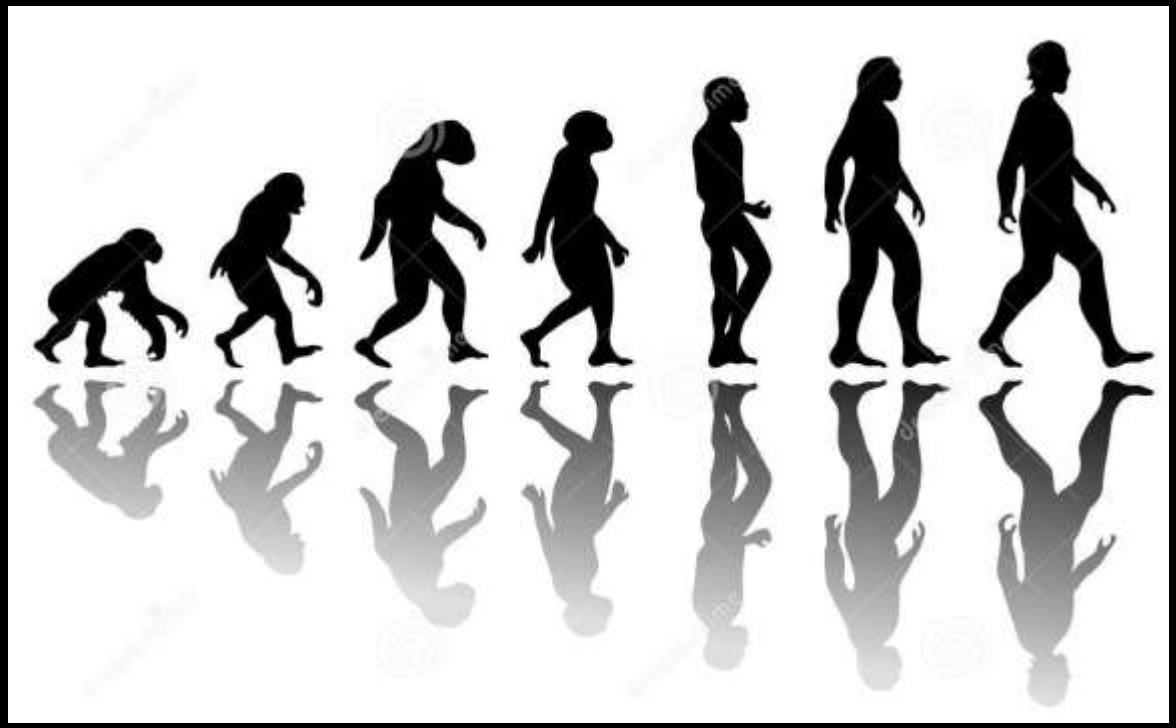
青眼看物联网安全:
小议控制理论的挑战与机遇

Security Issues of CPS: A Control-Theoretic Perspective

陈杰
香港城市大学电子工程系

Cyber Control Security: Where are we?

众说纷纭



莫衷一是

Attack

Detection

Defense

2008 2017 。。。 。。。

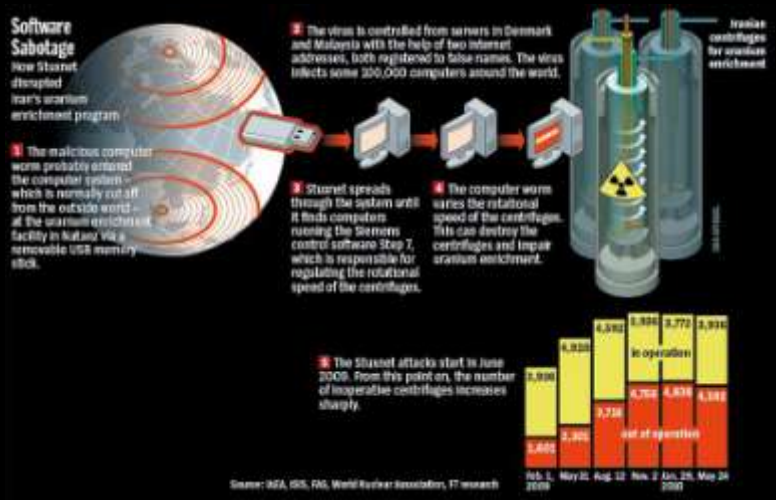
- Lots of hype, low SNR, not much to tell.
- He says, she says … Lack of consensus.
- Opportunities? Time to jump in the water?

Issues to Consider

- Is it a real problem or just “fake news”?
- What are the fundamental issues?
- Can the existing control theory handle it?

