Integration of Resilient Control and Economic Incentives

Saurabh Amin

in collaboration with the FORCES team

FORCES Kickoff Meeting
Washington, D.C., April 12th, 2013

Outline

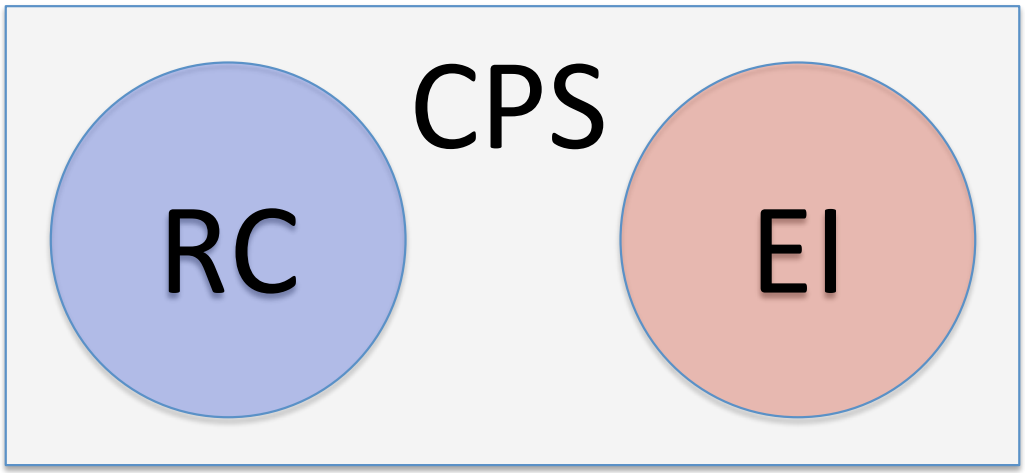
1. CPS resilience: the FORCES approach
2. Resilient Control (RC)
   - Electricity networks
   - Transportation networks
   - Water networks
3. Economic Incentive (EI) Mechanisms
4. RC + EI Validation

Outline

A dichotomy in CPS

Resilient Control (RC) tools

Primarily driven by the technological developments with a view of distributed sensing of phenomena, change detection and fault diagnosis, and closed-loop control over sensor-actuator networks.



Economic Incentives (EI) tools

Primarily driven by the strategic interactions of human decision makers within systemic societal institutions with a view of aligning individually optimal allocations with socially optimal ones.

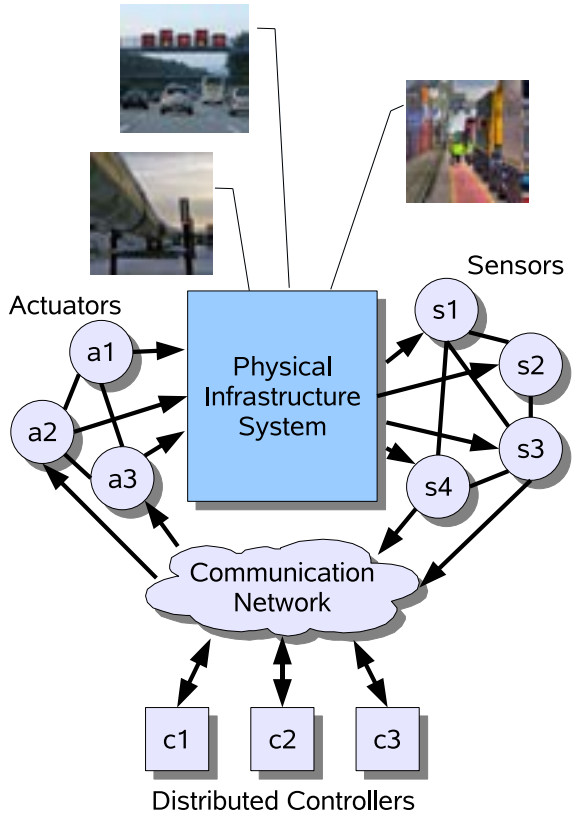
Sensor Webs → Action Webs

New functionalities

- State awareness
- Real-time closed-loop control
- Demand management
- Incident management

Need for RC + EI integration

1. Off-the-shelf IT devices
   ⇒ software bugs & hardware flaws
2. Open networks
   ⇒ accessible by strategic attackers
3. Multi-party management
   ⇒ incentives for misbehavior
4. Large # of field devices
   ⇒ increased attack surface



Large-scale critical infrastructures are Cyber-Physical Systems (CPS)

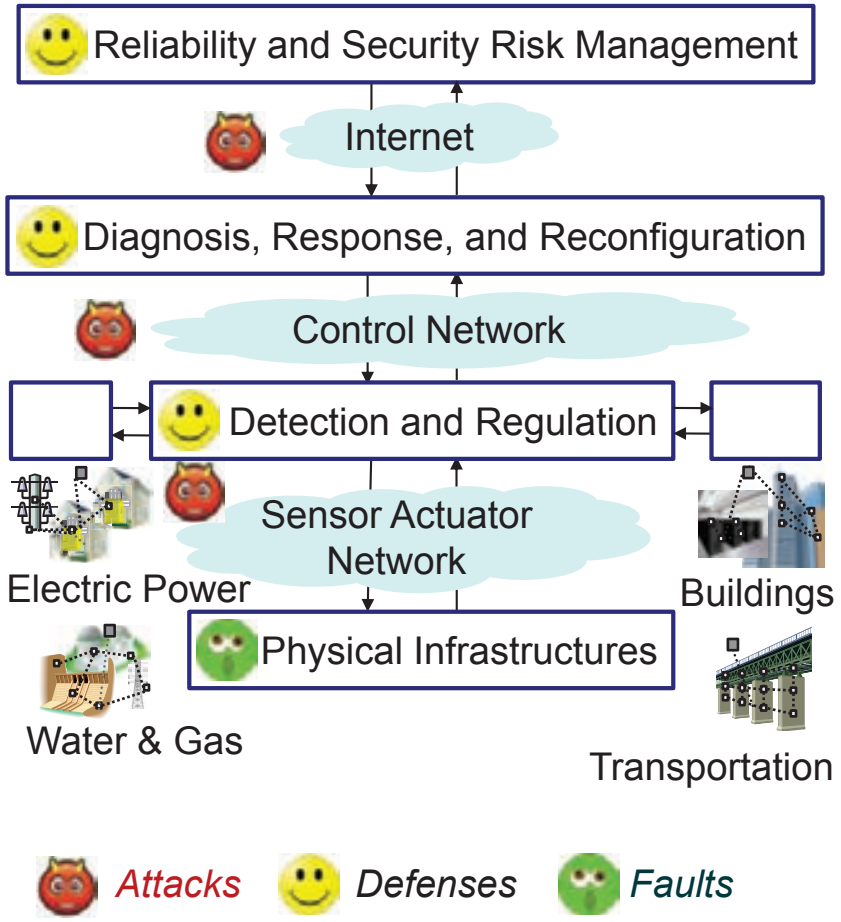
FORCES approach to high-confidence CPS

Theory of robust control

- Assessment, diagnosis, & response
- Stealthy attack diagnosis
- Attack-resilient control

Theory of incentive mechanisms

- Information deficiencies
- Individual vs. social incentives
- Interdependent network risks



Dichotomy of RC and EI is no longer suited for ensuring resilient CPS.

FORCES infrastructure domains

CPS Environments	RC	EI
Road traffic operations	Distributed traffic control (metering & control)	Congestion pricing and traveler incentives
Airport and airspace operations	Robust air traffic scheduling and routing	Strategic allocation of airport & airspace resources
Electricity transmission & bulk-power operations	Wide-area monitoring, state estimation, and MPC	Transmission planning & cost allocation
Electricity distribution & demand management	Distributed load control, control of smart appliances	Incentives for peak-shaving & reducing price volatility

Cyber-attacks and privacy threats

Integrity: A1 & A3

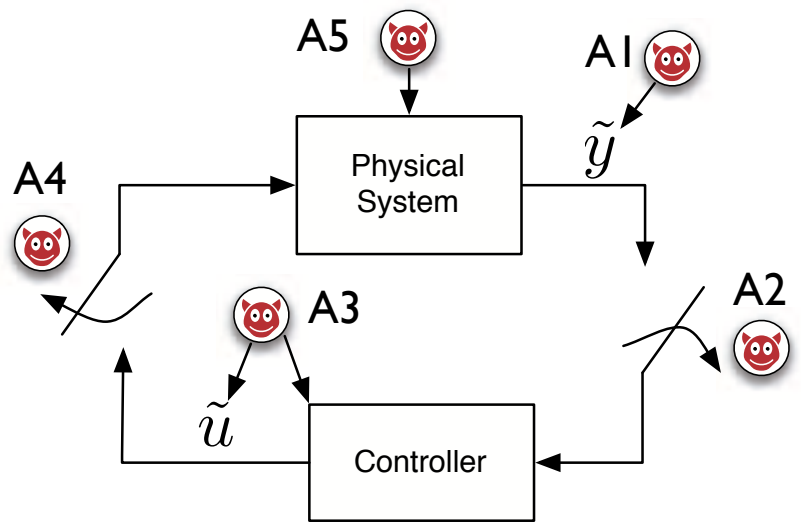
- Deception causes lack of integrity
- Trustworthiness of CPS data

Availability: A2 & A4

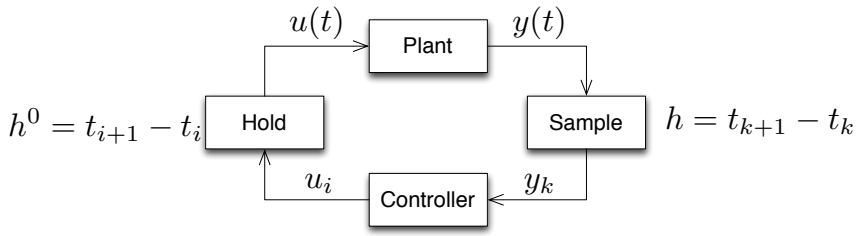
- Denial-of-service (DoS) causes lack of availability
- Accessibility of CPS components

Privacy

- Disaggregate usage data collection causes lack of privacy
- Minimization of privacy-sensitive data



Deception & DoS attacks to CPS



Privacy-preserving sampling of CPS

Claim #1: Cyber attacks $\neq$ Random faults

Attackers

- Malicious insiders
- Computer hackers
  - cyber criminals, cyber warriors, hacktivists, rogue hackers, spies

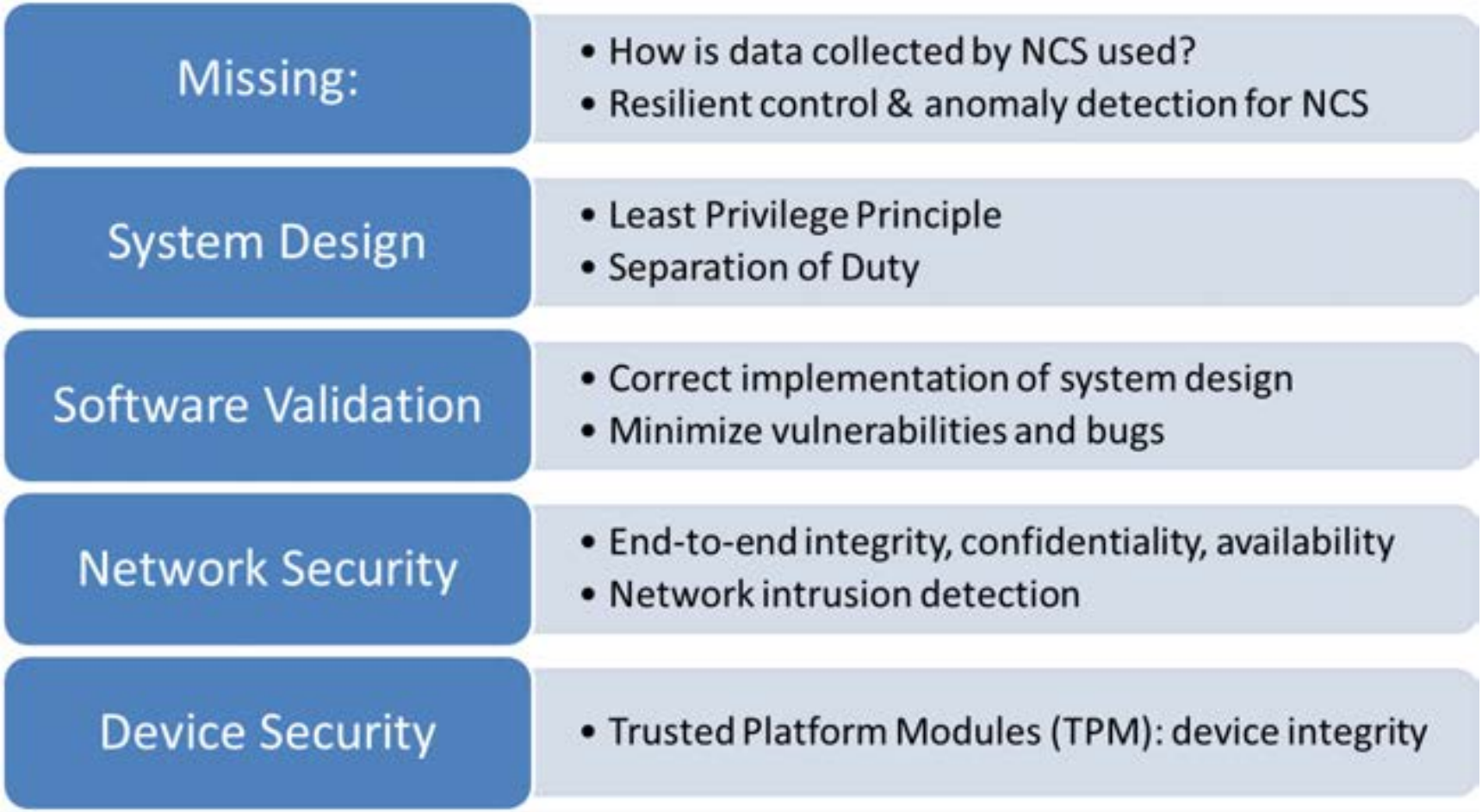
Attacker may manipulate CPS data

- Time between telemetry requests can be used for malicious traffic injection
- Both malicious and legitimate traffic can travel through encrypted tunnels

