Games of Partial Information in Cyber-Physical Systems Security

João P. Hespanha

Work in collaboration with
Denis Garagic, Guosong Yang, Radha Poovendran

Outline

UC SANTA BARBARA engineering

Partial information games

- ➼ Sensor manipulation
  - Sensor-reveal game
  - Existence/computation of Nash equilibrium
  - Data-driven approach to detection

with D. Garajic (BAE)

[based on CDC'20, WeC09.4]

- ➼ Asymmetric information
  - online learning of the attacker's best response

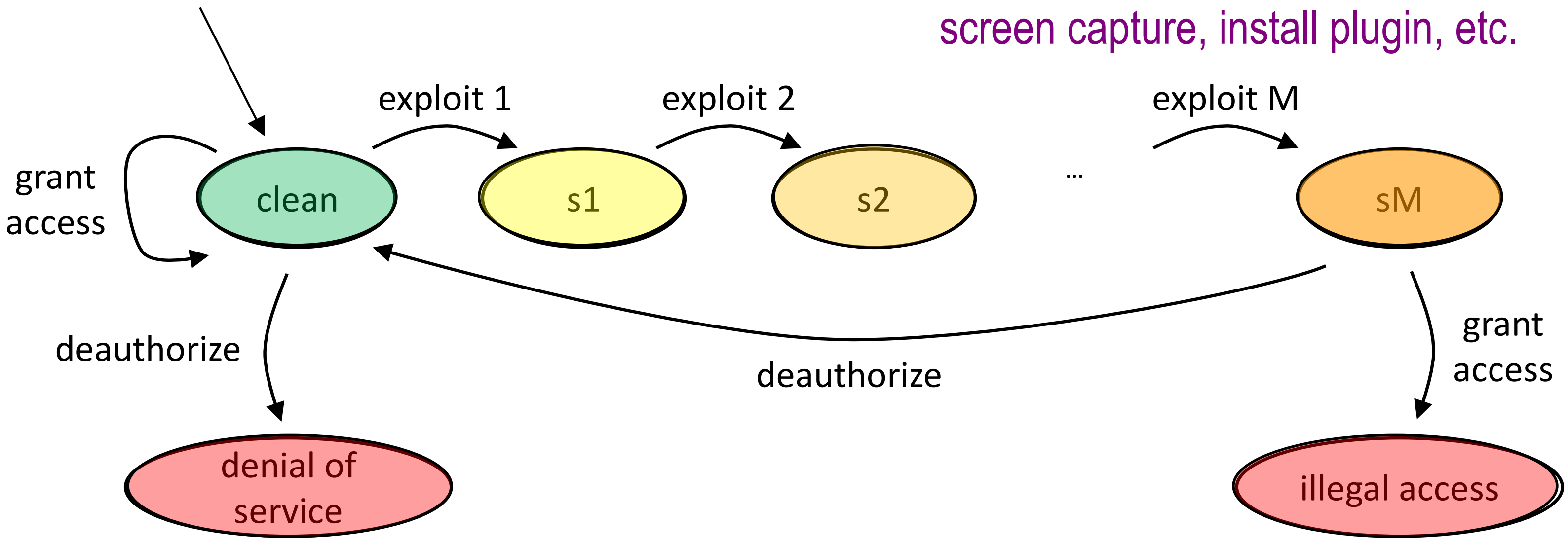
with G. Yang (UCSB),
R. Poovendran (UW)

[based on CDC'20, FrB09.3]

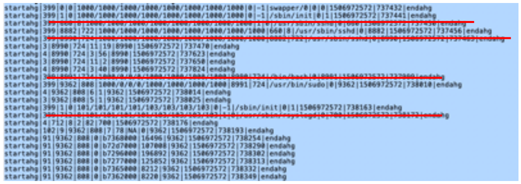
Motivation – Cyber attack detection

- Legitimate users can request access to a secure service by providing appropriate credentials.
- Attacker explores *sequence of exploits* to get access to credentials that would give her illegal access.

scan network ports, gain user access, escalade user to admin access, execute arbitrary code, screen capture, install plugin, etc.