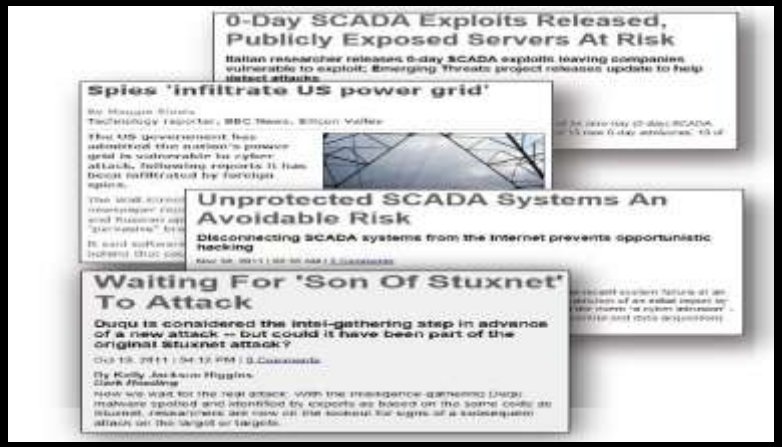
Cyber Attacks in Public Media

Power blackout, Ukraine, 2015



Stuxnet attack on Iranian nuclear plant, 2010

Turkish oil pipeline, 2008



US water distribution system

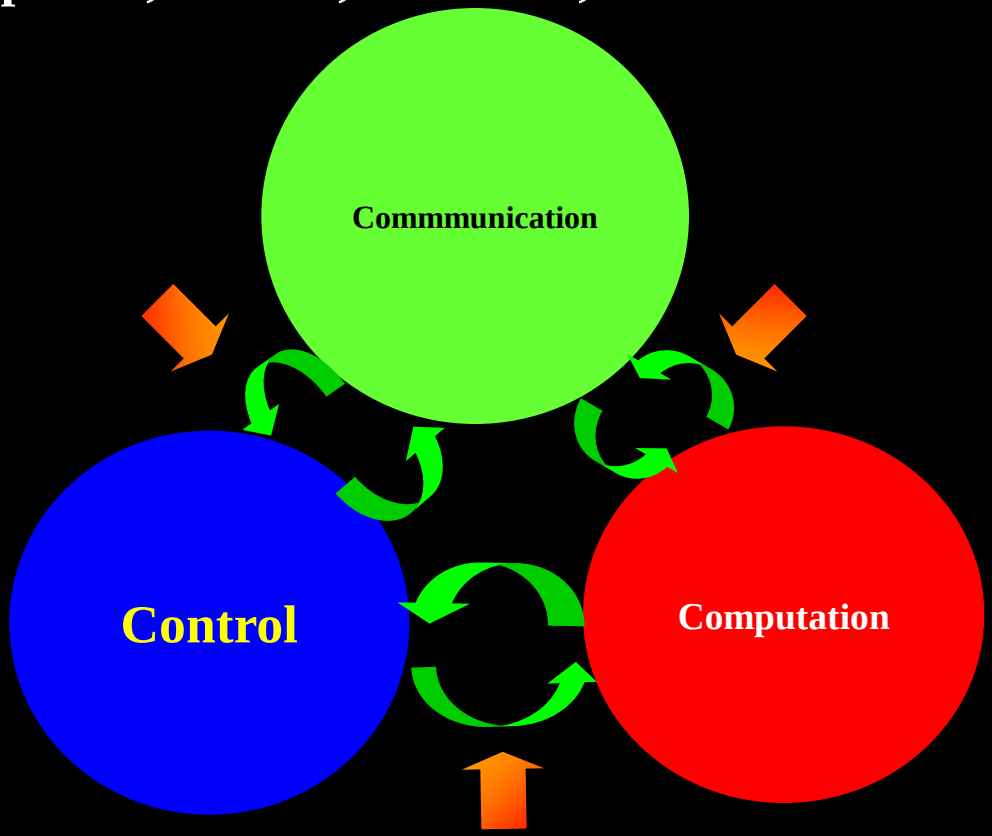
UK power plant, etc

It does seem to be … A Real Problem!

- Motivated by real-world cyber threats.
- Exhibits fundamental differences from conventional control problems:
  - System robustness: Continuous vs abrupt.
  - System faults: Benigh vs. malicious, independent vs. coordinated.
  - Internet theft: Different levels of damage.
- Motivated by needs for new research problems?

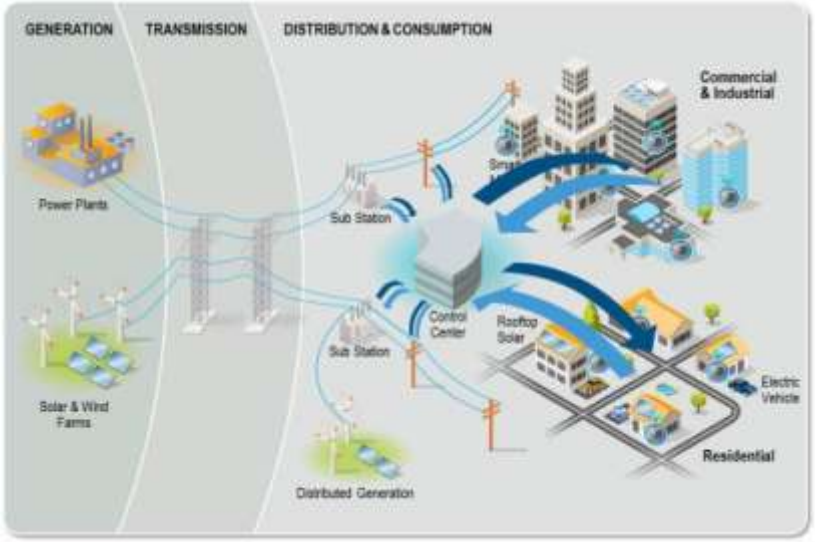
How so? A Glimpse into CPS

A vast networked and distributed system interconnecting physical plants, computers, sensors, actuators, and communication networks



Interconnection exposes the system to threats; malicious agents can gain access through cyber components-computers, communication networks-to launch attacks on sensors and actuators.

Examples of CPS



A typical CPS:

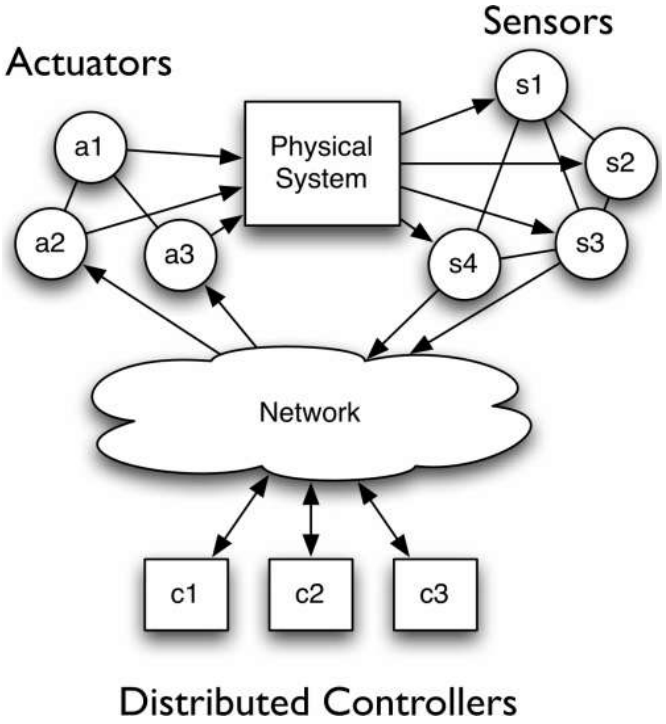
A long-distance, wide-area generation, transmission, and distribution system consisting of electronic field devices, communication networks and control centers.

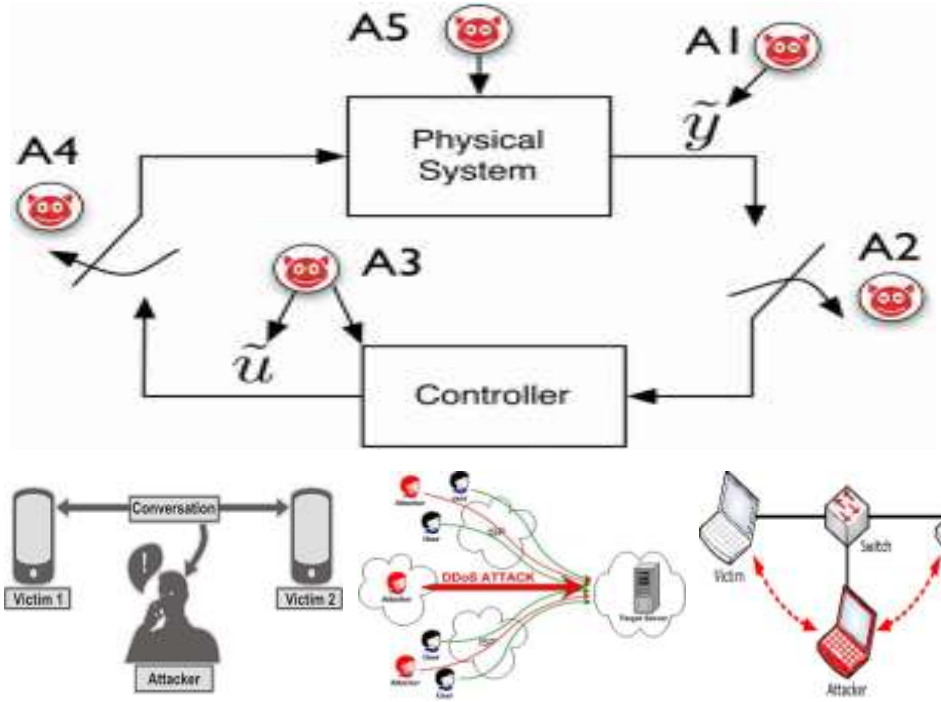
- **Water distribution systems**
- **Transportation networks**
- **Air traffic control**
- **… Many more safety-critical infrastructures!**