A. Cárdenas, S. Amin, S. Sastry, et al. [ASIACCS]

S. Amin, X. Litrico, S. Sastry, A. Bayen. [HSCC '10]

Claim #2: IT security is necessary but not sufficient



A. Cárdenas, S. Amin, S. Sastry. [HotSec '08]

A. Cárdenas, S. Amin, G. Schwartz. [HiCoNS'12]

Claim #3: CPS operators underinvest in security

Stuxnet worm [’10-’11]

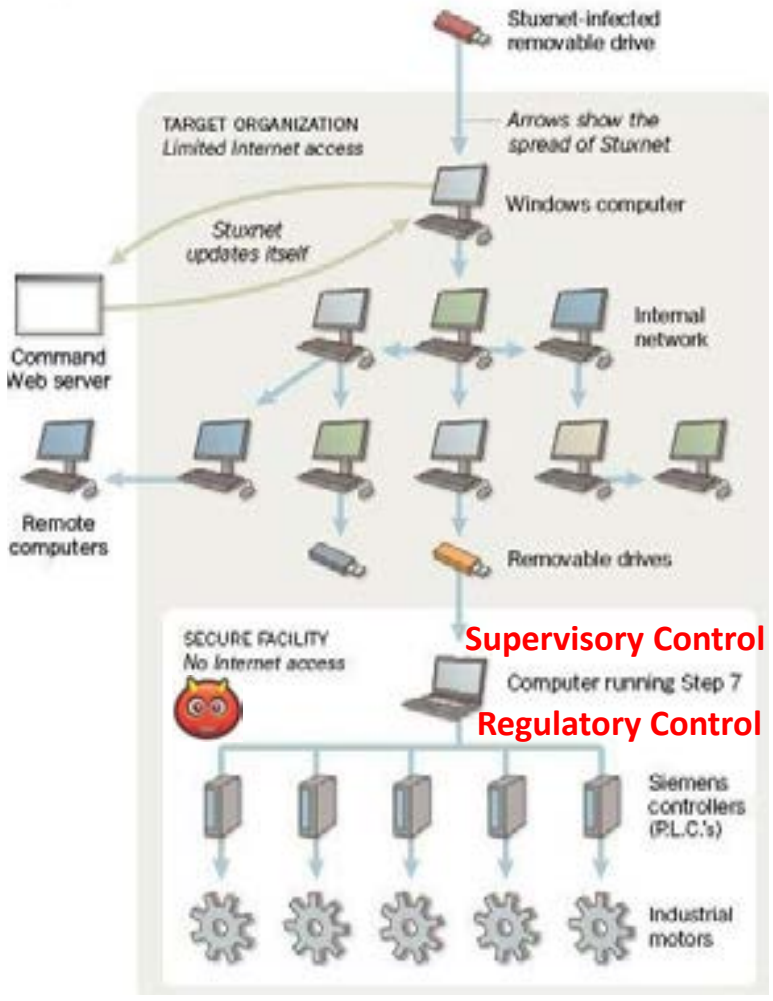
- Targets SCADA systems
- Four zero-day exploits, windows rootkit, antivirus evasion, p-2-p updates, network infection routines
- Reprograms PLC code
- Information stealing: Duqu [’11-’12]

Network induced risks

- Security is a public good
- Infrastructures are privately managed
- Individual & social incentives differ

S. Amin, G. Schwartz, S. Sastry.

GameSec ’10, CDC ’11, Automatica



Source: Symantec, NYT

Claim #4: Reliability-Security failures are non-isolable

Hacker Apparently Triggers Illinois Water Pump Burnout

Attack illustrates the extent to which industrial control systems are Internet-connected, yet lack basic password checks or access controls.

By **Mathew J. Schwartz** InformationWeek
November 21, 2011 11:45 AM

Federal authorities are investigating a hack that resulted in the burnout of a water pump at the Curran-Gardner Township Public Water District in Illinois. Located west of Springfield, Ill., the utility serves about 2,200 customers.

A hacker apparently exploited a supervisory control and data acquisition (SCADA) system that managed the water pump and set the pump to continually turn on and off. Only after the pump failed, earlier this month, did plant operators discover that their systems had been exploited, apparently in September. The attack appeared to have been launched from a server based in Russia.

DHS, FBI Dispute Illinois Water Hack

Feds say their preliminary investigation finds no evidence of stolen credentials or foreign attackers.

By **Mathew J. Schwartz** InformationWeek
November 23, 2011 12:41 PM

The Department of Homeland Security and FBI on Tuesday issued a joint statement disputing that an Illinois water utility's industrial control systems were recently hacked.

The DHS's Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) and the FBI cautioned that findings issued by the DHS Illinois State Fusion Center--aka the Illinois State Terrorism and Intelligence Center (STIC)--"were intended to be initial raw reporting and not conclusive in nature."

G. Schwartz, S. Amin, et al. [Allerton '11], S. Amin, G. Schwartz, S. Sastry. [CDC'11]

Claim #5: Security legislation needs a scientific base

Cybersecurity Act S.2105 *vs.* SECURE IT Act S. 2151

- S.2105 [Lieberman et al.]: DHS to access risks and vulnerabilities to critical infrastructures. Recommends a *regulation* that requires private companies owning designated critical infrastructure to *certify* that their cybersecurity capabilities rise to an appropriate level.
- S. 2151 [McCain et al.]: Federal contractors *required* to inform the government about cyber threats. Provides *liability protections* for the private sector to share cyber threat information through established channels and the Department of Commerce.

Big questions: Regulations? Incentives? Privacy laws?

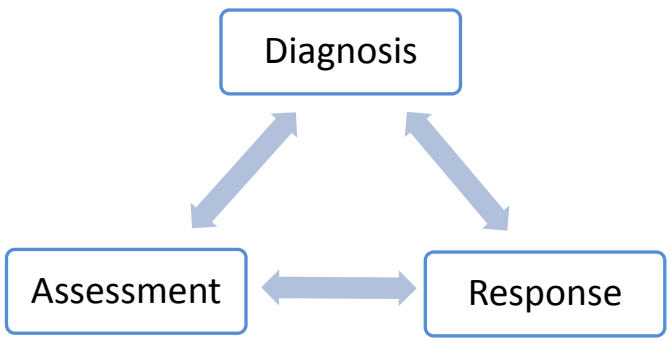
R. Böhme, G. Schwartz. [WEIS'10]

G. Schwartz, B. Johnson, S. Sastry [Work-in-progress]

Outline

Resilient control for CPS security

1. Threat assessment
   - How to model attacker and his strategy?
   - Consequences to the physical infrastructure
2. Attack diagnosis
   - How to detect manipulations of sensor-control data?
   - Stealthy [undetected] attacks
3. Attack resilient control
   - Design of resilient control algorithms?
   - Fundamental limitations



Outline

Indian Blackout of 2012

- 620*M* people without power
- 10*x* severe that US blackout of 2003

Control + Incentive issues:

1. Overdraw by utilities
2. High loading
3. Weak transmission
4. Mis-operation of protection systems

Electricity Theft: India

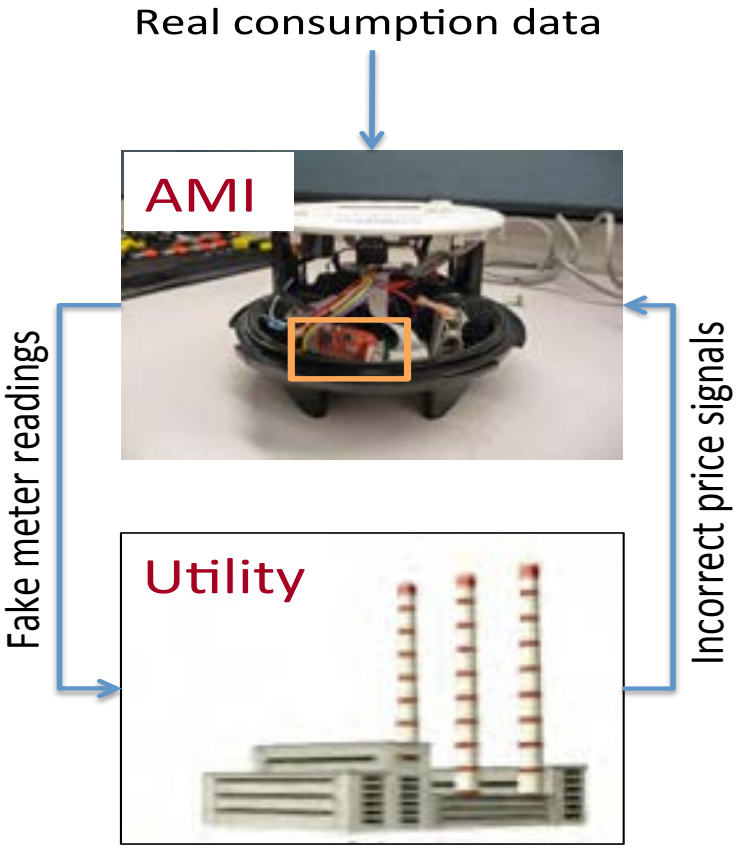
A man stands on a stepladder to fix tangled overhead electric power cables at a residential area in Noida, India, June 1, 2011 (Parivartan Sharma/Courtesy Reuters).

World Bank Reports: $\sim 30 - 50\%$ electricity is stolen in some jurisdictions

Deception attacks to AMIs

Stealthy attacker

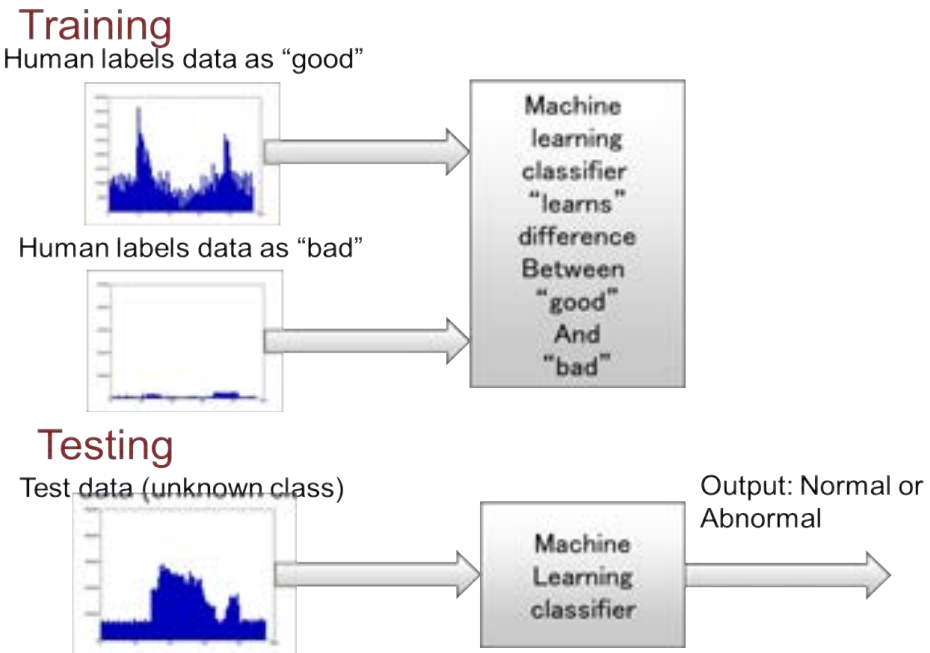
- Knows/learns CPS parameters
- Adapts to diagnosis algorithm
- Injects malicious data after obtaining unauthorized access
- Achieves his goal yet evade detection



- Real data: $Y_1, \ldots, Y_n$
- Fake readings: $\hat{Y}_1, \ldots, \hat{Y}_n$
- Attack model: $\hat{Y}_i = \hat{Y}_i + a_i$

