}_k \sim \mathbb{N}(\beta_{k,t}, \Sigma_{k,t}) \tag{11}$$

where

$$\beta_{k,t} = CA(A - KCA)^{t-k-1} \Delta e_k + E\xi_{t-1} - \sum_{j=k}^{t-2} CA(A - KCA)^{t-2-j} KE\xi_j$$

and $\Sigma_{k,t} = \Sigma_r$. Hence, we calculate the KL divergence in r_t^a and r_t^0,

$$\begin{aligned} \mathbb{D}_{KL}(r_t^a || r_t^0 \ | \mathcal{I}_k) &= \frac{1}{2} \left[Tr(\Sigma_r^{-1} \Sigma_{k,t}) + log \frac{|\Sigma_r|}{|\Sigma_{k,t}|} - p + \beta_{k,t}^T \Sigma_r^{-1} \beta_{k,t} \right] \\ &= \frac{1}{2} \beta_{k,t}^T \Sigma_r^{-1} \beta_{k,t} \leq \delta, \quad t = k+1, ..., N+1 \end{aligned} \tag{12}$$

Reformulate the original problem

From the knowledge of Kalman filter, the attacker's estimation error $\widetilde{e}_t \sim \mathbb{N}(0, (I - KC)P)$ is orthogonal to $\widetilde{x}_t$, where $\mathbb{E}[(x_t - \widetilde{x}_t)^T \widetilde{x}_t] = 0$.

$$J = \mathbb{E}\left\{\sum_{t=1}^{N+1} \|x_t - x^*\|_{Q_t}^2\right\} \Longleftrightarrow J = \mathbb{E}\left\{\sum_{t=1}^{N+1} [Tr((I_n - KC)PQ_t) + \|\widetilde{x}_t - x^*\|_{Q_t}^2]\right\}$$

Problem 2.

$$\begin{aligned} \min_{\mathcal{F}(\mathcal{I}_0),..,\mathcal{F}(\mathcal{I}_N)} \quad & \bar{J} = \mathbb{E}\left\{\sum_{t=1}^{N+1} \|F_1 \bar{\theta}_t\|_{Q_t}^2\right\} \\ \text{s.t.} \quad & \bar{\theta}_{t+1} = \bar{\mathcal{A}}\bar{\theta}_t + \bar{\mathcal{B}}\xi_t + \bar{\mathcal{H}}\widetilde{r}_{t+1}, \quad t = 0, ..., N \\ & \beta_{k,t}^T \Sigma_r^{-1} \beta_{k,t} \leq \delta, \quad k = 1, ..., N, t = k+1, ..., N+1 \\ & \beta_{k,t} = CA(A - KCA)^{t-k-1} F_2 \bar{\theta}_k + E\xi_{t-1} - \sum_{j=k}^{t-2} CA(A - KCA)^{t-2-j} KE\xi_j \end{aligned}$$

where add the target state x^* to the virtual state $\bar{\theta}_t = [\widetilde{x}_t^T \ \widehat{x}_t^{a\,T} \ \Delta e_t^T \ (x^*)^T]^T$, and $F_1 = [I_n \ 0 \ 0 \ -I_n], F_2 = [0 \ 0 \ I_n \ 0]$.

We discuss the attack strategy at time k

- the attacker has deployed the attack signals $\{\xi_0, ..., \xi_{k-1}\}$
- the attacker has known the $\{\widetilde{r}_0, ..., \widetilde{r}_k\}$ and can calculate $\bar{\theta}_k$
- the attacker aims to design an optimal $\{\xi_k, ..., \xi_N\}$

Define $\{\bar{\theta}_{k+1}, ..., \bar{\theta}_{N+1}\}$ as a vector Θ_{k+1}^{N+1}, $\{\xi_k, ..., \xi_N\}$ as Ξ_k^N, $\{\widetilde{r}_{k+1}, ..., \widetilde{r}_{N+1}\}$ as $\widetilde{R}_{k+1}^{N+1}$, $\{\beta_{k,k+1}, ..., \beta_{k,N+1}\}$ as Γ_{k+1}^{N+1}, and then we obtain

$$\Theta_{k+1}^{N+1} = \mathbb{A}_k \bar{\theta}_k + \mathbb{B}_k \Xi_k^N + \mathbb{H}_k \widetilde{R}_{k+1}^{N+1}$$
$$\Gamma_{k+1}^{N+1} = \mathbb{C}_k \bar{\theta}_k + \mathbb{D}_k \Xi_k^N \tag{13}$$