Initiatives around the Globe

- **US Homeland Security Control Systems Security Program, 2008**
- **Japan National Control Systems Security Center, 2012**
- **EU Network and Information Security Agency White Paper on Industrial Control Security, 2013**
- **China National Development and Reform Commission 发改委国家信息安全工业控制专项, 2013**
- **中共中央网络安全和信息化领导小组，2013**

Attack Models

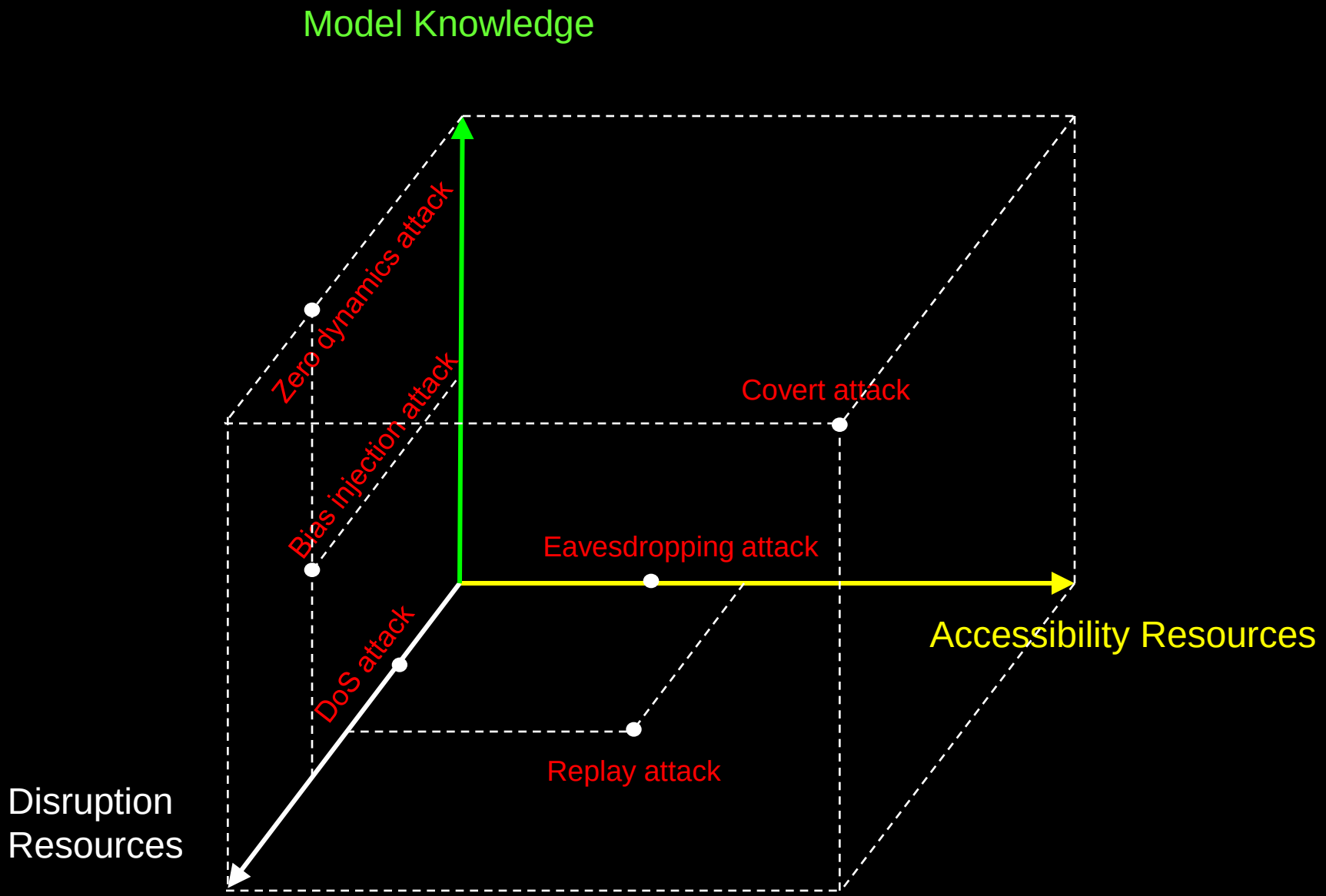




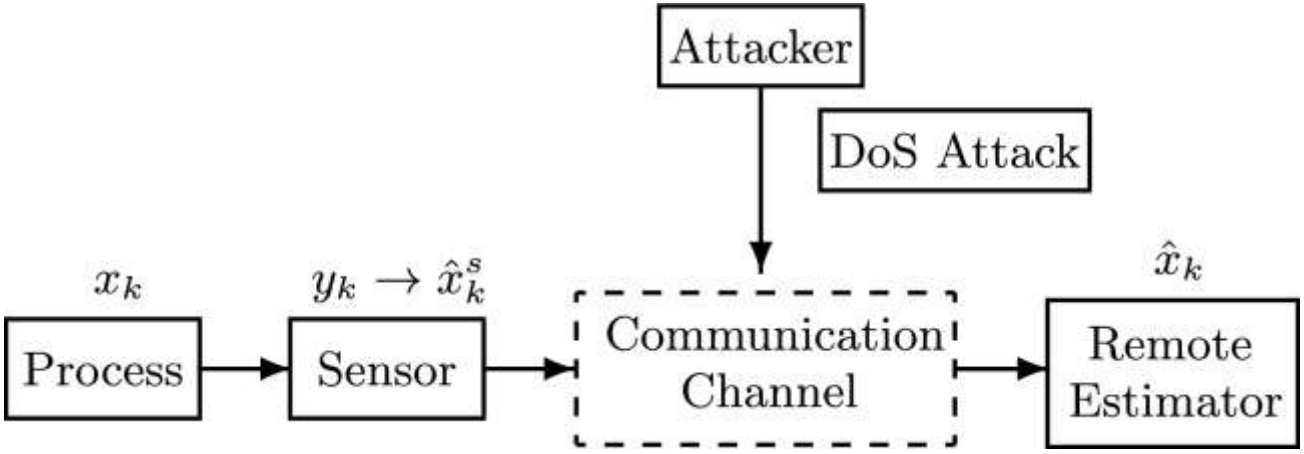
- **Confidentiality attack (Replay attack)**
- **Availability attack (DoS attack)**
- **Integrity attack (False data injection)**

All attempting to change a system's behavior, to steer a system's states away from its normal operating range.