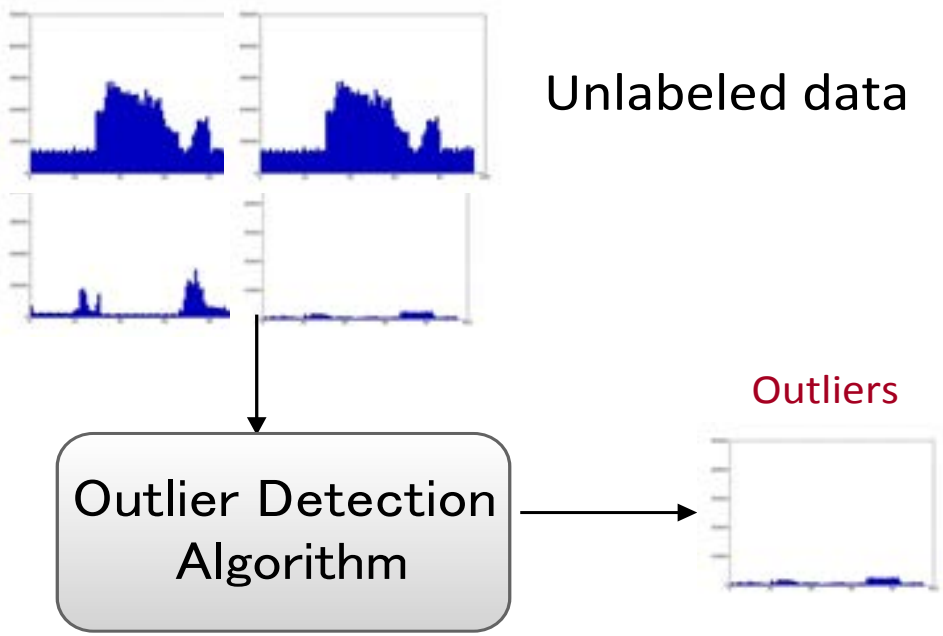
Previous work on energy usage profiles

Supervised learning



- Difficult to obtain “attack” data
- Difficult to generalize to new “smart” attacks

Unsupervised learning



- More false alarm rates
- Easier to attack and difficult to tune

Detection of stealthy attacks

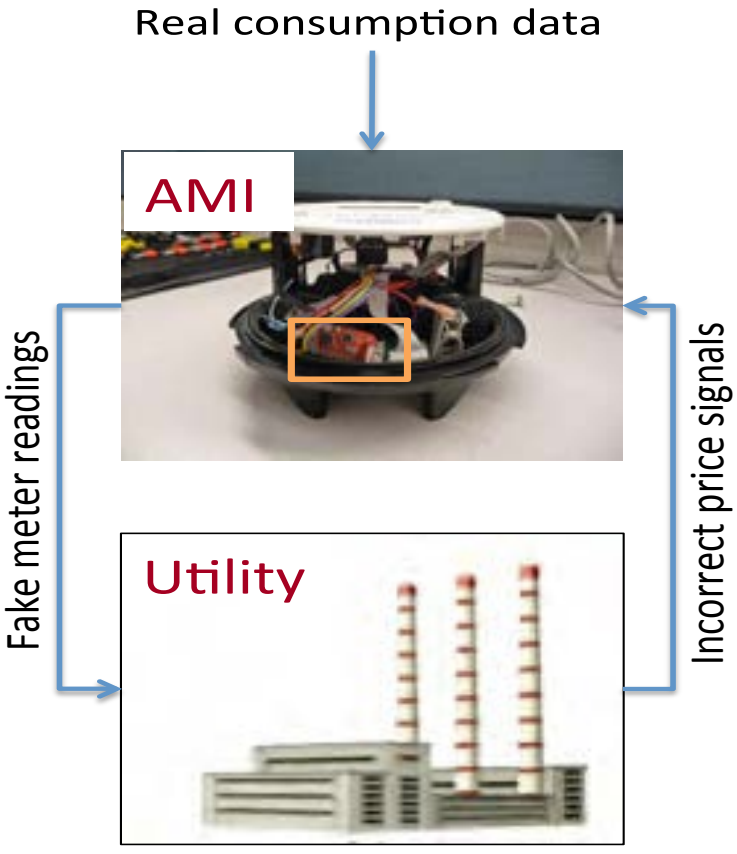
Adversary's goal

$$\min_{\hat{Y}_1,\ldots,\hat{Y}_n} f(\hat{Y}_1,\ldots,\hat{Y}_n)$$
$$g(\hat{Y}_1,\ldots,\hat{Y}_) \leqslant 0$$

E.g.: Minimize energy bill while not being detected by a classifier

Anomaly detection

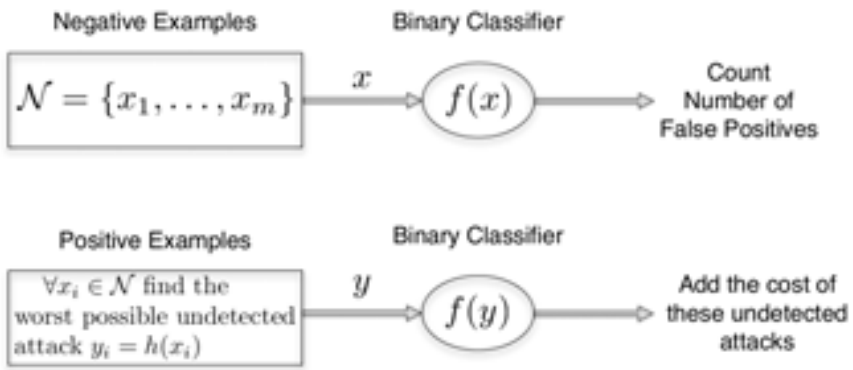
- Cumulative sum (CUSUM)
- Exponential weighted moving average (EWMA)
- Local outlier factor (LOF)
- Generalized likelihood ratio (GLR)



- Real data: $Y_1,\ldots,Y_n$
- Fake readings: $\hat{Y}_1,\ldots,\hat{Y}_n$
- Attack model: $\hat{Y}_i = \hat{Y}_i + a_i$

Learning with good data and attack invariants

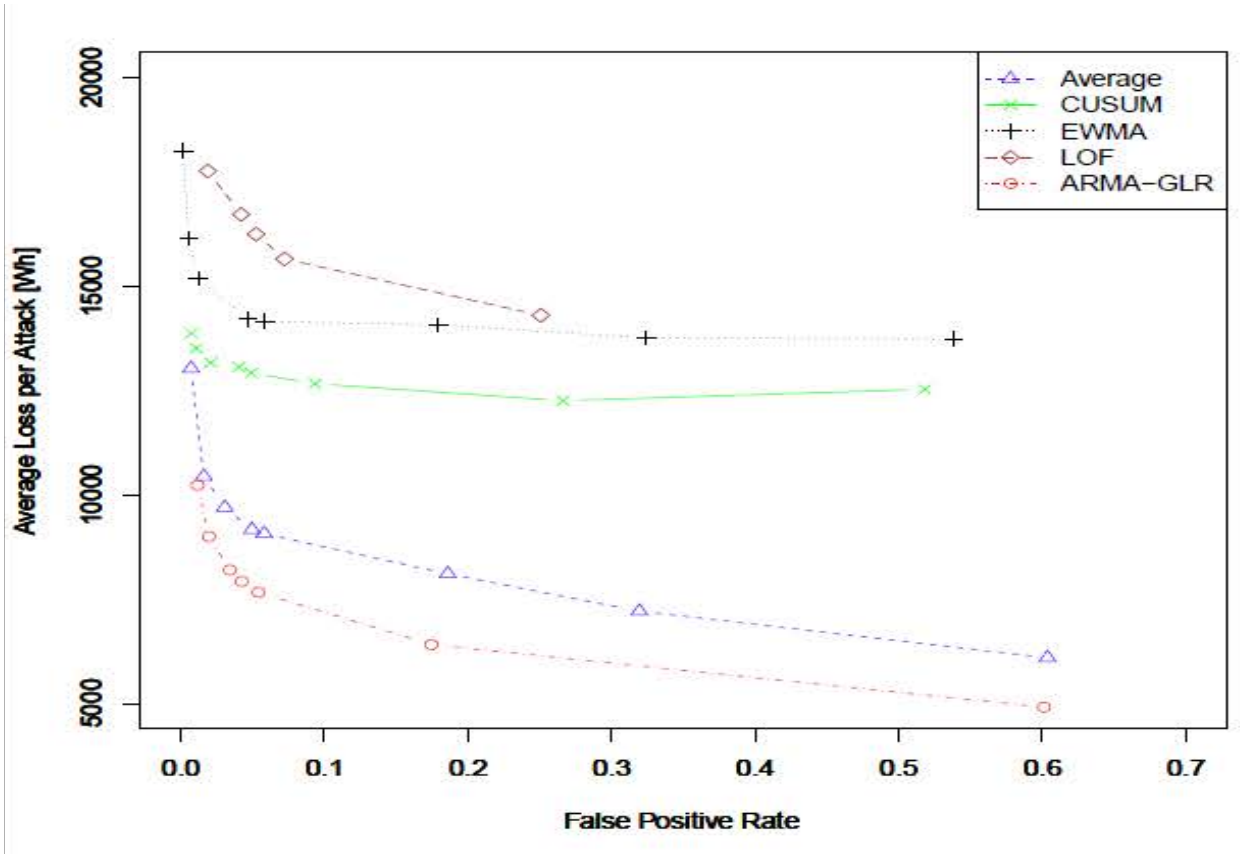
- We only have “good” data
  - No access to “attack” data
  - Train only one class (“good” data)
- We know “attack invariant”
  - Known attacker objective: minimize energy bill while not being detected by classifier
  - Use composite hypothesis testing to select attack probability distribution
- Find worst possible undetected attack for each classifier, and compute the corresponding cost (e.g., kWh lost).



A. Cárdenas et al. 2012

Evaluation

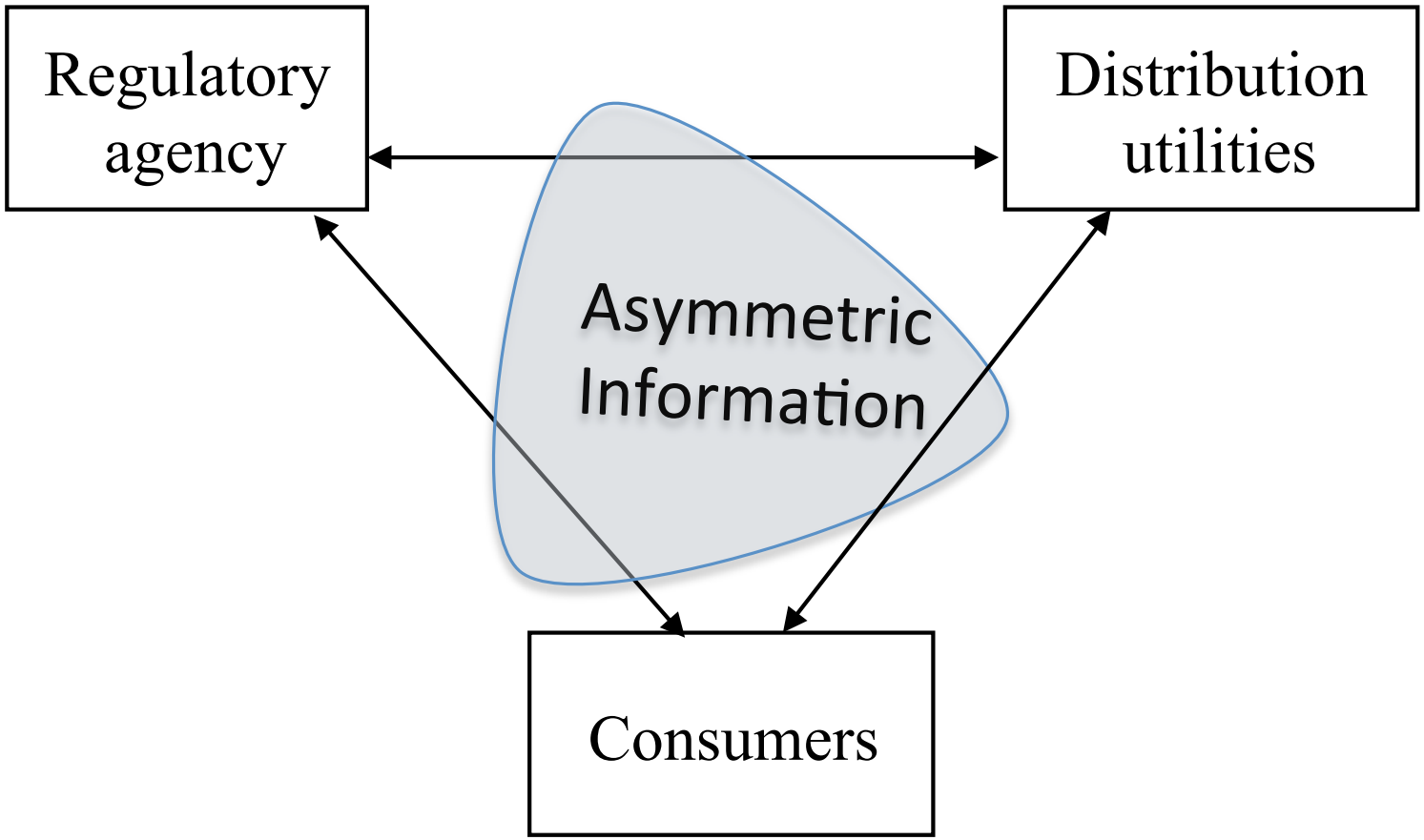
Autoregressive moving average (ARMA) based generalized likelihood ratio (GLR) provides maximum diagnostic ability for “stealthy attacks”. A. Cárdenas, et. al. ’12



Work in progress

- Attacker mistraining classifier
- Detect other anomalies (e.g., consumer on vacation)

Regulated electricity distribution: players



All parties have hidden (private) information. E.g: distributor knows his costs & consumer demand better than regulator.