where

$$\mathbb{A}_k = \begin{bmatrix} \bar{\mathcal{A}} \\ \bar{\mathcal{A}}^2 \\ \vdots \\ \bar{\mathcal{A}}^{N+1-k} \end{bmatrix}, \mathbb{B}_k = \begin{bmatrix} \bar{\mathcal{B}} & 0 & \cdots & 0 \\ \bar{\mathcal{A}}\bar{\mathcal{B}} & \bar{\mathcal{B}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathcal{A}}^{N-k}\bar{\mathcal{B}} & \bar{\mathcal{A}}^{N-k-1}\bar{\mathcal{B}} & \cdots & \bar{\mathcal{B}} \end{bmatrix}, \mathbb{H}_k = \begin{bmatrix} \bar{\mathcal{H}} & 0 & \cdots & 0 \\ \bar{\mathcal{A}}\bar{\mathcal{H}} & \bar{\mathcal{H}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathcal{A}}^{N-k}\bar{\mathcal{H}} & \bar{\mathcal{A}}^{N-k-1}\bar{\mathcal{H}} & \cdots & \bar{\mathcal{H}} \end{bmatrix},$$

$$\mathbb{C}_k = \begin{bmatrix} CAF_2 \\ CA(A-KCA)F_2 \\ \vdots \\ CA(A-KCA)^{N-k}F_2 \end{bmatrix}, \mathbb{D}_k = \begin{bmatrix} E & \cdots & 0 & 0 \\ \vdots & \ddots & \cdots & \cdots \\ -CA(A-KCA)^{N-2-k}KE & \cdots & E & 0 \\ -CA(A-KCA)^{N-1-k}KE & \cdots & -CAKE & E \end{bmatrix}$$

Find an optimal attack strategy at time $k \Longleftrightarrow$ Solve a QCQP problem

Problem 3.

$$\begin{aligned} \min_{\Xi_k^N} \quad & \bar{J}_k = 2\bar{\theta}_k^T \mathbb{A}_k^T \mathbb{Q}_k \mathbb{B}_k \Xi_k^N + \left\| \mathbb{B}_k \Xi_k^N \right\|_{\mathbb{Q}_k}^2 \\ \text{s.t.} \quad & \left\| \mathbb{C}_k \bar{\theta}_k + \mathbb{D}_k \Xi_k^N \right\|_{\mathbb{S}_i}^2 \leq \delta, \ i = 1, .., N-k \end{aligned} \tag{14}$$

where

$$\mathbb{Q}_k = diag(F_1^T Q_{k+1} F_1, ..., F_1^T Q_{N+1} F_1)$$

$$\mathbb{S}_i = diag(0, ..., 0, \Sigma_r^{-1}{}_{-i_{th}}, 0, ..., 0), \ i = 1, .., N-k$$

The algorithm gives the optimal solution $\xi_t = \mathcal{F}(\mathcal{I}_t)$ to problem 1

Algorithm 1 Optimal attack policy $\mathcal{F}(\mathcal{I}_k)$

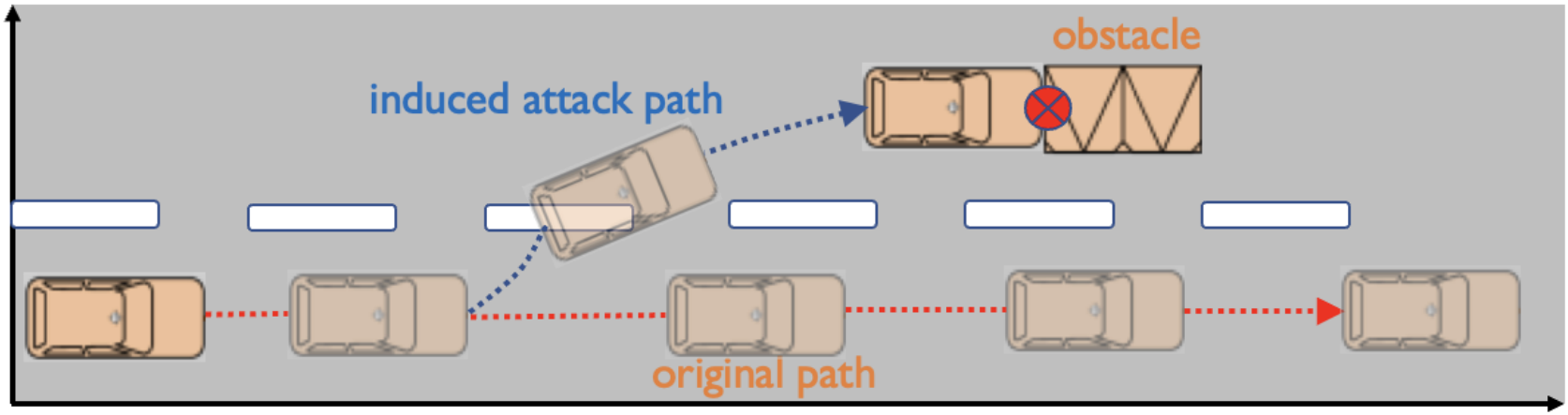
1: $\mathcal{I}_0 = \{u_{-\infty}, .., u_{-1}, y_{-\infty}, ..., y_0\}$, $\xi_k = 0$ and $k = 0$
2: **while** $1 \leq k \leq N$ **do**
3: Attack the sensor with optimal attack ξ_{k-1}
4: Measure the output y_k and update $\mathcal{I}_k = \{\mathcal{I}_{k-1}, u_{k-1}, y_k, \xi_{k-1}\}$
5: Calculate the θ_k from $\mathcal{F}(\mathcal{I}_k)$ and find optimal Ξ_k^N by solving problem 3
6: $k \leftarrow k+1$
7: **end while**