The Dangerous Cyber World



DoS Attack



Attack typically modeled as an additive or a modulating Bernoulli process; requires no system knowledge, but SNR a critical factor.

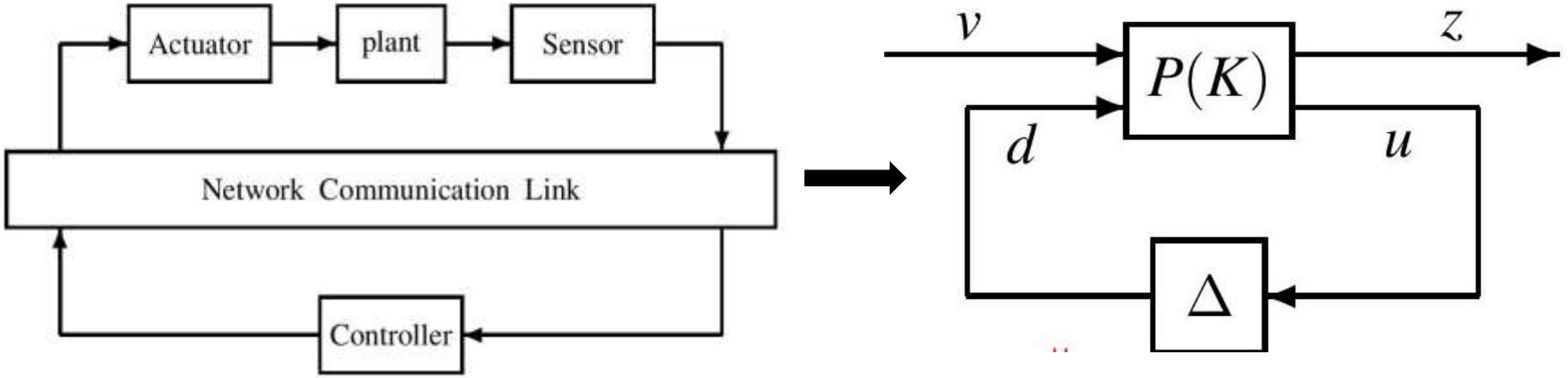
- **Detection: Reformulated as a networked estimation problem**

  (A number of such problems are solved by Cheng, Chen, and Shi)

- **Defense: Reformulated as a networked control problem**

- **Stochastic robust control and game theories can be applicable**

Defense of DoS Attack: A Networked Control Problem



- **DoS attacks formulated as Bernoulli processes, and reformulated as structured multiplicative stochastic uncertainties**

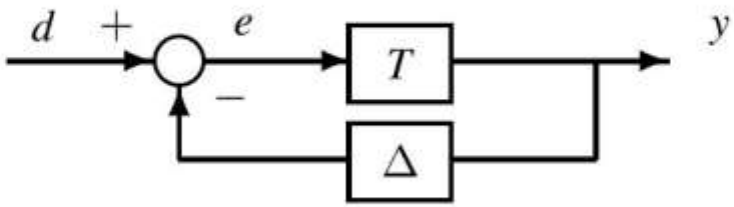
$$\Delta(k) = diag\{\Delta_1(k), \dots, \Delta_n(k)\}$$

$\Delta_i(k)$: *Uncorrelated i.i.d. random sequences with variances* σ_i^2

- **Defense: Reformulated as a robust/mean-square/variance control problem subject to multiplicative stochastic uncertainties**

Mean-Square Small Gain Theorem

(Willems & Blankenship, 1971; Hinrichsen & Pritchart, 1996; Lu & Skelton, 2002; Elia, 2005)



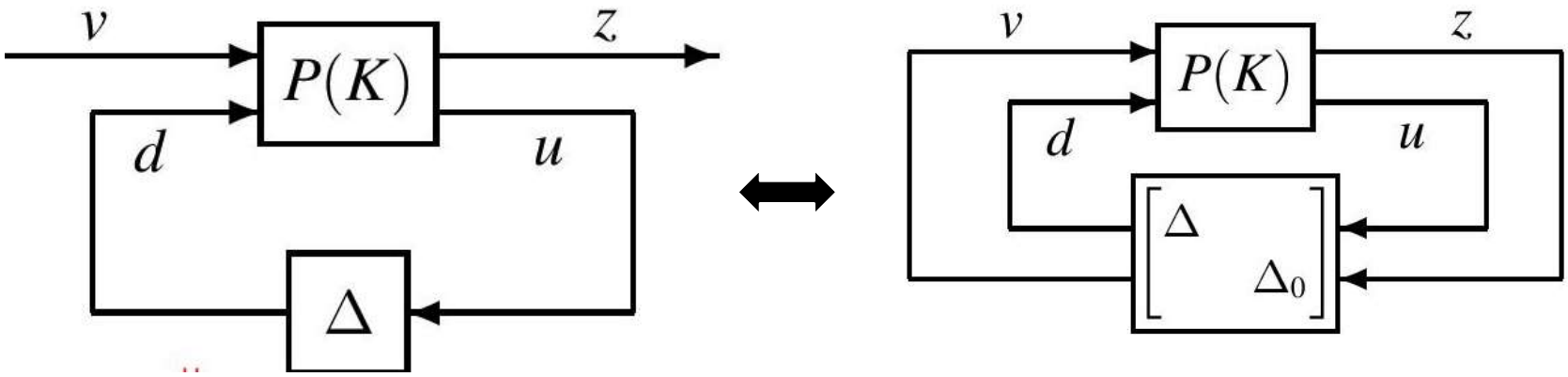
Mean Square Small Gain Theorem : Let T be a stable LTI system, and Δ be given by $\Delta(k) = \text{diag}\,(\Delta_1(k), \cdots, \Delta_m(k))$. Under Assumptions 1-3, the system is mean-square stable if and only if

$$\rho(W) < 1,$$

where

$$W = \begin{bmatrix} \|T_{11}\|_2^2 & \cdots & \|T_{1m}\|_2^2 \\ \vdots & \cdots & \vdots \\ \|T_{m1}\|_2^2 & \cdots & \|T_{mm}\|_2^2 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_m^2 \end{bmatrix}.$$

