Secure Detection in Adversarial Environments: the Price of Security

Xiaoqiang Ren

School of EEE
Nanyang Technological University, Singapore

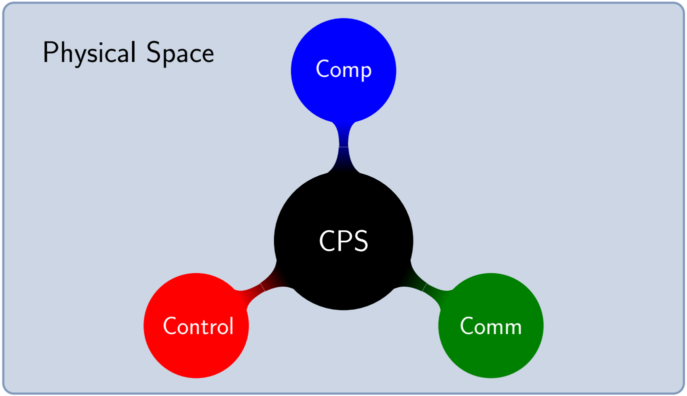
Joint work with Jiaqi Yan and Yilin Mo

Outline

1. Research Background: CPS Security
2. Trade-off Between Efficiency and Security
3. Conclusion

Cyber-Physical System

- Cyber-Physical System (CPS) refers to the embedding of computation, communication and control into physical spaces.



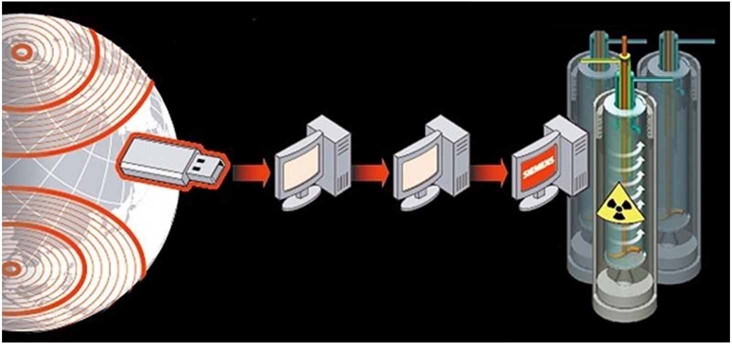
- Applications: aerospace, chemical processes, civil infrastructure, manufacturing, transportation, internet of things.

Security Threats for the CPS

Extensive use of widespread sensing and networking makes the CPSs vulnerable to malicious attacks.

1. Devices have low computation capability
2. Legacy hardware and software: not designed with security in mind
3. Complex interaction between the physical space and cyber space
4. CPS cannot be shutdown easily during the attack: economical reasons, inertia, . . .
5. Critical CPS requires high reliability/provable performance
6. . . .

Stuxnet



Stuxnet is the first discovered malware that spies on and subverts industrial control systems. It was discovered in June 2010.

2015 Ukraine Power Outage

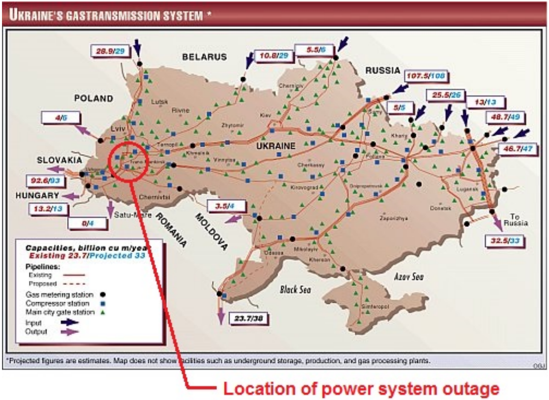


Figure: A successful attack on CPS can have devastating effects.