My Research Topic

Experimental result

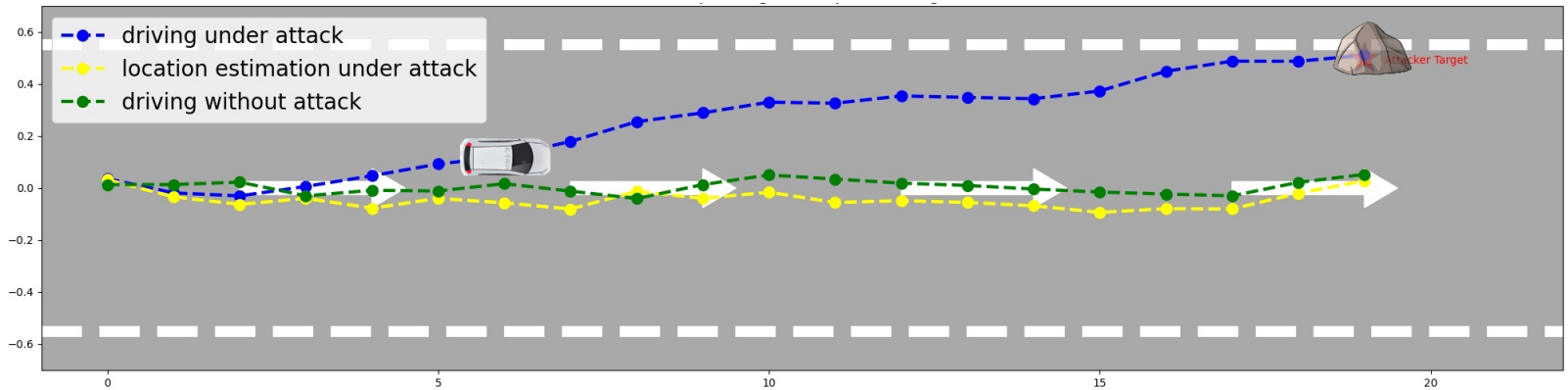
Simulate **the lane following task** in autonomous driving

- the autonomous car keeps driving in the middle of the lane line
- in the horizontal direction, the car maintains a constant speed
- in the vertical direction, the car controls the distance from the midline
- the car uses GPS as the only location observer, and equips a χ^2 detector
- the attacker implements GPS spoofing attack