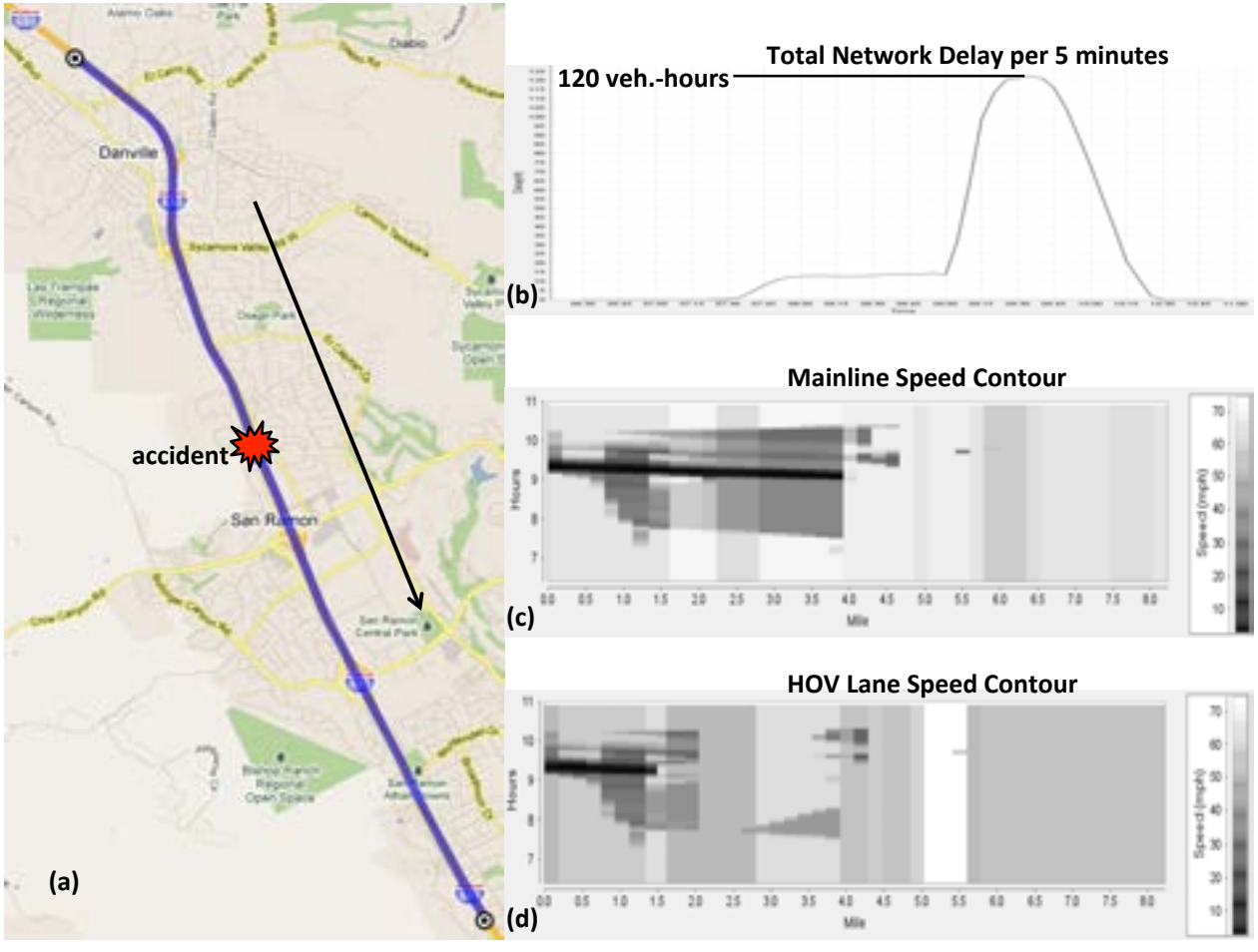
Incentive regulation for deploying diagnostic systems?

Outline

RC problem: Active management of traffic incidents

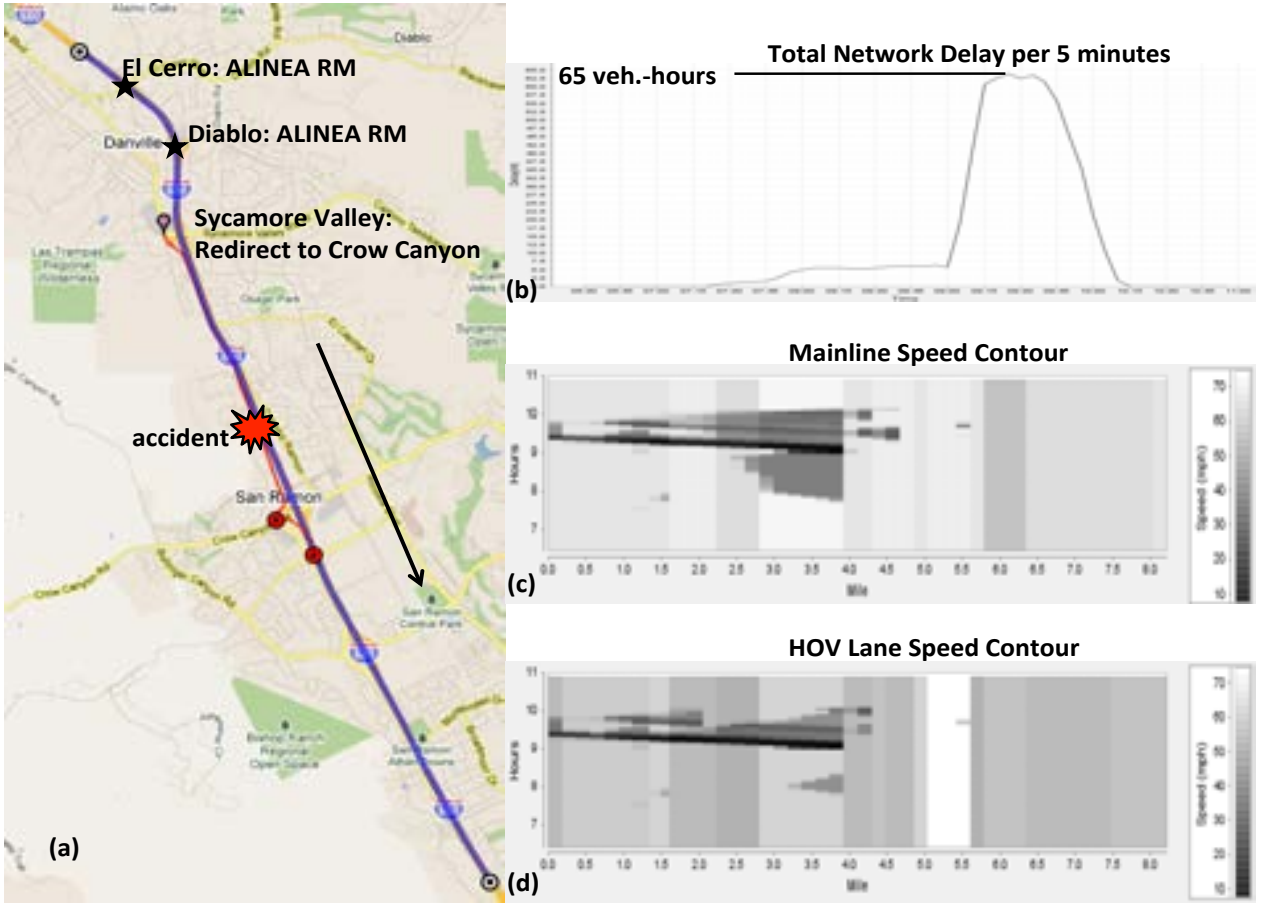
Random failure

CA Highway patrol report: Accident on I680 (11/15/2010) blocking two right lanes near post-mile 35 upstream of the offramp to Crow Canyon Rd.



RC Strategy I

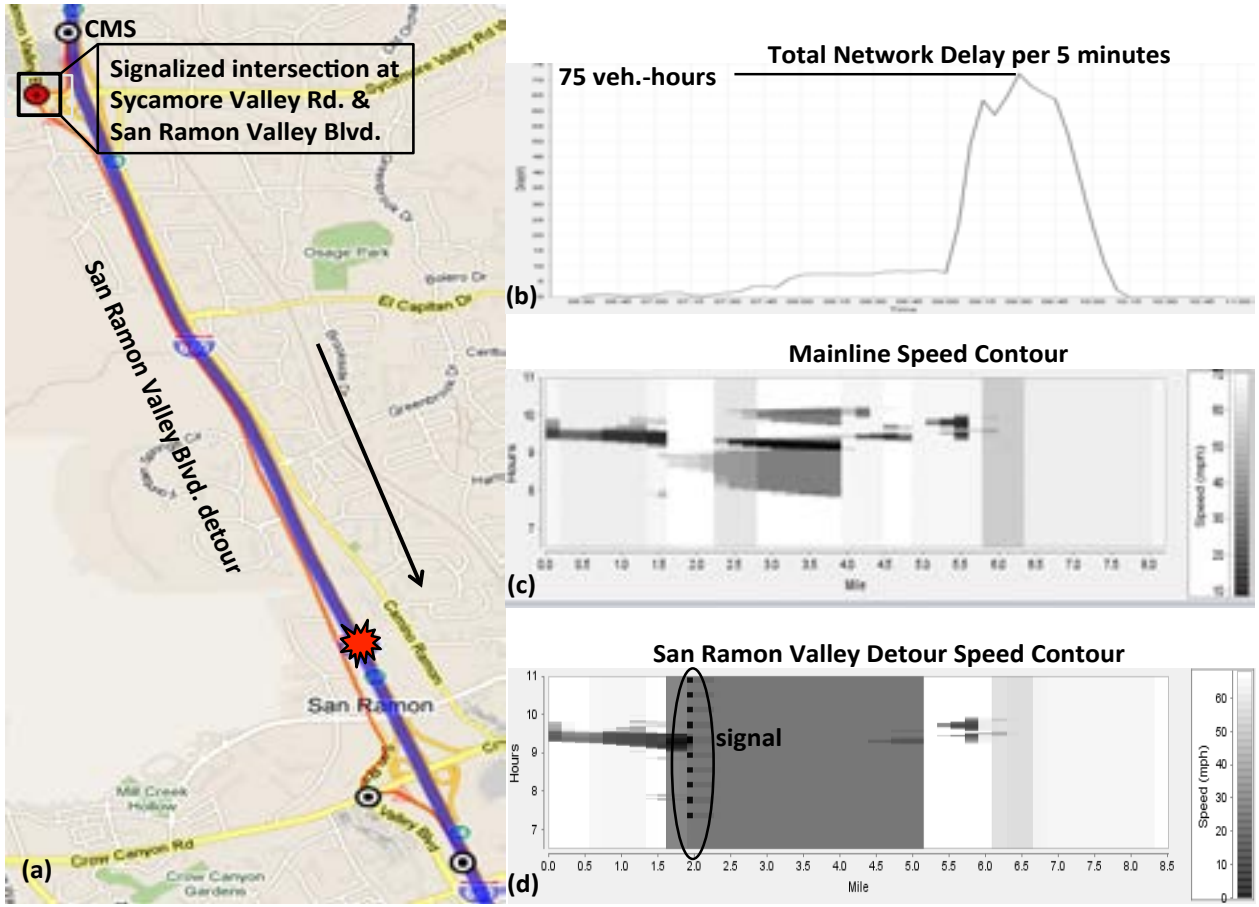
- Open HOV lane for everyone upstream of the accident
- Start ramp metering (queue control on El Cerro & Diablo ramps)
- Redirect traffic from Sycamore Valley Rd. to Crow Canyon Rd.