Defense of DoS Attack: A Networked Stabilization Problem

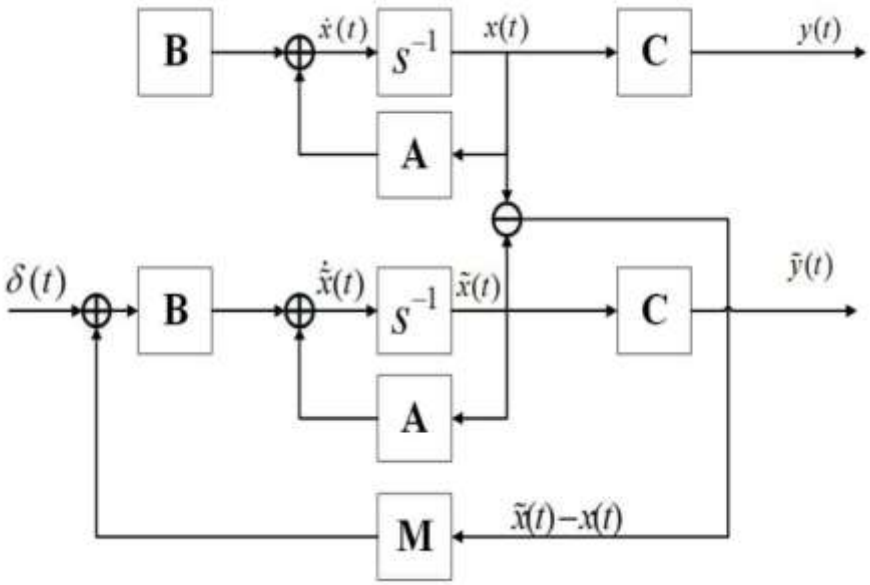


- **Mean-square stabilization and optimal control problems can be solved in the same framework.**
- **Defense of DoS attack: Reformulated as a mean-square optimal control problem.**
- **Challenge: Distributed mean-square optimal control.**

T. Qi, J. Chen, W. Su, and M. Fu, "Control under Stochastic Multiplicative Uncertainties: Part 1, Fundamental Conditions of Stabilizability," *IEEE TAC*, vol. 62, no. 3, March 2017, pp. 1269-1284.

W. Su, J. Chen, M. Fu, and T. Qi, "Control under Stochastic Multiplicative Uncertainties: Part 2, Optimal Design for Performance," *IEEE TAC*, vol. 62, no. 3, March 2017, pp. 1285-1300.

Integrity Attack



$$d(t) = M\Delta x(t) + \delta(t)$$

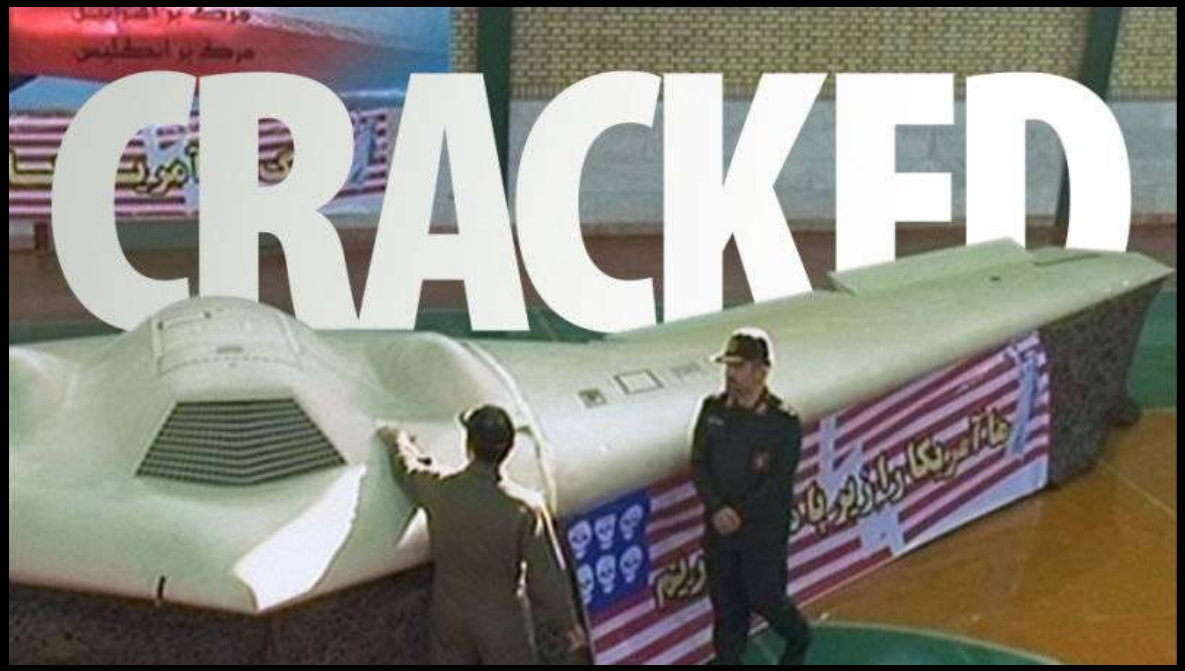
$$\Delta x(t) = \tilde{x}(t)\text{-}x(t)$$

$$\begin{aligned} \max \quad & f(d) \\ s.t. \quad & s(d) \leq \epsilon \\ & ||d||_0 \leq n \end{aligned}$$

s(d): Stealthiness, deceptiveness of the attack
$||d||_0$: Cardinality, accessibility to attacker
f(d): Destructiveness, attack performance

Attacker knows the system model, and launches attacks by injecting a false signal and additionally tampering with the system's states, while hiding its identity.

Combined DoS and Integrity Attacks



US most advanced spy drone RQ-170 downed by

Iranian Cyberwarfare Unit, 2011

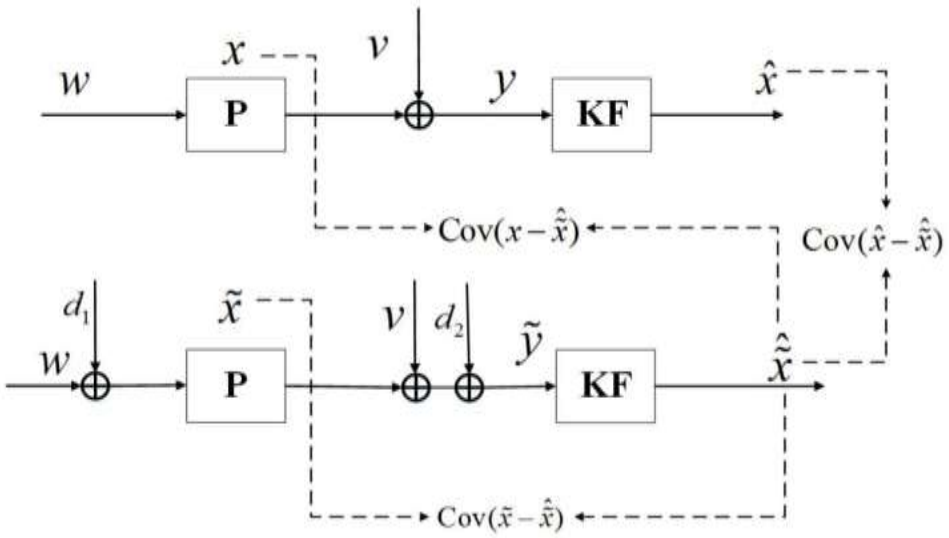
Means of Cyber Security

- **Passive: Mature and effective**
  - **Encryption**
  - **Coding**
  - **Firewall**
- **Active: Secured control by estimation and control**
  - **Detection by estimation algorithm**
  - **Defense by control algorithm**
  - **Perhaps system structure plays a more important role**

Issues of Cyber-Secured Control

- **Performance degradation**
- **Detectability and detection**
- **Prevention and defense**