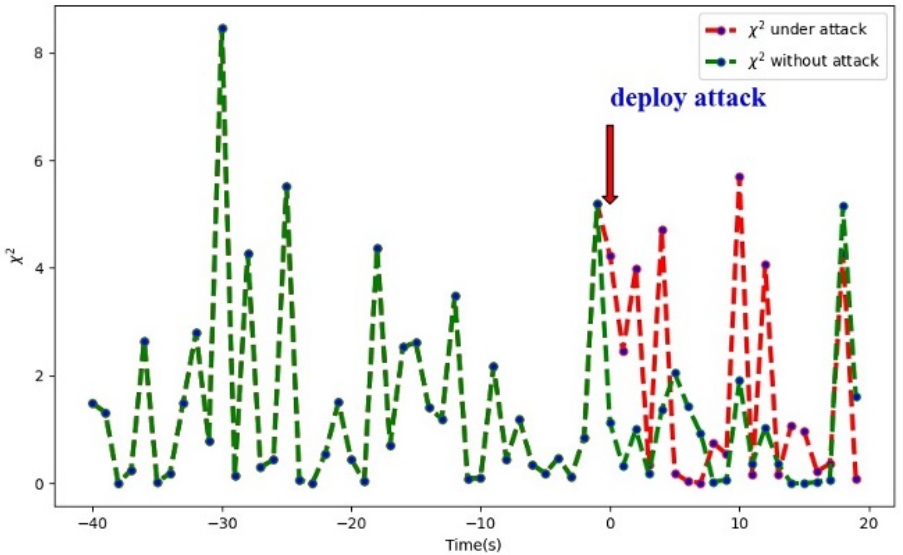


Experiment 1

- GPS spoofing attack against traditional χ^2 detector

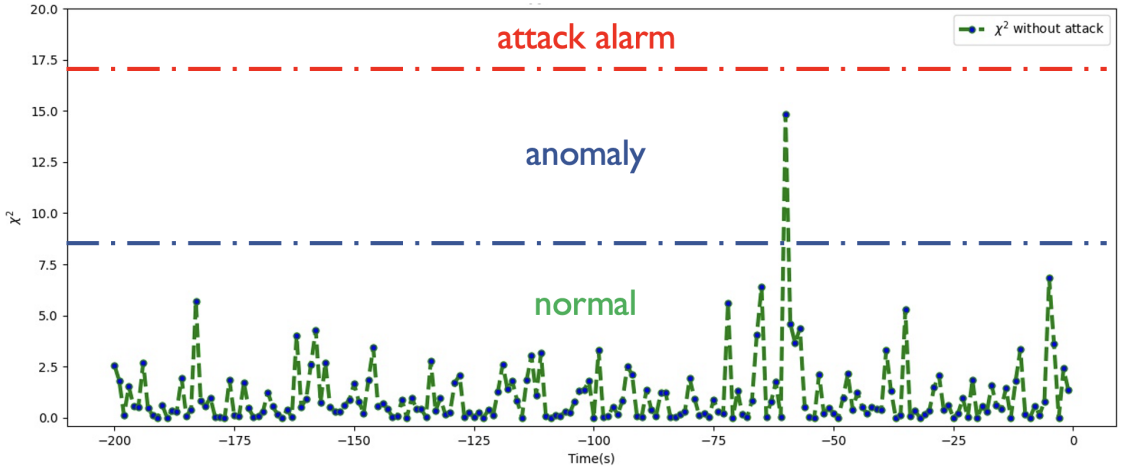


- traditional χ^2 detector



The detector with exceptions handling

- disadvantages of the traditional χ^2 detector
  - if the threshold of χ^2 detector is large, false negative rate is high
  - if the threshold of χ^2 detector is small, false alarm rate is high

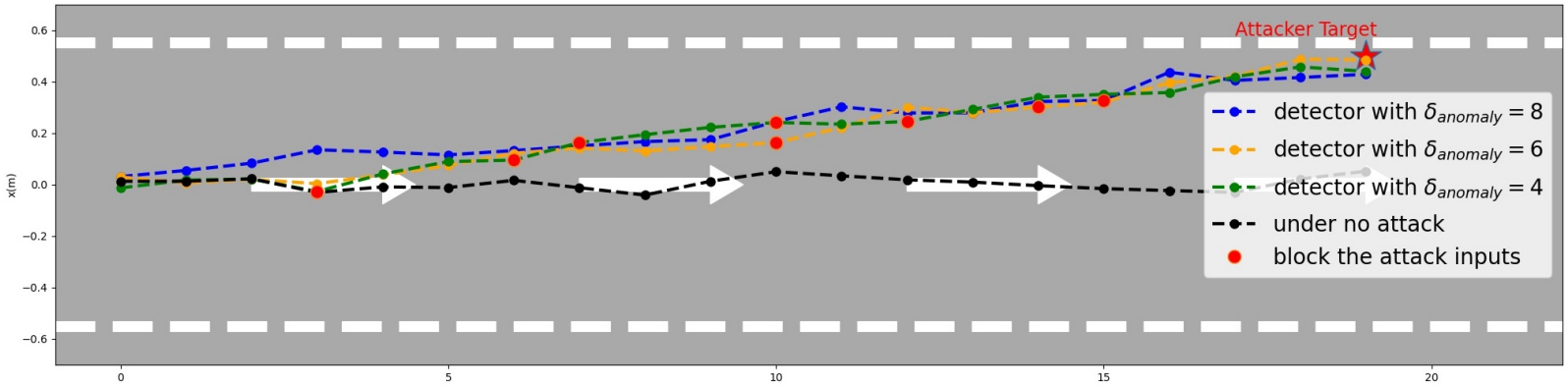


- advanced χ^2 detector

$$\text{detector}(\chi^2) = \begin{cases} \text{update state estimation}, & \chi^2 \leq \delta_{anomaly} \\ \text{ignore this measurement}, & \delta_{anomaly} < \chi^2 \leq \delta_{attack} \\ \text{send an attack warning}, & \delta_{attack} < \chi^2 \end{cases}$$