RC significantly improves performance and reliability

RC Strategy II

- Strategy I plus
- Diverge traffic to a parallel arterial (San Ramon Valley Boulevard) using changeable message signs (CMS)

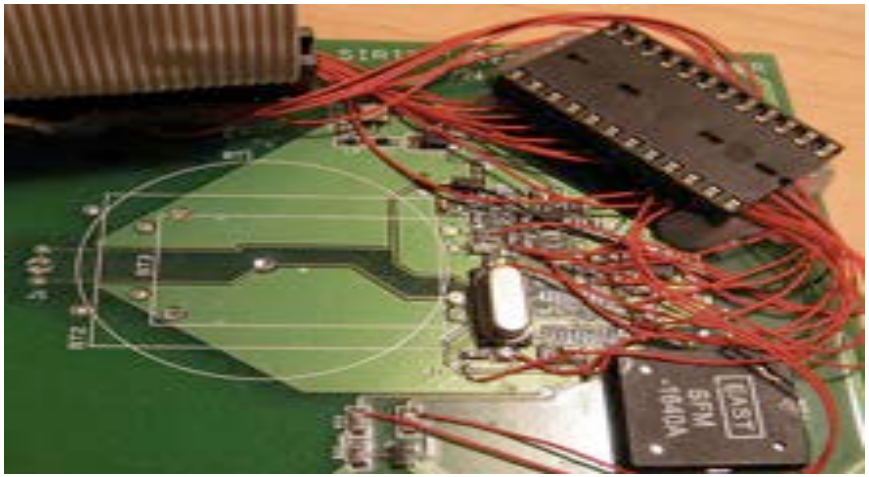


- Strategy I is expected to be better than Strategy II
- Freeway incident management requires coordination with arterial traffic management

Cyber-attacks to transportation infrastructures

Hackers: Road signs near MIT *(2008)*

Insiders: LA traffic control *(2008)*



Hackers: Tolling system*(2008)*

UCSD-UW Demo: Car hacking *(2011)*

Tools for real-time assurance

- Trustworthy information
  - Advisories (congestion levels) and alerts (incidents)
- Resilient control for safe and efficient operation under
  - (Non-)recurrent congestion
  - Incidents
  - DoS and deception attacks
- Operational strategy selection
  - Automatic control-based enforcement strategies
  - Pricing strategies to manage network congestion

Outline

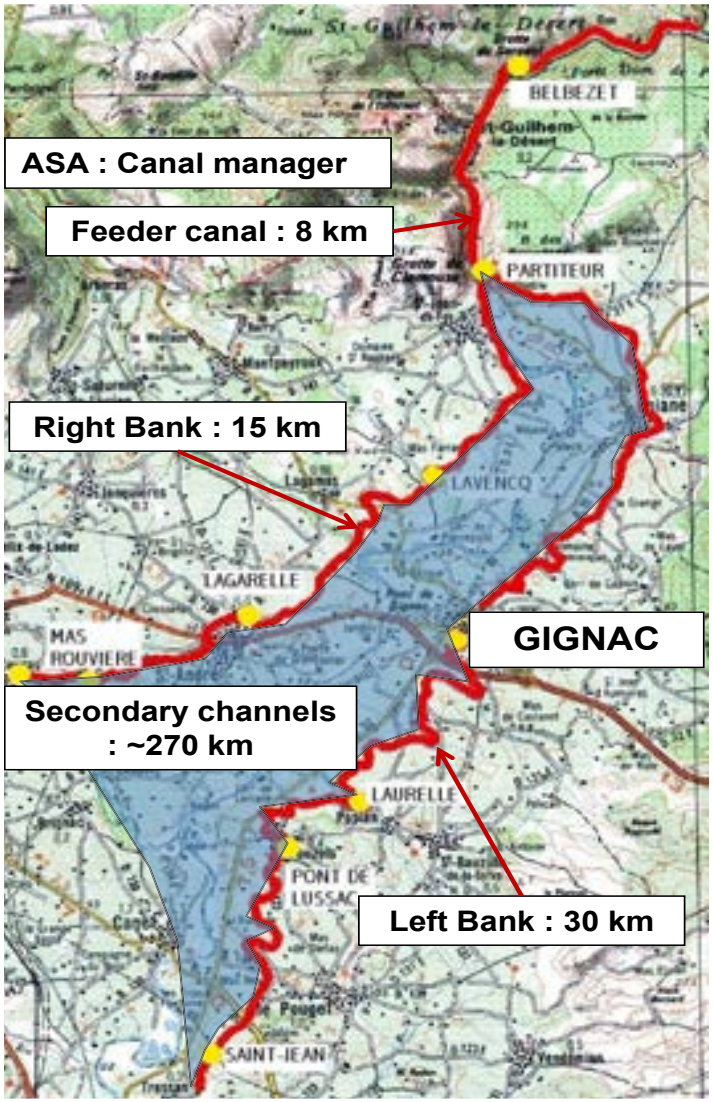
Gignac water SCADA system

SCADA components

- Level & velocity sensors
- PLCs & gate actuators
- Wireless communication

Gignac Le canal victime d'actes de vandalisme à répétition

Depuis le 21 juin, le canal de Gignac est victime d'actes malveillants sur l'ouvrage de l'aqueduc de l'Aurelle (derrière le lagunage de Popian) : effondrement du radier du canal puis dégradation des réparations mises en place (retrait des boulots de serrage, mettant gravement en péril la pérennité de l'aqueduc).
L'ouvrage de l'Aurelle permet la continuité du transport de l'eau vers les parcelles du périmètre irrigué situé sur les communes de Pouzols, Le Pouget, Tressan et Puilacher, soit près de 900 ha, pour lesquels l'apport d'eau estival est essentiel.
Ces agissements ont fait l'objet de constats par les brigades de gendarmerie et de plaintes contre X. Il est à noter que l'intégralité du patrimoine de l'Association syndicale autorisée du canal de Gignac est un ouvrage public, dont la destruction, la dégradation ou la détérioration peuvent faire l'objet de poursuites et être punies de trois ans d'emprisonnement et de 45 000 € d'amende.

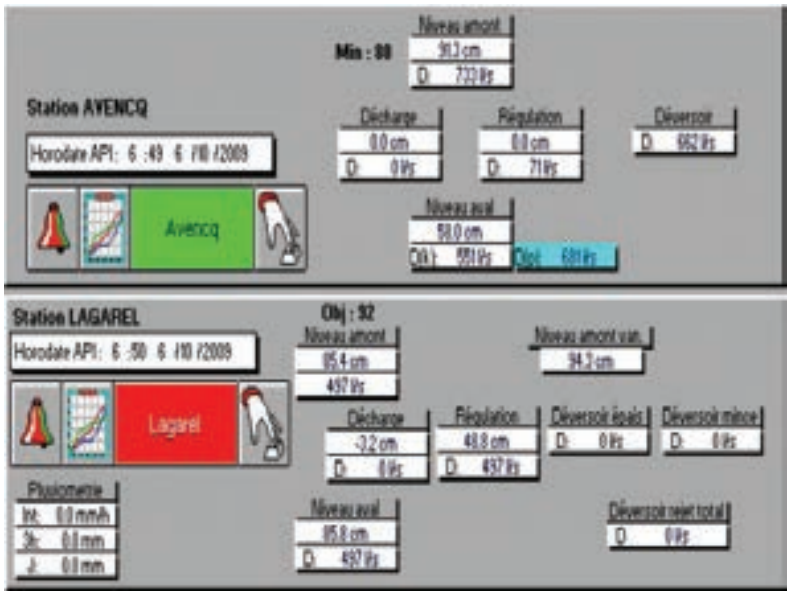


Regulatory control of canal pools

Control objective

- Manipulate gate opening
- Control upstream water level
- Reject disturbances (offtake withdrawals)

SCADA interface



Avencq cross-regulator

Defender and attacker models

Defender

- Estimate Model [Freq. Domain]

$$\hat{y}_i^d = \frac{a_i^d}{s} e^{-\tau_i s} \hat{q}_{i-1} - \frac{a_i^d}{s} [\hat{q}_i + \hat{p}_i]$$

Parameters: a_i^d, τ_i, Laplace variable: s

- Design robust decentralized PI control

$$\hat{q}_{i-1} = \kappa_{i-1\,i} \hat{y}_i^d, \quad \hat{q}_i = \kappa_{ii} \hat{y}_i^d$$

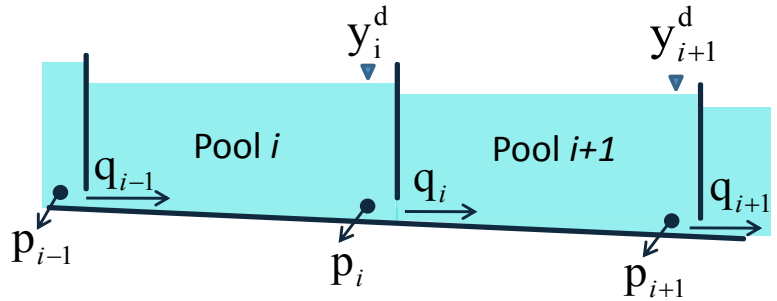
Controllers: $\kappa_{i-1\,i}, \kappa_{ii}$