Performance Degradation



$d_1(k)$: sensor attack

$d_2(k)$: actuator attack

$$\mathrm{P}(k)=Cov(\hat{x}(k)-\hat{\tilde{x}}(k))$$

covariance between the normal estimated state and the state under attack

$$\tilde{x}(k+1) = A\tilde{x}(k) + w(k) + d_1(k)$$
$$\tilde{y}(k)=C\tilde{x}(k)+v(k)+d_2(k)$$

- **Performance degradation can be quantified by deterministic, probabilistic and mean square measures.**
- **Can incorporate stealthiness measure to formulate a constrained optimization problem.**

Example: First-Order System

A=a (|a|<1), C=c, $P_d < \infty$, l: $Kalman\ filter\ gain$

False signal injection **The error covariance under attack converges if and only if**

$$P_d < \infty$$

$$\sup_k P(k) = \frac{[1 + a^2(1 - lc)](lc)^2}{(1 - a^2)[1 - a^2(1 - lc)][1 - a^2(1 - lc)^2]} P_d$$

Feedback attack **The error covariance under attack diverges if and only if**

$$\rho\left(\begin{bmatrix} (1 - lc)a - lcM & lca \\ -M & a \end{bmatrix}\right) > 1$$

Undetectable Attacks

- **An attack $d(t)$ is said to be undetectable if for any initial states $x(0)$ and $\tilde{x}(0)$, the system's output satisfies the relation**

$$y(\tilde{x}(0), 0, t) = y(x(0), d(t), t)$$

An undetectable attack is perfectly stealthy and cannot be observed from output measurements.

- **Attack to linear system**

$$y(x(0) - \tilde{x}(0), d(t), t) = 0$$

This means that the attack is "blocked" from the system's output, reminiscent of zeros of a systems.

Zero Dynamics Attack and Defense

Zero attack The attack signal $d(t) = e^{zt}d_0$ is undetectable if and only if

$$\begin{bmatrix} zI - A & B \\ C & 0 \end{bmatrix} \begin{bmatrix} x(0) - \tilde{x}(0) \\ d_0 \end{bmatrix} = 0$$

i.e., when z is a transmission zero of $G(s)$ with zero direction d_0.

Defense of zero attack It suffices to add more columns, i.e., more sensors, to the output matrix C, so that the matrix

$$\begin{bmatrix} zI - A & B \\ C & 0 \end{bmatrix}$$

is full row rank.

This changes the system structure and can effectively remove the undesirable zero.

Positive Distributed Systems

$$\begin{bmatrix}\dot{\xi}(t)\\ \dot{\varsigma}(t)\end{bmatrix}=\begin{bmatrix}0 & I\\ -L & -K\end{bmatrix}\begin{bmatrix}\xi(t)\\ \varsigma(t)\end{bmatrix}+Bd(t)$$

$$y(t)=C\begin{bmatrix}\xi(t)\\ \varsigma(t)\end{bmatrix}$$

- **A model for power grid, with representing the phases and frequencies and L, K lumping together inductances**
- **A model for a multi-agent system with double integrator agents:**

$$\dot{\xi}_i(t)=\varsigma_i(t)$$

$$\dot{\varsigma}_i(t)=\boldsymbol{u_i}(\mathbf{t})$$

$$\boldsymbol{u_i}(\mathbf{t})=-\boldsymbol{k_i}\varsigma_i(t)+\sum_{j\in N_i}a_{ij}\,(\xi_j(\mathbf{t})-\boldsymbol{\varsigma_i}(\mathbf{t}))$$

Defense of Second-Order Positive Systems

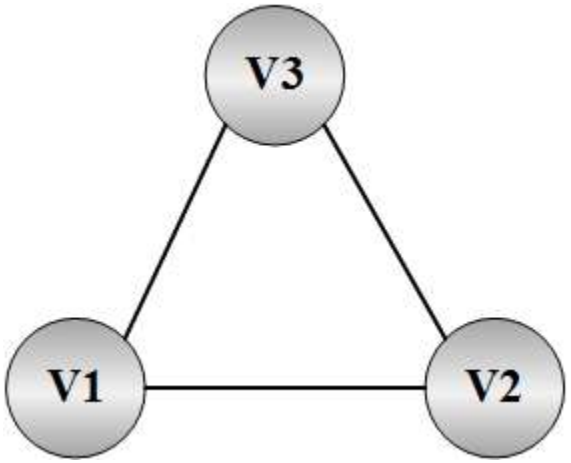
Zero attack Any single attack $d(t) = e^{zt}e_i$ at the i –th actuator can be defended by placing a single sensor at any node j over a strongly connected graph, i.e., with $B = [e_i^T \;\; 0^T]^T$ or $B = [0^T \; e_i^T]^T$, it is always possible to select $\mathrm{C} = [e_j^T \;\; 0^T]^T$ or $\mathrm{C} = [0^T \; e_j^T]^T$ such that

$$\begin{bmatrix} zI - A & B \\ C & 0 \end{bmatrix}$$

is full row rank, where $A = \begin{bmatrix} 0 & I \\ -L & -K \end{bmatrix}$.

Redesign plant structure so as to remove the zero.

An Example



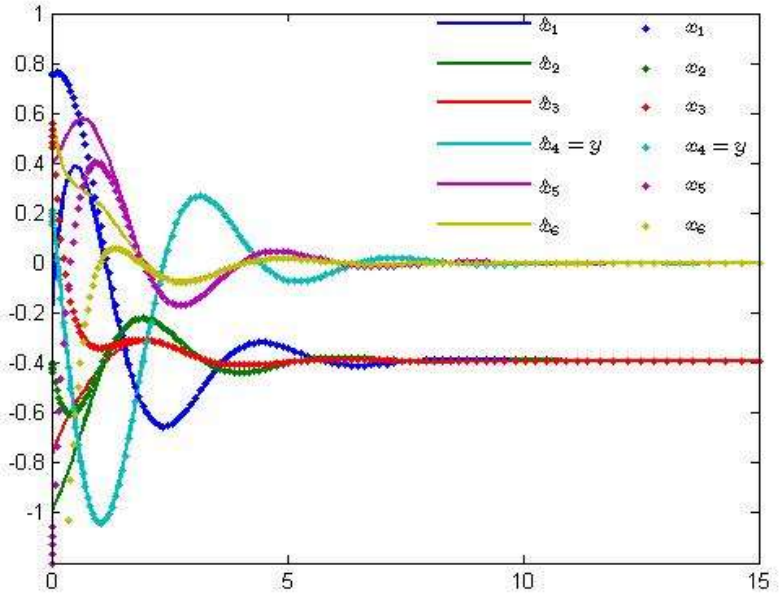
$$L = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \mathrm{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$\begin{bmatrix} \dot{\xi}(t) \\ \dot{\varsigma}(t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ -L & -K \end{bmatrix} \begin{bmatrix} \xi(t) \\ \varsigma(t) \end{bmatrix} + Bd(t)$$