Experiment 2

- GPS spoofing attack against advanced χ^2 detector



- advanced χ^2 detector

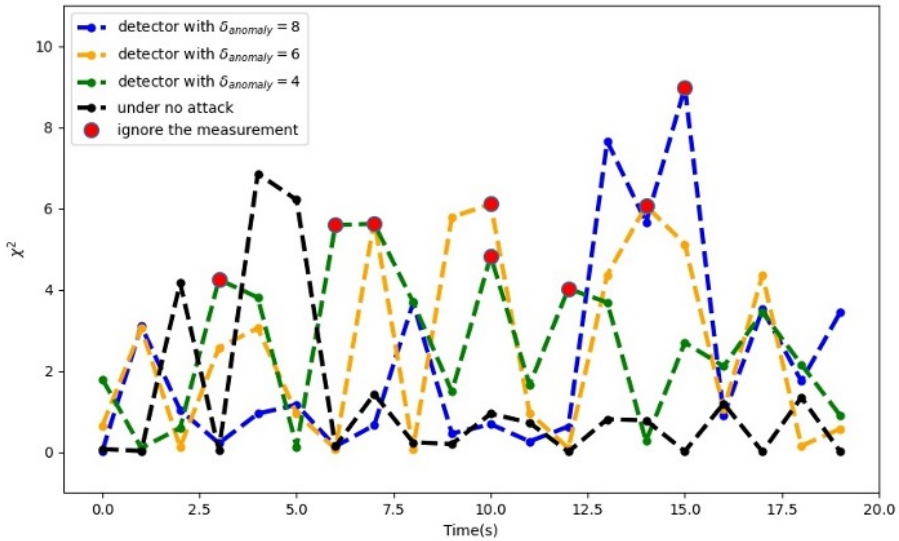
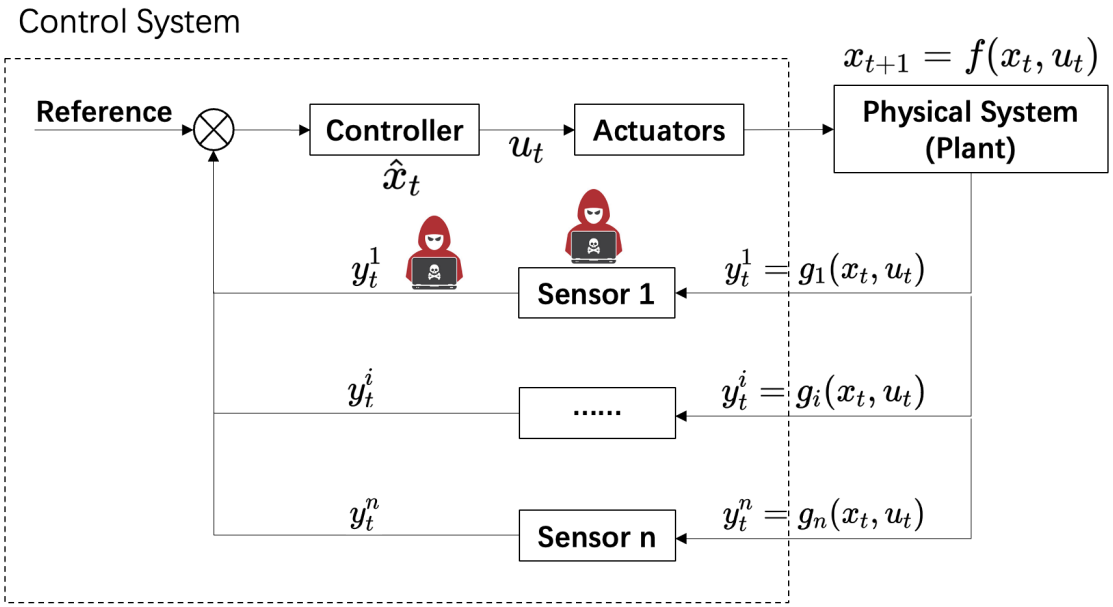