Industrial Control Systems

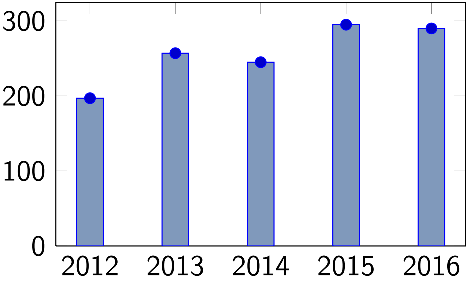


Figure: Reported Number of ICS Incidents by Fiscal Year

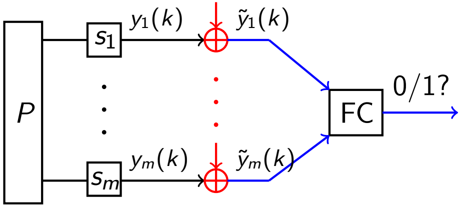
In FY 2016, ICS-CERT (Industrial Control Systems Cyber Emergency Response Team) received and responded to 290 incidents as reported by asset owners and industry partners.

Hardening CPS Security using Control Theory

- System Modelling
- Attack Modelling
- Intrusion Detection and Isolation
- Resilient Algorithm Design
- Fundamental Limitations
- Security Investment
- ...

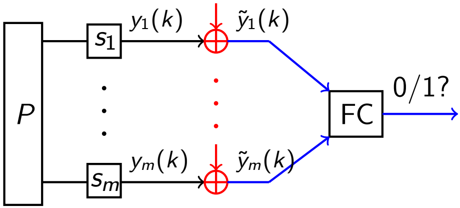
Outline

Binary Hypothesis Testing Under Attack



- Up to n sensors' measurements arbitrarily manipulated
  1. Compromising the sensors' hardware/software
  2. Hijacking the communication from sensors
  3. Physical attacks
- The system knows n, but does not know what sensors are compromised.

Motivating Example: Classic Probability Ratio Test



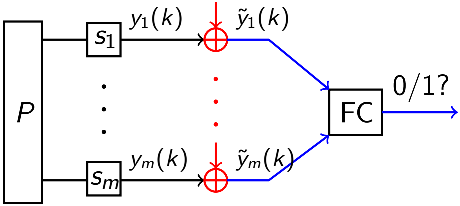
- At each time k, classic probability ratio test runs as

$$\theta = \begin{cases} 0 & \text{if } \sum_{t=1}^{k} \sum_{i=1}^{m} L(\tilde{y}_i(t)) \leq 0 \\ 1 & \text{if } \sum_{t=1}^{k} \sum_{i=1}^{m} L(\tilde{y}_i(t)) > 0, \end{cases}$$

where $L(\tilde{y}_i(k))$ is the log-likelihood ratio.

☞ Optimal without attacks

Motivating Example: Classic Probability Ratio Test



- At each time k, classic probability ratio test runs as

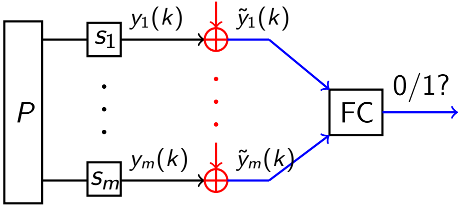
$$\theta = \begin{cases} 0 & \text{if } \sum_{t=1}^{k} \sum_{i=1}^{m} L(\tilde{y}_i(t)) \leq 0 \\ 1 & \text{if } \sum_{t=1}^{k} \sum_{i=1}^{m} L(\tilde{y}_i(t)) > 0, \end{cases}$$

where $L(\tilde{y}_i(k))$ is the log-likelihood ratio.

☞ Optimal without attacks

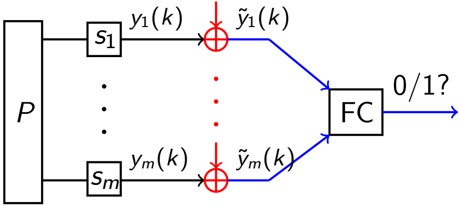
not secure at all

Motivating Example: Trimmed Mean Algorithm



- At each time k, trimmed mean algorithm runs as
  1. Remove the measurements with the largest n and smallest n log-likelihood ratios;
  2. Apply classic probability ratio test to the remaining $m - 2n$ data

Motivating Example: Trimmed Mean Algorithm



- At each time k, trimmed mean algorithm runs as
  1. Remove the measurements with the largest n and smallest n log-likelihood ratios;
  2. Apply classic probability ratio test to the remaining $m - 2n$ data

too conservative?