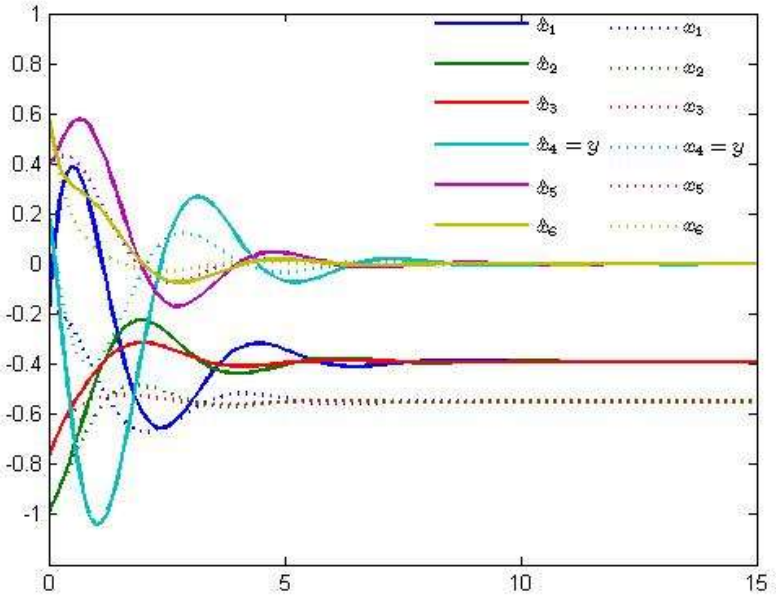
$$y(t) = C \begin{bmatrix} \xi(t) \\ \varsigma(t) \end{bmatrix}$$

- **Select $B = [1\ 0\ 0\ 0\ 0\ 0]^T$, $\mathrm{C} = [0\ 0\ 0\ 1\ 0\ 0]^T$.**
- **Zeros: $\{0, -2.6914, -1.1543 \pm 1.2059j\}$**
- **Attack signal: $d(t) = 2.583e^{-2.6914t}$**

Simulation Result

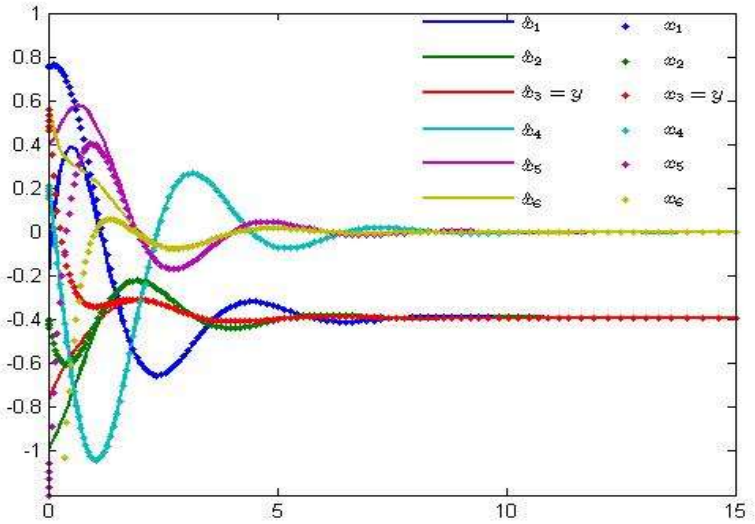


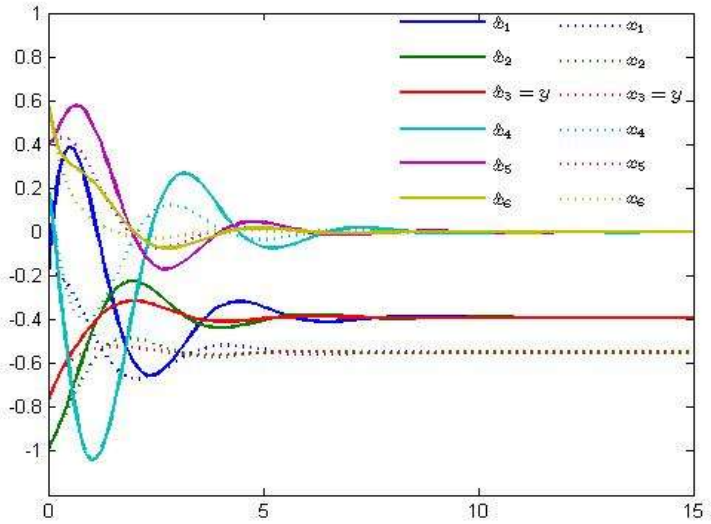
Attack (light blue) cannot be detected



System states under attack

Attack Detection





- **Select $B = [1\ 0\ 0\ 0\ 0\ 0]^T$, C $= [0\ 0\ 1\ 0\ 0\ 0]^T$.**
- **Zeros: $\{-1.0000, -2.6914, -1.0000 \pm 1.4142j\}$**
- **Attack signal: $d(t) = 2.583e^{-2.6914t}$**

A System-Theoretic Interpretation

- **The result shows that the system structure can be very useful in detecting and defending against "undetectable" attacks. This makes sense intuitively.**
- **The interpretation in terms of zero direction seemingly suggests a geometrical perspective.**

- **An attacker's goal:**
  - **Stealthy: Unobservable from the system's output.**
  - **Destructive: Make some states as controllable as possible.**

Geometrical Considerations

- **Controllability and observability Gramian:**

$$AL_c + L_c A^T = BB^T$$

$$A^T L_o + L_o A = C^T C$$

- **Diagonalization**

$$T_1 L_c T_1^T = \Sigma_c = diag\{\sigma_{c1}, \dots, \sigma_{cn}\}$$

$$T_2^T L_o T_2 = \Sigma_o = diag\{\sigma_{o1}, \dots, \sigma_{on}\}$$

Strongly controllable states: States corresponding to $\{\sigma_{c1}, \dots, \sigma_{cr_s}\}$

Weakly observable states: States corresponding to $\{\sigma_{o(r_w+1)}, \dots, \sigma_{on}\}$