Table of Contents

Future work

Multi-sensor system



Current limitations

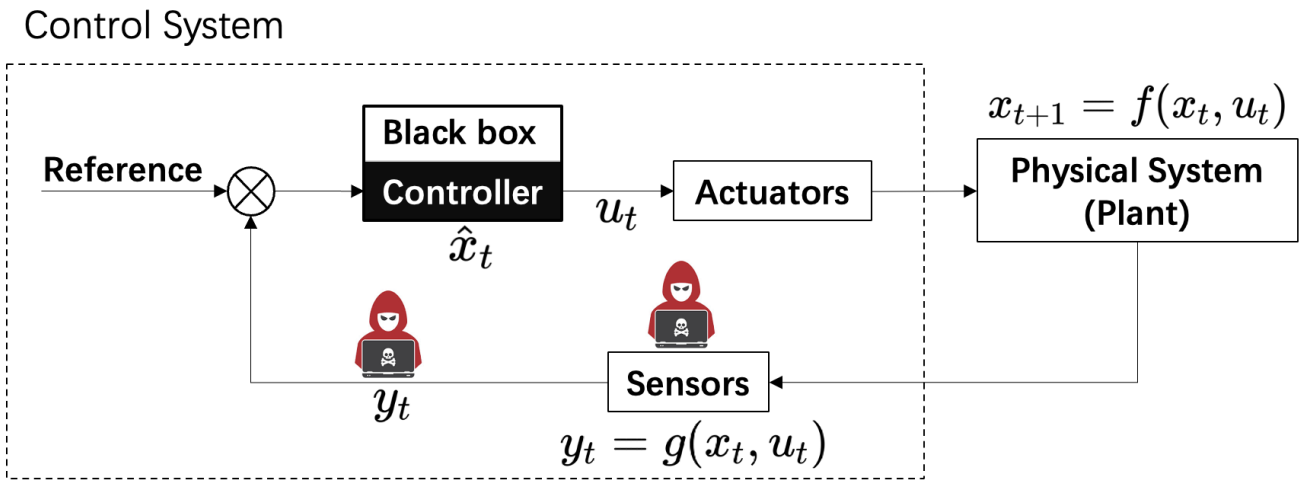
- only consider the system with single sensor
- other functioning sensors will mitigate the attack effect

Potential research approaches

- (attackers) model the impact of attacks on multi-sensor fusion
- (defenders) design new defense mechanisms with multi-sensors

Future work

Attacker without perfect system information



Current limitations

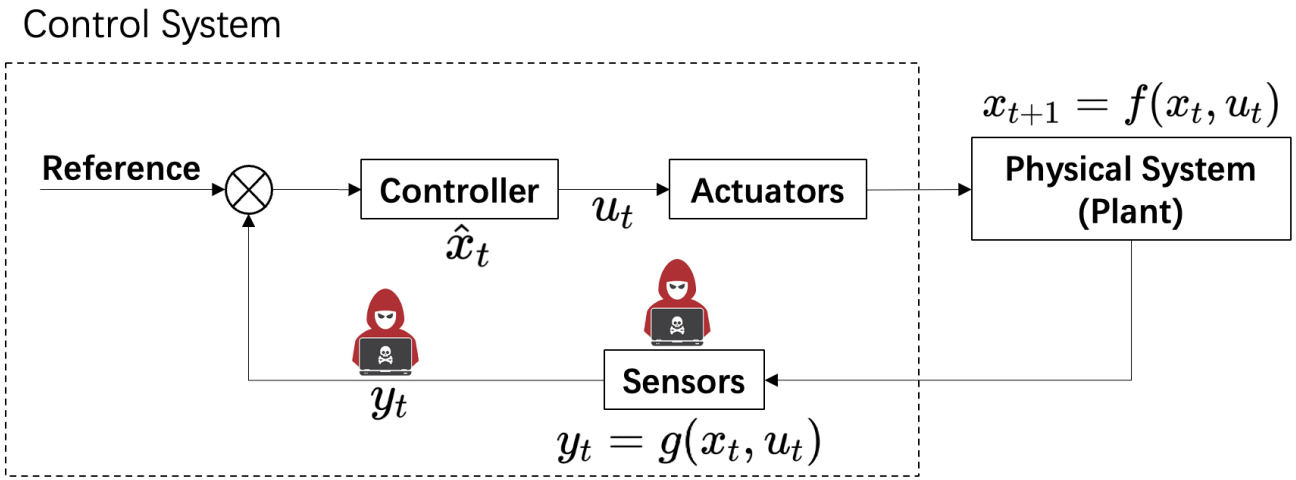
- the optimal attack strategy requires perfect system knowledge
- the distribution of system noise is unknown