Tradeoff Between Security and Efficiency

- Security: The performance of the information fusion algorithm when under attack

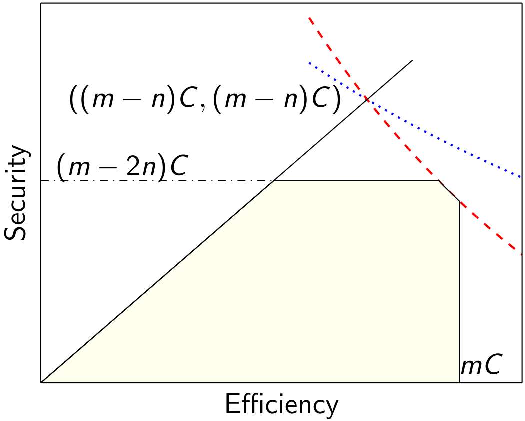
$$\liminf_{k\to\infty} -\frac{\log \max_{g,\theta} \Pr(f_k \neq \theta|\theta)}{k}$$

- Efficiency: The performance of the fusion algorithm when all sensors are benign.

$$\liminf_{k\to\infty} -\frac{\log \max_{\theta} \Pr(f_k \neq \theta|\theta)}{k}$$

- What is best achievable trade-off between security and efficiency?

Main Results



C: biggest contribution that one healthy sensor can provide

$$C \triangleq \liminf_{k\to\infty} -\frac{\log \max_\theta \Pr(f_k^* \neq \theta|\theta)}{k}$$

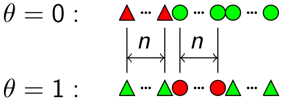
where f^* is the classic probability ratio test.

Proofs of Upper Bounds

- The best achievable efficiency is mC.
  - Classic probability ratio test

Proofs of Upper Bounds

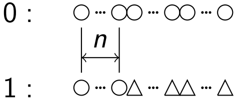
- The best achievable efficiency is mC.
  - Classic probability ratio test
- The best achievable security is $(m-2n)C$.
  - The achievability is deferred
  - The limit is shown by construct the following attack strategy.



green/red: healthy/compromised sensors
circle/triangle: different distributions

Fundamental Limits of Trade-off

- Consider the following two hypotheses:



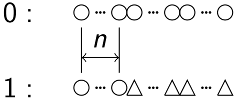
Suppose that we aim to find a detector such that the following is minimized.

$$\Pr(f = 1|0) + \phi \Pr(f = 0|1).$$

- Bayesian detection theory $\Longrightarrow$ fundamental relation between $\Pr(f = 1|0)$ and $\Pr(f = 0|1)$.

Fundamental Limits of Trade-off

- Consider the following two hypotheses:



  Suppose that we aim to find a detector such that the following is minimized.

$$\Pr(f = 1|0) + \phi \Pr(f = 0|1).$$

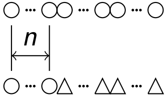
- Bayesian detection theory $\Longrightarrow$ fundamental relation between $\Pr(f = 1|0)$ and $\Pr(f = 0|1)$.
- Efficiency $\leq \Pr(f = 1|0)$, Security $\leq \Pr(f = 0|1)$

Fundamental Limits of Trade-off

- Consider the following two hypotheses:

0 :

1 :



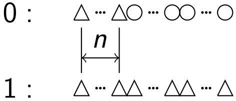
Suppose that we aim to find a detector such that the following is minimized.