Attacker

- Compromise y_i^d and inject g_i

$$\tilde{y}_i^d = y_i^d + g_i$$

- Regulate p_i to steal water

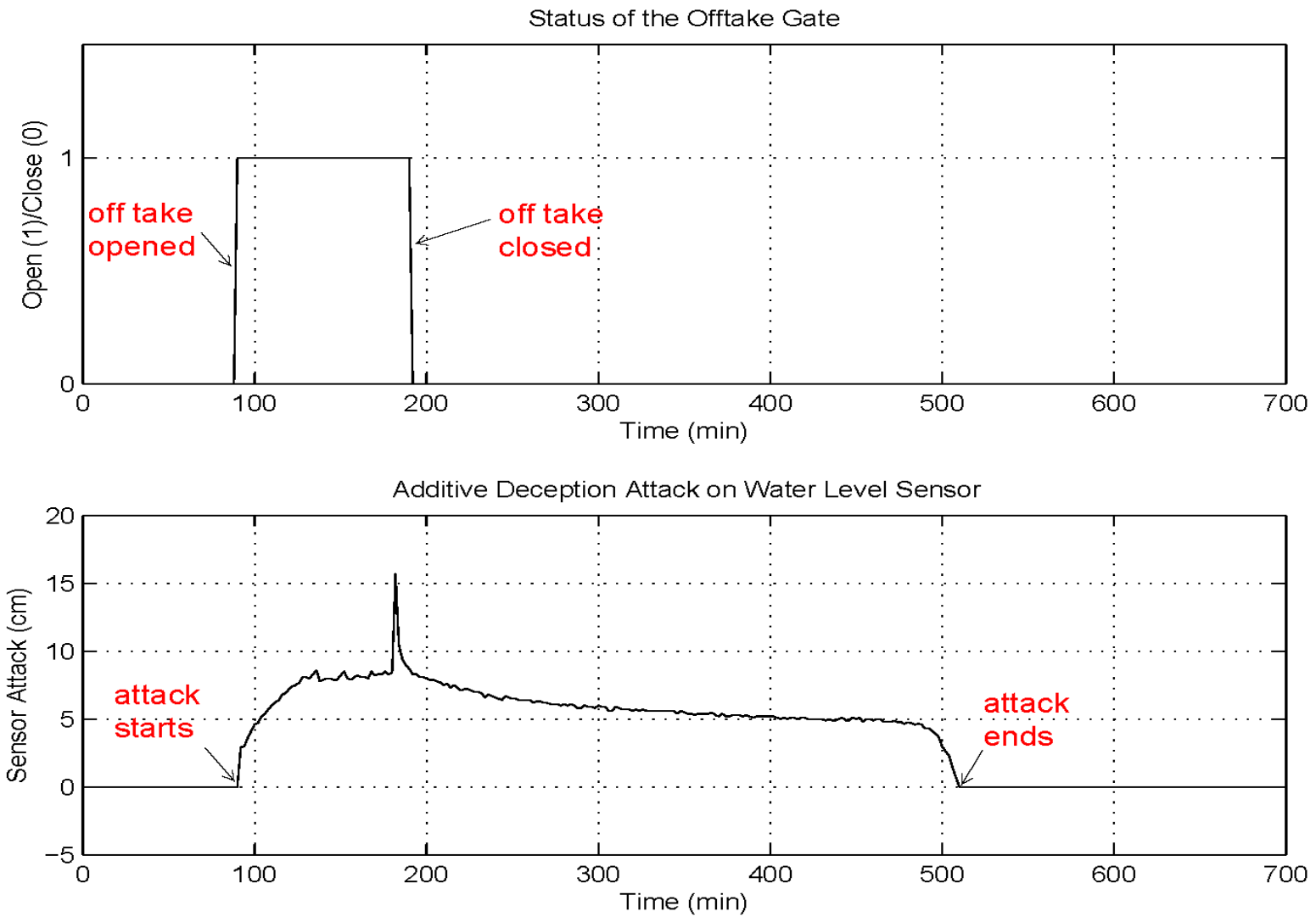


Test site before attack

Test site after attack

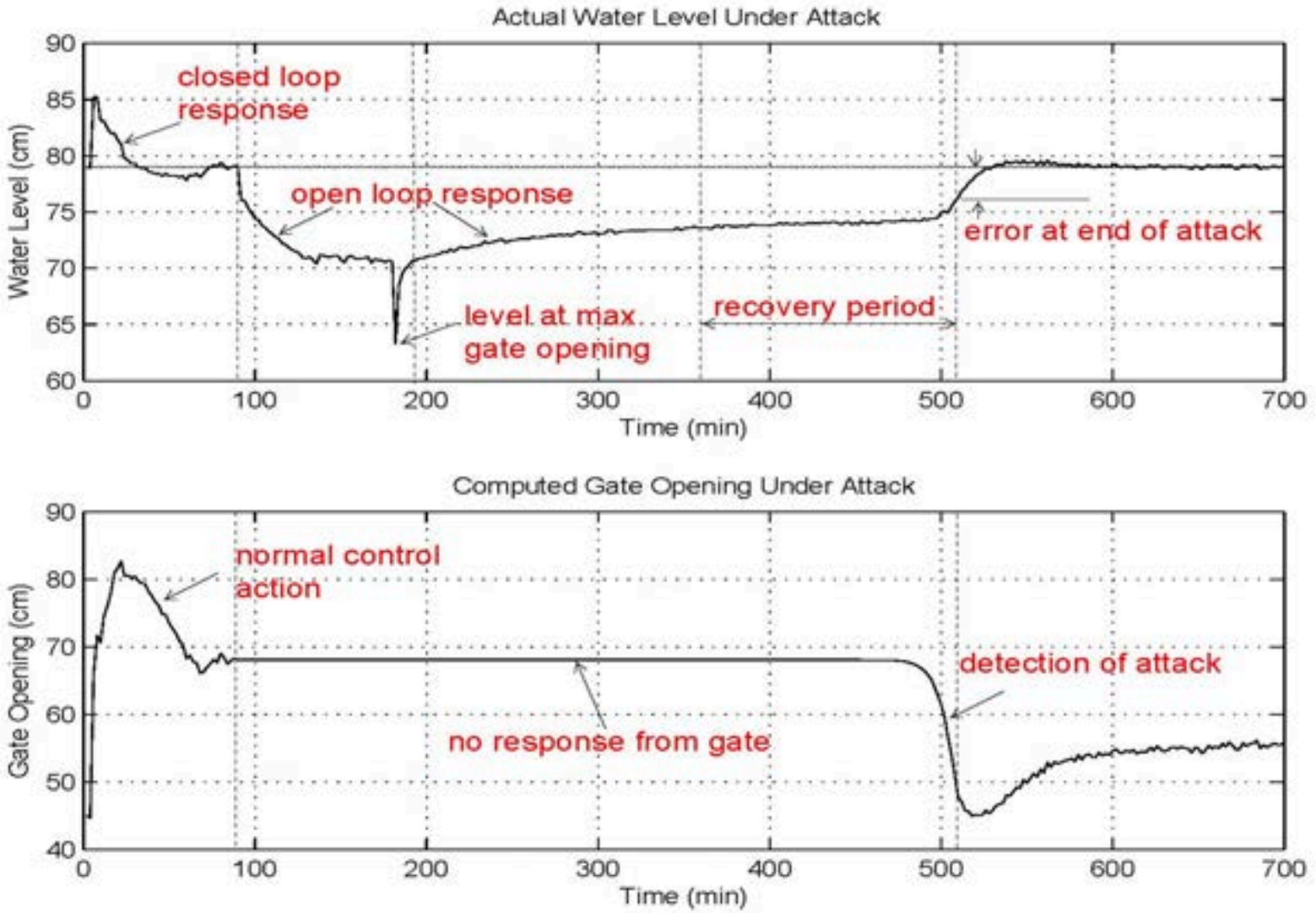
Cyber-attack on the Avencq canal pool

Field operational test (October 12th, 2009)