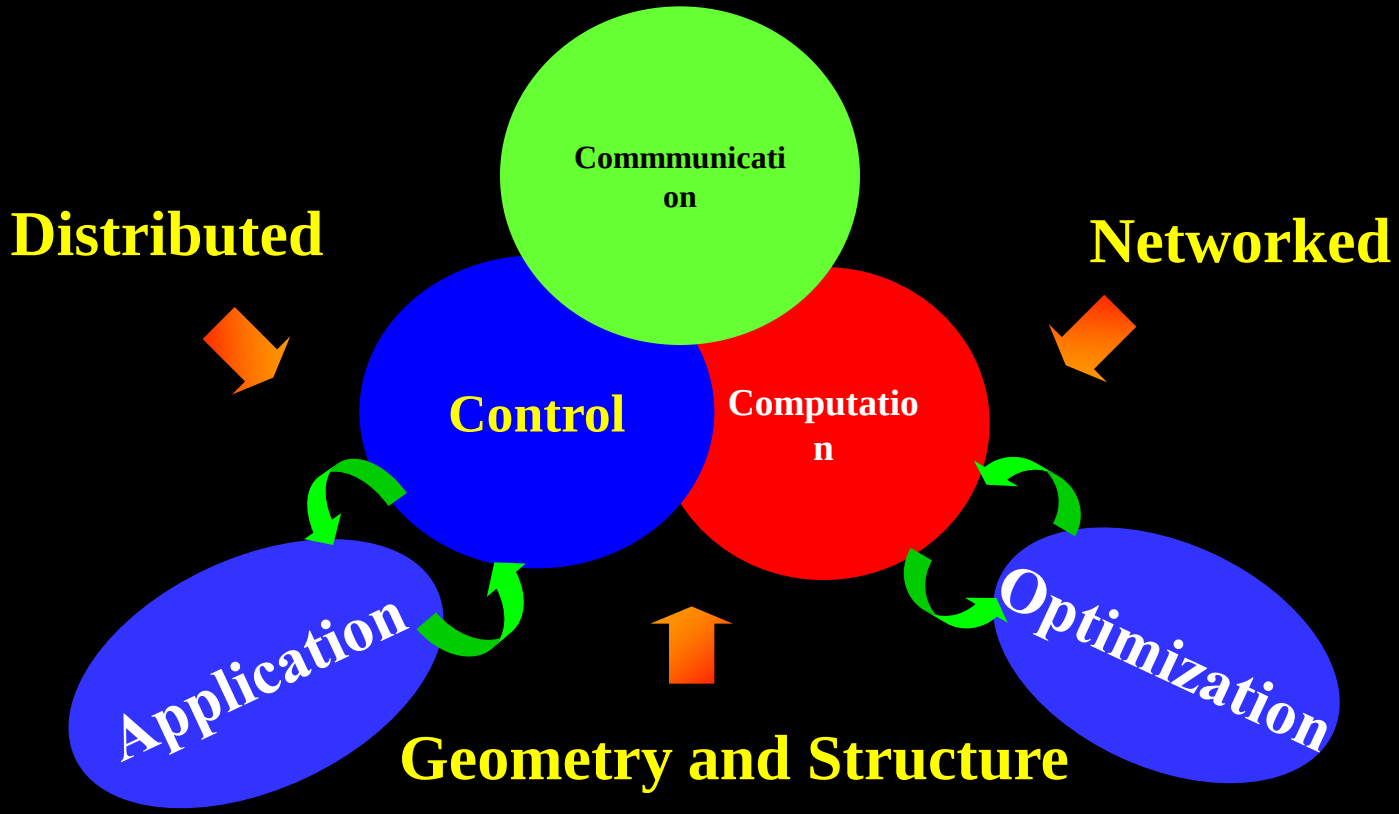
- **Strongly controllable subspace** C_s **and weakly observable subspace** C_w

Vulnerable States $C_s \cap C_w$

Research Thrusts:

To develop fundamental scientific understanding

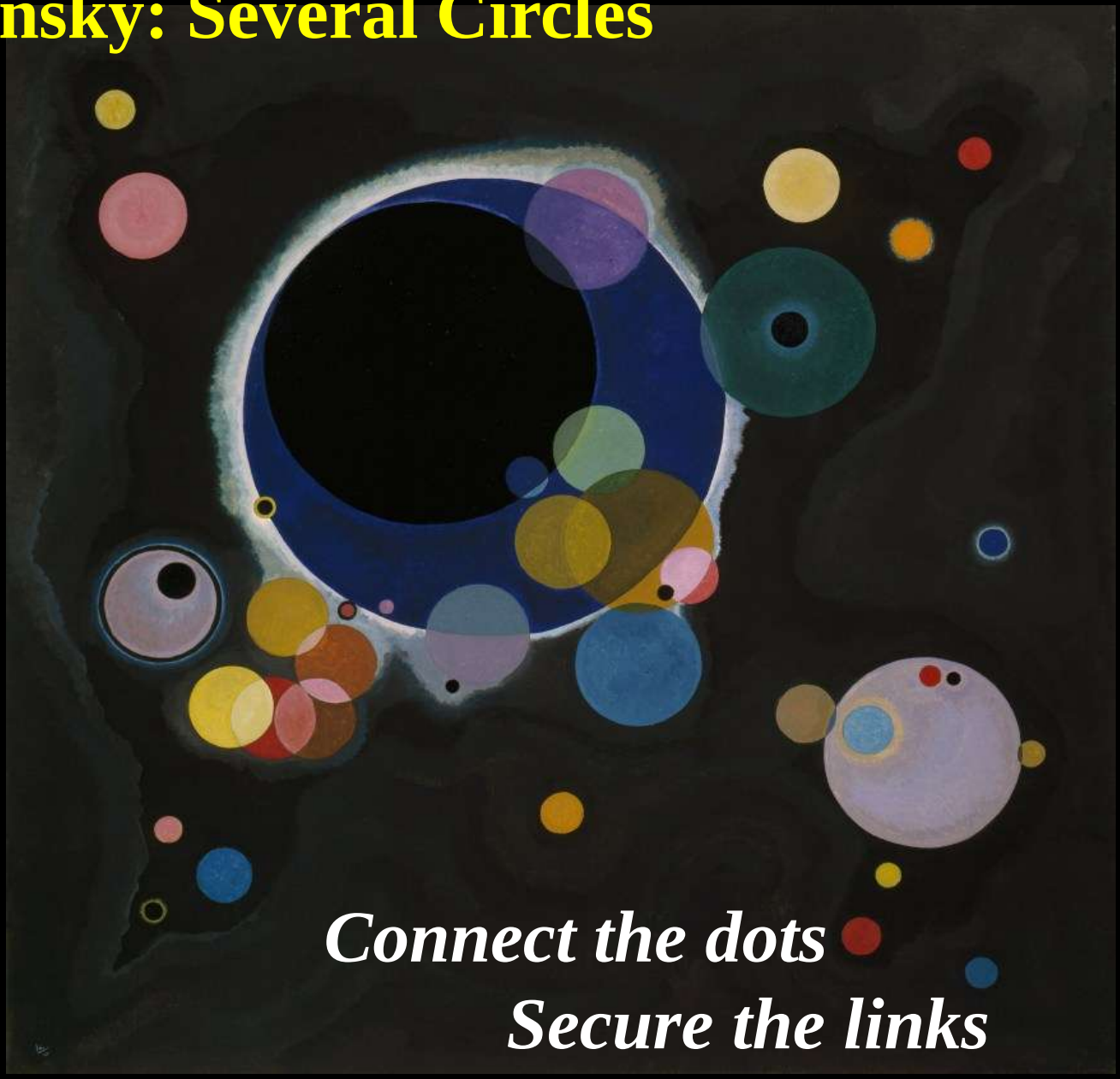
To develop enabling engineering tools and algorithms



Interdisciplinary/Collaborative Approach:

- **Understanding the practical problems**
- **Understanding the real issues**
- **Working with computer science/engineering, communications people**

Kandinsky: Several Circles



Connect the dots

Secure the links

Connect the world