Potential research approaches

- (attackers) estimate the control law of the system
- (attackers) design optimal attack strategy with inaccurate system information

Future work

Nonlinear system



Current limitations

- have difficulty in modeling and analyzing nonlinear systems
- no closed-formed expression exists in most cases

Potential research approaches

- (attackers) apply learning-based approaches to design attack strategies
- (attackers and defenders) utilize numerical methods to approximate models and analyze errors

Thank You

Questions

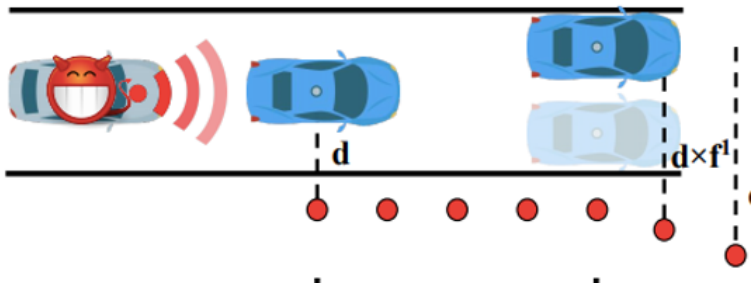
Contributions

- formulate the attack effect
  - system state
  - Kalman innovation under attack
- a stealthy attack with control objectives
  - attacks in autonomous driving: not stealthy
  - attacks in cyber physical system: without control objectives
- an online attack strategy
  - adjust the attack strategy according to the real situation
  - work well even when some attack signals are accidentally blocked

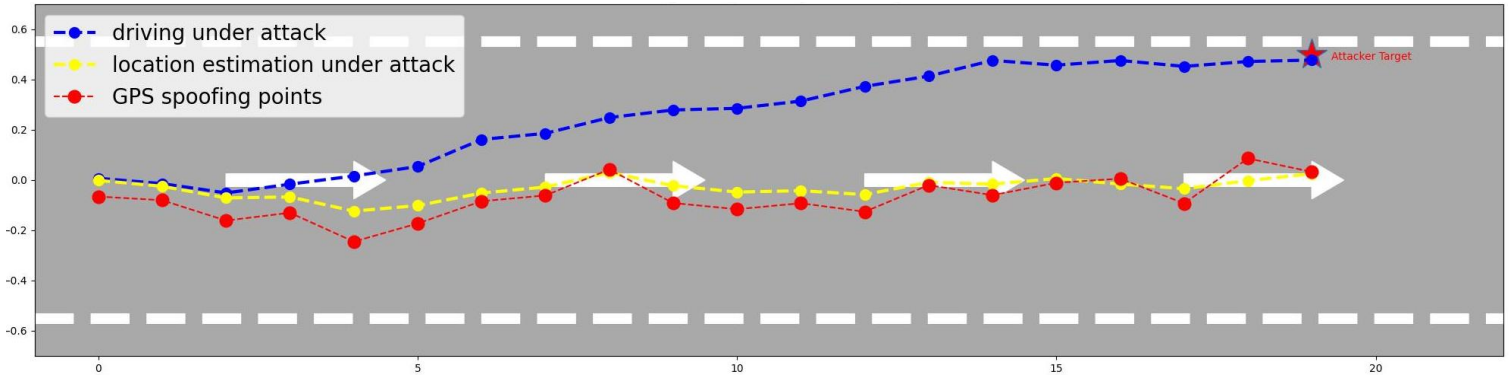
Questions

Compare with related work

- attacks in autonomous driving[1]



- our work



- improvement
  - design an attack sequence rather than one-step attacks
  - stealthy

[1] Junjie Shen et al. "Drift with Devil: Security of {Multi-Sensor} Fusion based Localization in {High-Level} Autonomous Driving under {GPS} Spoofing". In: *29th USENIX Security Symposium (USENIX Security 20)*. 2020, pp. 931–948.