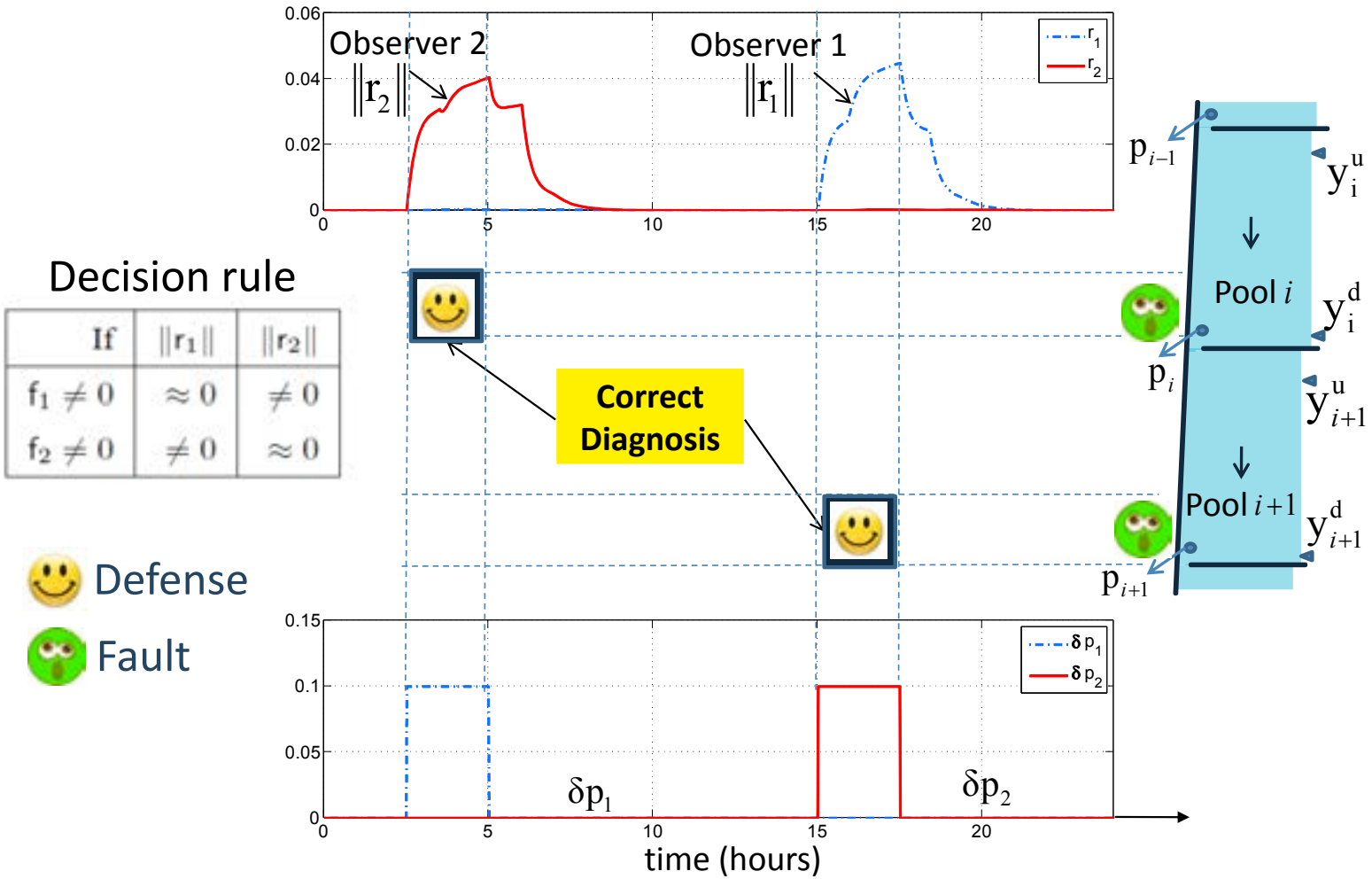
Cyber-attack on the Avencq canal pool

Successful attack

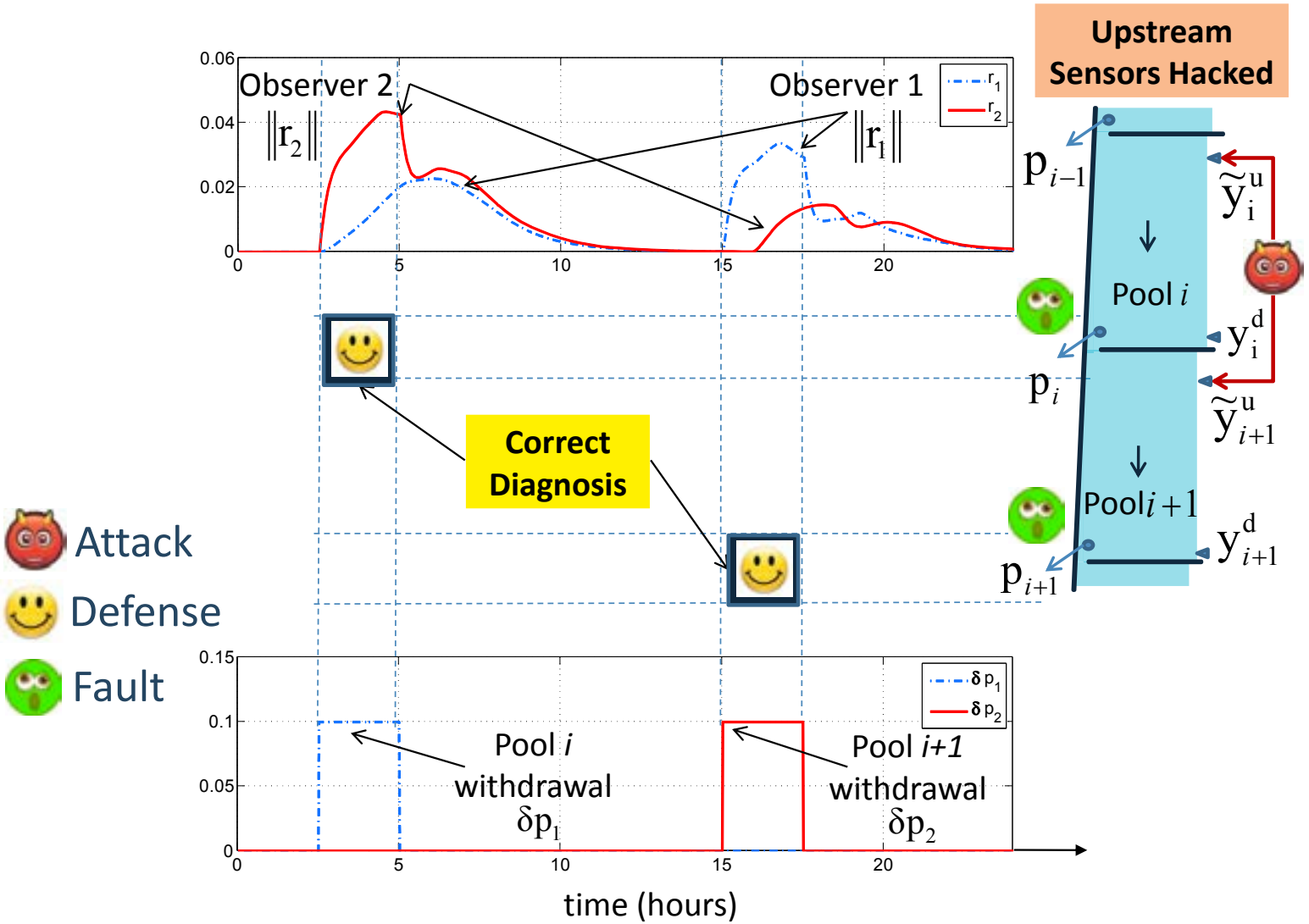


Model-based diagnosis scheme

Sensors: y^d_i, y^d_{i+1} and y^u_i, y^u_{i+1}

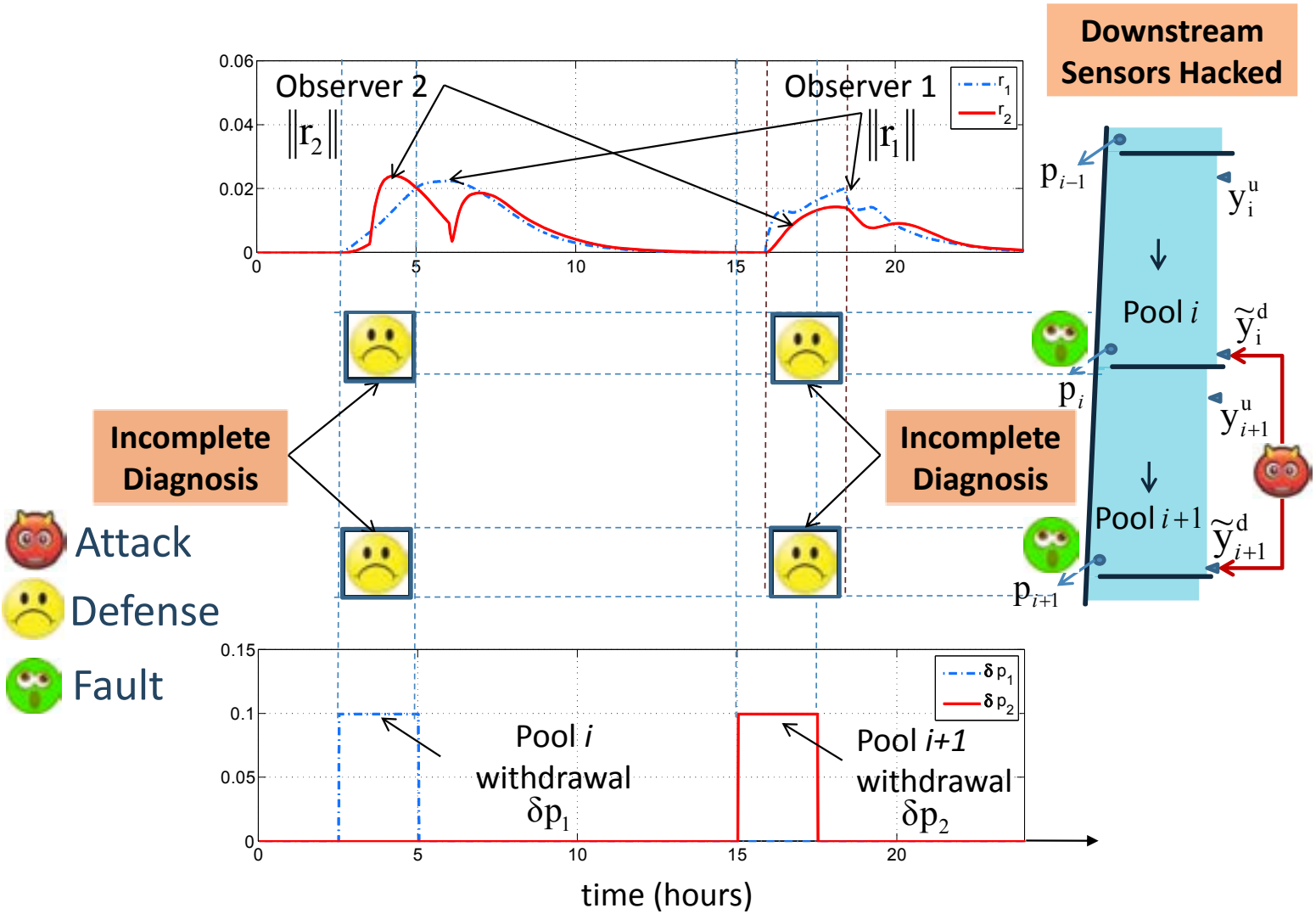


Attack diagnosis: upstream level sensors hacked



Correct diagnosis of withdrawal in both pools

Attack diagnosis: downstream level sensors hacked

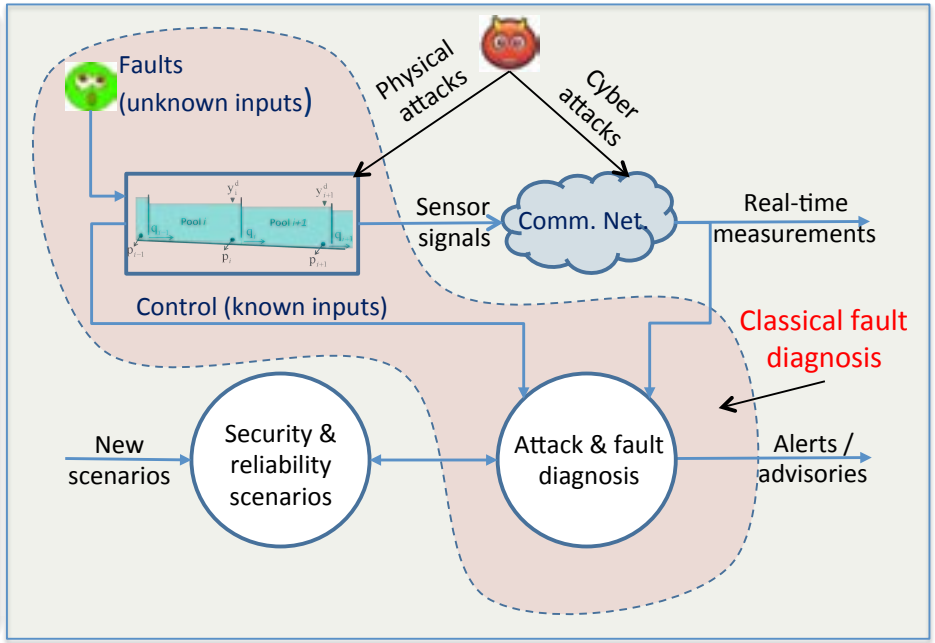


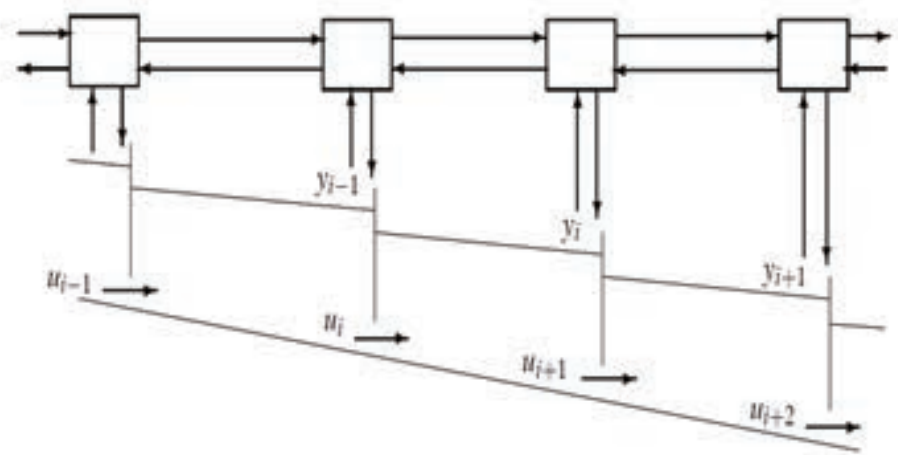
Withdrawal detected in both pools

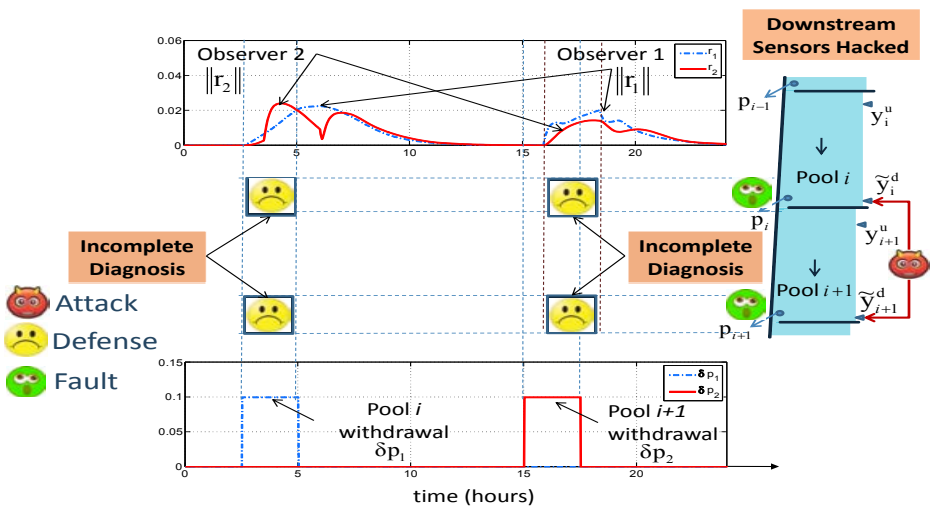
Model-based fault/attack diagnosis

Security implications

- Enhanced model (redundancy) improves detection
- Sensors located near offtakes are critical
- Localized sensor attacks do not lead to global degradation
- Multiple pool sensor attacks can evade detection [stealth]







Outline

Interdependent security (IDS) & incentives to secure

A problem of incentives

Due to presence of network-induced interdependencies, the individually optimal [Nash] security allocations are sub-optimal.

Interdependencies due to

- Network induced risks $\Rightarrow$ vulnerability to distributed DOS attacks
- Negative externalities
- Goal: Develop mechanisms to reduce CPS incentive sub-optimality

Courtesy: C. Goldschmidt (Symantec)

[Amin, Schwartz, Sastry, CDC '11, Automatica]

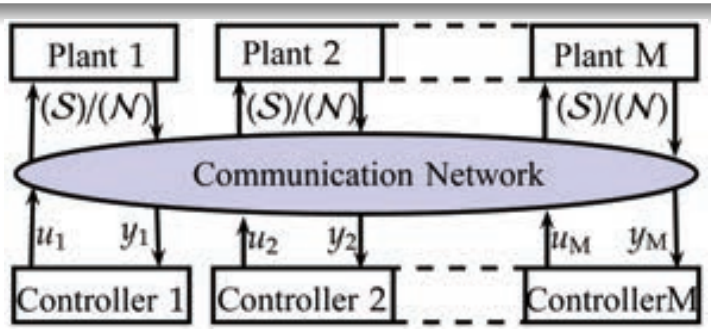
Interdependence for networked control systems (NCS)

NCS security & reliability

- Security failures (attacks S) & reliability failures (faults R) are difficult to distinguish
- Model for communication network failures F:

$$\begin{aligned}\Pr(S \cup R \mid F) &= \Pr(R \mid F) + \Pr(S \mid F) - \Pr(R \mid F)\Pr(S \mid F) \\ &= \underbrace{\Pr(R \mid F)}_{\text{direct failure (reliability)}} + \underbrace{(1 - \Pr(R \mid F))\Pr(S \mid F)}_{\text{indirect failure (security)}},\end{aligned}$$

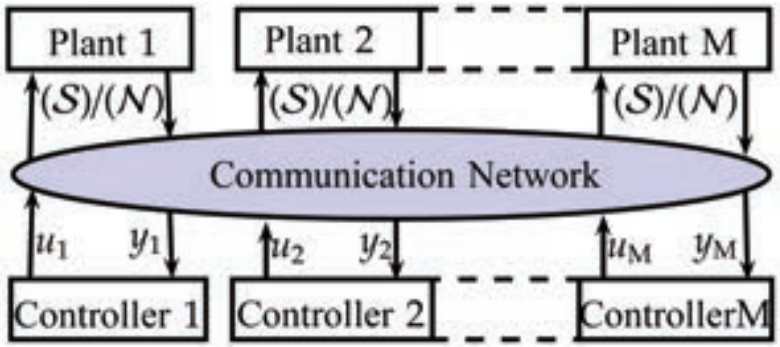
- Interdependence: $\Pr(S \mid F) = \alpha(\eta)$
  - $\alpha(\cdot)$: strictly increasing function
  - η: number of insecure players (NCS)



Network induced interdependencies

Environment summary

A game of $\mathbf{M}$ plant-controller systems (players)



For player i

1. (S) or (N) (Stage 1 choice variable)
   If (S) then i incurs per period security cost, $\ell^i \in [0, \infty)$

$$V^i := \begin{cases} S, & \text{player } i \text{ invests in security,} \\ N, & \text{player } i \text{ does not invest in security} \end{cases}$$

2. $u^i \in \mathbb{R}^m$ – control input (Stage 2 choice variable)

Model summary

Stage 1: Each player chooses (S) or (N)

- Failure probabilities depend on security choices
- Based on interdependent security model:

$$\Pr(\mathrm{S} \cup \mathrm{R} \mid \mathrm{F}) = \underbrace{\Pr(\mathrm{R} \mid \mathrm{F})}_{\text{direct failure (reliability)}} + \underbrace{(1 - \Pr(\mathrm{R} \mid \mathrm{F}))\Pr(\mathrm{S} \mid \mathrm{F})}_{\text{indirect failure (security)}},$$

- If (S), at each t, player incurs a per period (heterogeneous) security costs $\ell^i \in [0, \infty)$

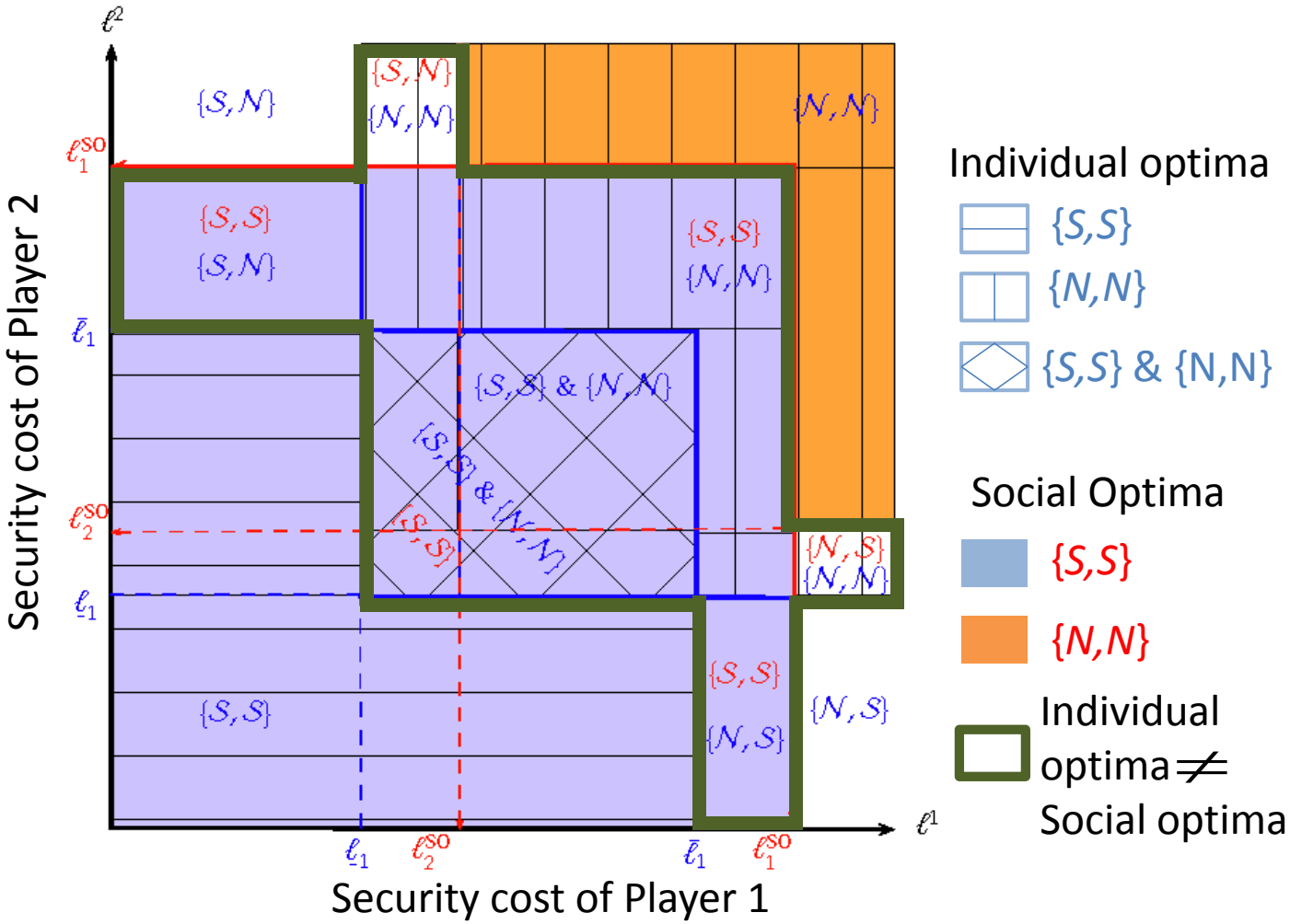
Stage 2: Each player is an operator of a NCS