- Attack detection based on sensor ("logs") that can be disabled by an attacker

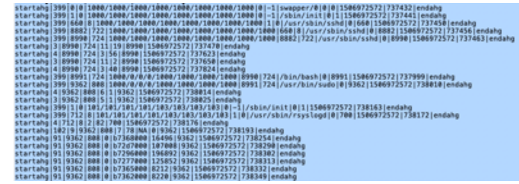


Bayesian Detection

UC SANTA BARBARA engineering

Problem formulation:

$$\theta = \begin{cases} 1 & \text{attack active} \\ 0 & \text{no attack} \end{cases}$$



cyber sensor logs

detection system wants to estimate $\hat{\theta}(y)$

to minimize $J_{\text{detec}} := A\,\mathrm{P}(\hat{\theta} = 1, \theta = 0) + B\,\mathrm{P}(\hat{\theta} = 0, \theta = 1)$

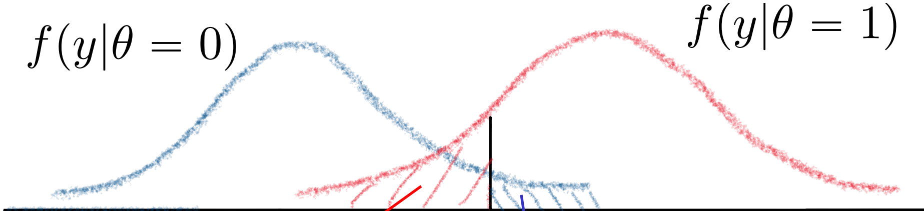
cost of false detections

cost of missed detections

classical Bayesian detection
(given a-priori distribution of θ)

1-D example:

$f(y|\theta = 0)$ $f(y|\theta = 1)$



L

$p_{\text{fn}}(L) := \mathrm{P}(\hat{\theta} = 0|\theta = 1)$

$p_{\text{fp}}(L) := \mathrm{P}(\hat{\theta} = 1|\theta = 0)$

Pick *L* in detector

$$\hat{\theta} = \begin{cases} 0 & y \leqslant L \\ 1 & y > L \end{cases}$$

to minimize

$$A\,\mathrm{P}(\theta = 0)p_{\text{fp}}(L) + B\,\mathrm{P}(\theta = 1)p_{\text{fn}}(L)$$

Adversarial Detection

Problem formulation:

$$\theta = \begin{cases} 1 & \text{attack active} \\ 0 & \text{no attack} \end{cases}$$

cyber sensor logs

detection system wants to estimate $\hat{\theta}(y)$

to minimize $J_{\text{detec}} := A\,\mathrm{P}(\hat{\theta} = 1, \theta = 0) + B\,\mathrm{P}(\hat{\theta} = 0, \theta = 1)$

cost of false detections

cost of missed detections

attacker wants to maximize

$$J_{\text{attack}} := R\,\theta - C\,\mathrm{P}(\hat{\theta} = 1, \theta = 1) + F\,\mathrm{P}(\hat{\theta} = 1, \theta = 0)$$

reward for attack

penalty for getting caught

reward for false detections (compromise confidence on detection mechanism)

classical Bayesian detection (given a-priori distribution of θ)

adversarial detection (attacker decides distribution of θ)

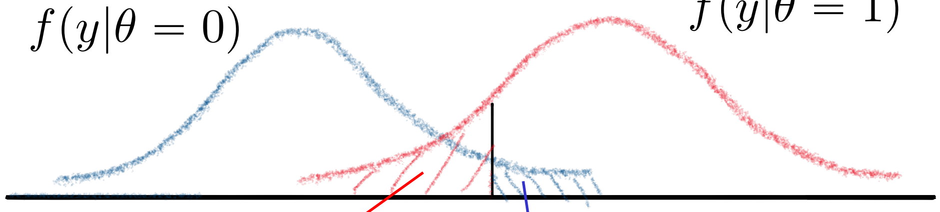
Adversarial Detection

Problem formulation:

$$\theta = \begin{cases} 1 & \text{attack active} \\ 0 & \text{no attack} \end{cases}$$