$$\Pr(f = 1|0) + \phi \Pr(f = 0|1).$$

- Bayesian detection theory $\Longrightarrow$ fundamental relation between $\Pr(f = 1|0)$ and $\Pr(f = 0|1)$.
- Efficiency $\leq \Pr(f = 1|0)$, Security $\leq \Pr(f = 0|1)$
- Vary ϕ

Fundamental Limits of Trade-off: Cont'd

- Consider the following two hypotheses:



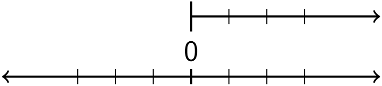
Achievability

There exists algorithms achieving the limits, i.e., the limit is tight.

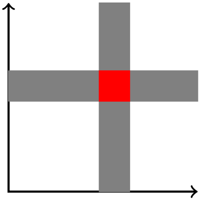
1. Each of the m measurements is mapped to *nonnegative* numbers by two functions l_0, l_1.
2. If there are $m-n$ values of l_0 whose sum is "small" enough, then choose $\hat{\theta} = 0$.
3. If there are $m-n$ values of l_1 whose sum is "small" enough, then choose $\hat{\theta} = 1$.
4. Compare the average of log-likelihood ratios with 0 to decide if $\hat{\theta} = 0$ or 1.

Intuitions of the Algorithm

- Nonnegative mapping.



- Safe kernel

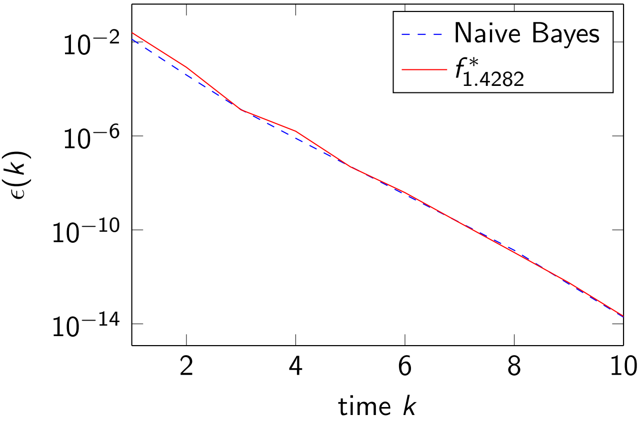


Gaussian Cases

- The best security $(m-2n)C$ and the best efficiency mC are achieved simultaneously
- Security is cost-free
  - ☞ Computational burden: $O(m)$ versus $O(m\log m)$
- More than Gaussian: “symmetric” distributions. There exists a constant a such that for any Borel measurable set $\mathcal{A}$,

$$\mu(a+\mathcal{A}) = \nu(a-\mathcal{A}).$$

Non-Asymptotic Performance



Secure Sensors

- A subset of sensors are well protected and cannot be compromised.
- Trade-offs?
- Similar ideas to prove limits and design algorithm?

Fundamental Trade-off when There are Secure Sensors

m_s normal sensors are replaced with secure ones.

- $2n \leq m - m_s$: nothing affected
  - ☞ The redundancy of the $m - m_s$ normal sensors is enough
- $2n > m - m_s$: the trade-off limit remains, and the maximum security level is increased from $\max(0, (m - 2n)C)$ to $m_s C$

☞ Do nothing or secure more than $m - 2n$ sensors

Detection Algorithm when There are Secure Sensors

1. Mapping by nonnegative functions l_0, l_1.
2. Sum l_0 of the m_s secure sensors and any $m - m_s - n$ of l_0 of the $m - m_s$ normal sensors, if there exist one "small" enough, then choose $\hat{\theta} = 0$.
3. Sum l_1 of the m_s secure sensors and any $m - m_s - n$ of l_1 of the $m - m_s$ normal sensors, if there exist one "small" enough, then choose $\hat{\theta} = 1$.
4. Compare with 0.

Conclusion

- We indeed can design algorithms that perform "well" whether or not the attacker is present
- In some cases, the cost of security is zero

Thank you for your time!