Questions

Compare with related work

- attacks in cyber physical system[2]

$$J = J_c + J_d$$
$$= \mathbb{E}\left\{\sum_{t=0}^{N} \|x_t - x^*\|_{Q_t}^2\right\} + \alpha * \mathbb{E}\left\{\sum_{t=0}^{N} \|r_t\|_{R_t}^2\right\}$$

- our work

$$\begin{aligned} \min \quad & J = \mathbb{E}\left\{\sum_{t=0}^{N} \|x_t - x^*\|_{Q_t}^2\right\} \\ \text{s.t.} \quad & \mathbb{D}_{KL}(r_t^a \parallel r_t^0) \leq \delta, t = 0, ..., N \end{aligned}$$

- improvement
  - no need to choose α
  - decouple the cost function and constraint

[2] Yuan Chen, Soummya Kar, and José MF Moura. "Cyber-physical attacks with control objectives". In: *IEEE Transactions on Automatic Control* 63.5 (2017), pp. 1418–1425.

Questions

Why design attacks?

- raise people's concerns, and increase people's research interest

- as a basic work to understand the system's vulnerability
- provide reasonable verification for new detection mechanisms

Questions

Success rate of the experiment

$$\text{detector}(\chi^2) = \begin{cases} \text{update state estimation}, & \chi^2 \leq \delta_{anomaly} \\ \text{ignore this measurement}, & \delta_{anomaly} < \chi^2 \leq \delta_{attack} \\ \text{send an attack warning}, & \delta_{attack} < \chi^2 \end{cases}$$

- **success rate**: $|x_N - x^*| < 0.05$
- **Detection rate**: $\chi^2 > \delta_{attack} = 15$

	Success rate	Detection rate
$\delta_{anomaly} = 10$	91%	2%
$\delta_{anomaly} = 8$	83%	13%
$\delta_{anomaly} = 6$	62%	33%
$\delta_{anomaly} = 4$	35%	59%

Questions

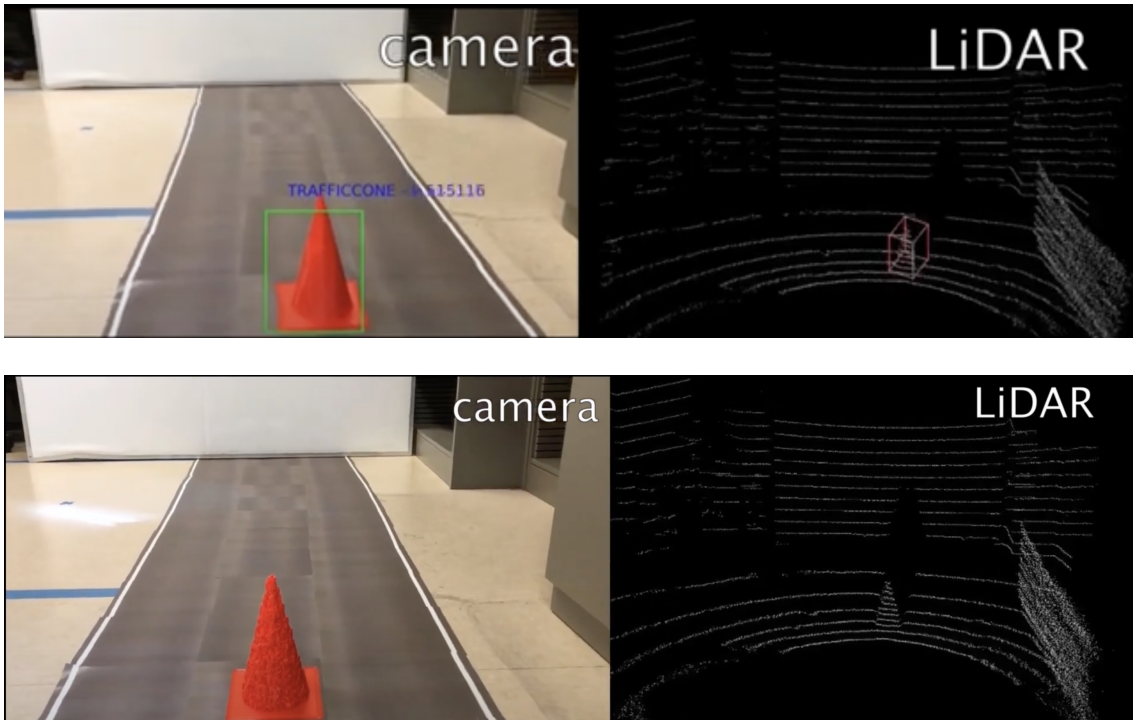
Non-Gaussian noise

- Gaussian noise is very important for theoretical research
- approximate it to Gaussian noise
- other methods, like particle filter

Questions

sensor attacks in autonomous driving

- camera and Lidar



- How they affect the system state?