- A standard model of NCS *and* unreliable communications

Individual optima [Nash equilibria] and social optima

Increasing incentive case

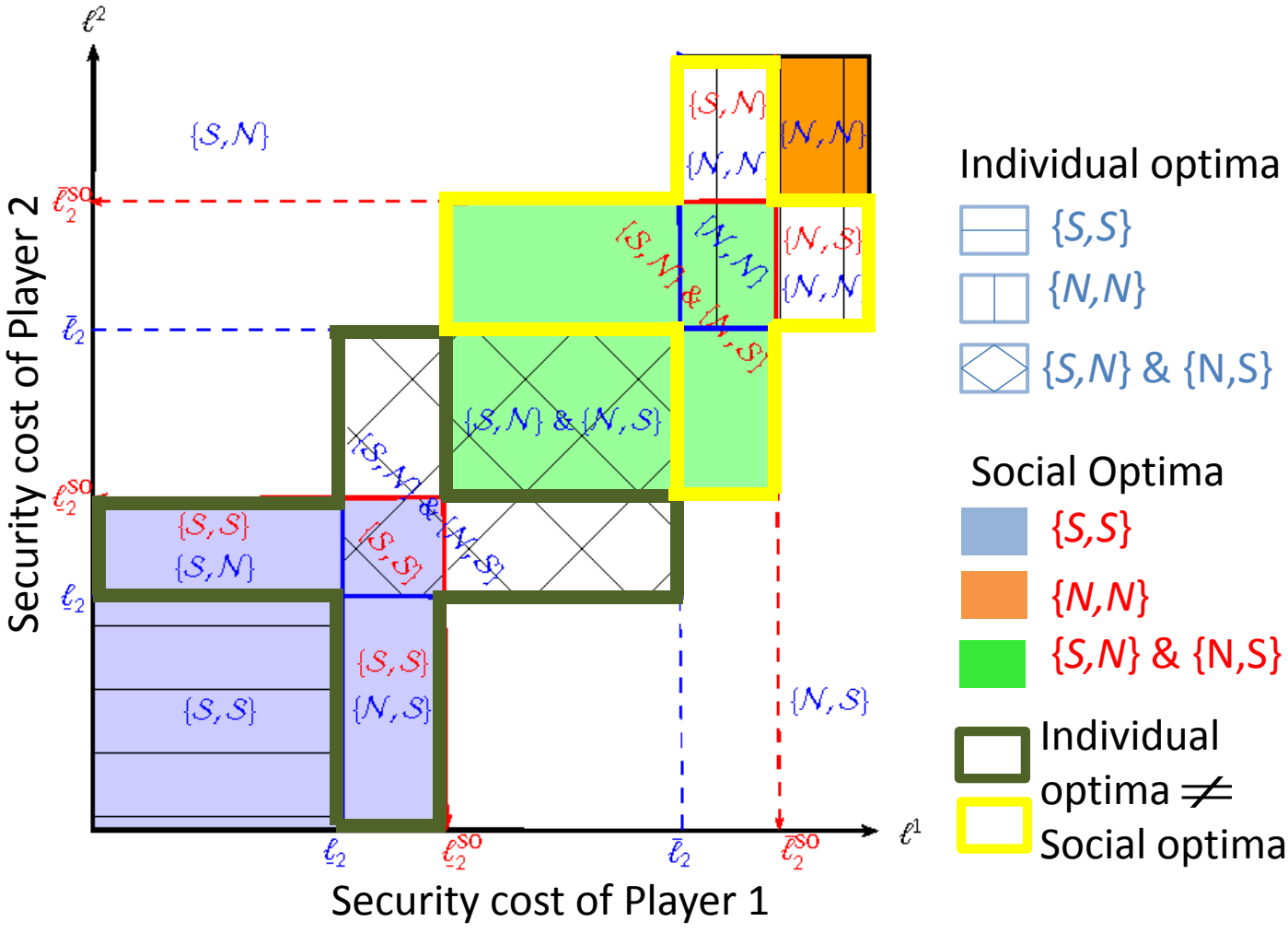
If a player secures, other player gain from securing *increases*



Individual optima [Nash equilibria] and social optima

Decreasing incentive case

If a player secures, other player gain from securing *decreases*



Outline

Current testbeds

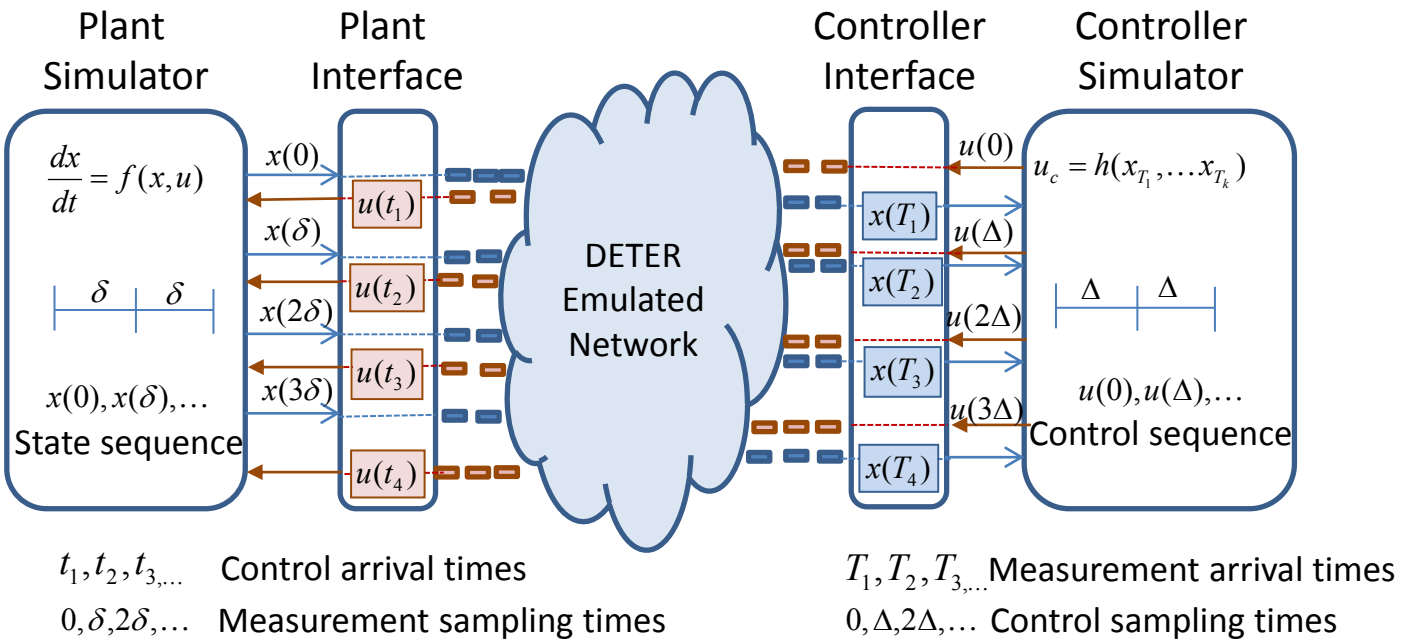
C2 Wind Tunnel

- Provides multi-modeling and simulation environment to evaluate performance of command and control (C2) centers
- Supports modeling of human performance and man-machine interaction

DEfense Test and Evaluation Research (DETER)

- Large scale testbed for simulation of internet attacks and defenses on complex networked systems
- Consists of approx. 300 computers and routers
- Used by DHS in year 2006 to simulate simultaneous attacks on critical infrastructures

DETER integration with CPS dynamics

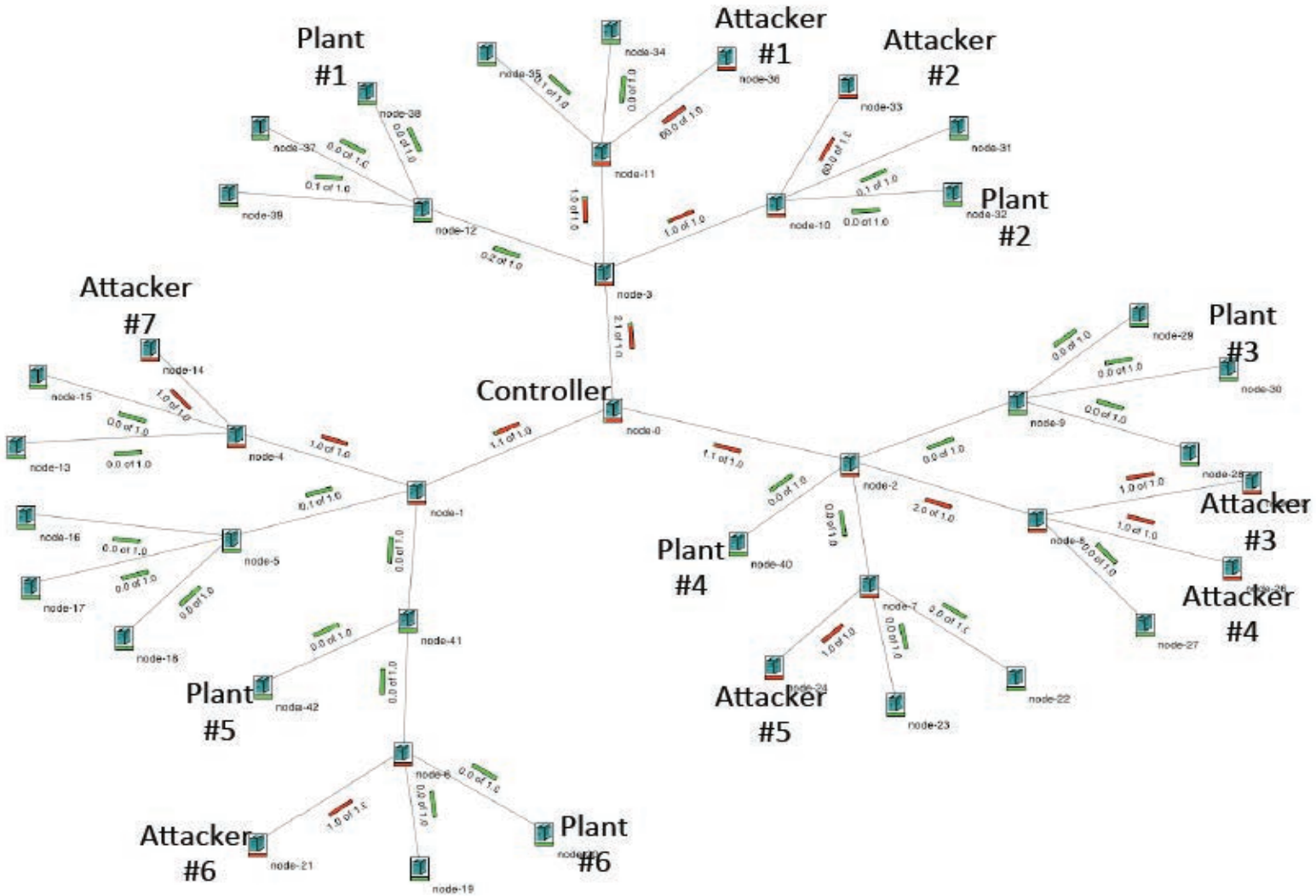


Effects of network topology and traffic

- Background [web] + foreground [control] + malicious [attack] traffic
- Plant-controller locations relative to compromised nodes
- Empirical distributions of delay, packet loss, and jitter

[A. Hussain, S. Amin '12]

CPS security experimentation using DETER



Multi-plant, single-controller CPS in spanning tree topology

FORCES: Looking forward