1-D example:

$f(y|\theta = 0)$ $f(y|\theta = 1)$



L

$p_{\text{fn}}(L) \coloneqq \mathrm{P}(\hat{\theta} = 0 | \theta = 1)$

$p_{\text{fp}}(L) \coloneqq \mathrm{P}(\hat{\theta} = 1 | \theta = 0)$

Detector picks *L* in

$$\hat{\theta} = \begin{cases} 0 & y \leqslant L \\ 1 & y > L \end{cases}$$

to minimize

$$A\,\mathrm{P}(\theta = 0) p_{\text{fp}}(L) + B\,\mathrm{P}(\theta = 1) p_{\text{fn}}(L)$$

attacker wants to select $\mathrm{P}(\theta = 1)$ to maximize (mixed policy)

$$J_{\text{attack}} \coloneqq R\,\mathrm{P}(\theta = 1) - C\big(1 - p_{\text{fp}}(L)\big)\,\mathrm{P}(\theta = 1) + F p_{\text{fn}}(L)\,\mathrm{P}(\theta = 0)$$

reward for attack

penalty for getting caught

reward for false detections

Adversarial Detection

Problem formula

$\theta =$

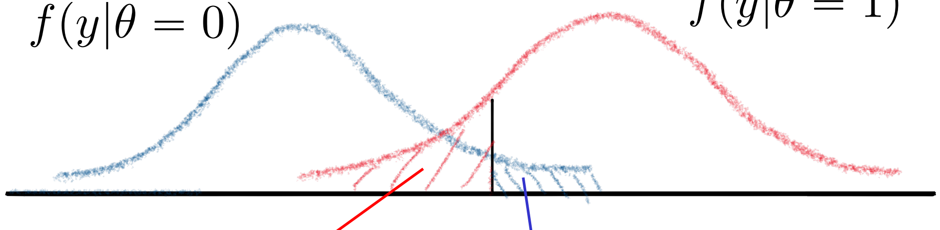
- Classical/adversarial formulations assume given & fixed distribution of data y (given θ)

but…

- In attacks detection, opponent will likely manipulate data (e.g., disable logs, engage in “diversions”)

1-D example:

$f(y|\theta = 0)$ $\quad f(y|\theta = 1)$



L

$p_{\text{fn}}(L) \coloneqq \mathrm{P}(\hat{\theta} = 0|\theta = 1)$

$p_{\text{fp}}(L) \coloneqq \mathrm{P}(\hat{\theta} = 1|\theta = 0)$

Detector picks *L* in

$$\hat{\theta} = \begin{cases} 0 & y \leqslant L \\ 1 & y > L \end{cases}$$

to minimize

$$A\,\mathrm{P}(\theta = 0)p_{\text{fp}}(L) + B\,\mathrm{P}(\theta = 1)p_{\text{fn}}(L)$$

attacker wants to select $\mathrm{P}(\theta = 1)$ to maximize (mixed policy)

$$J_{\text{attack}} \coloneqq R\,\mathrm{P}(\theta = 1) - C\big(1 - p_{\text{fp}}(L)\big)\,\mathrm{P}(\theta = 1) + F p_{\text{fn}}(L)\,\mathrm{P}(\theta = 0)$$

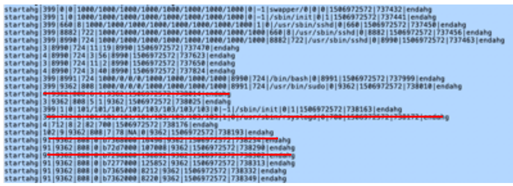
reward for attack

penalty for getting caught

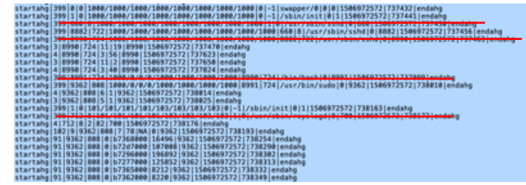
reward for false detections

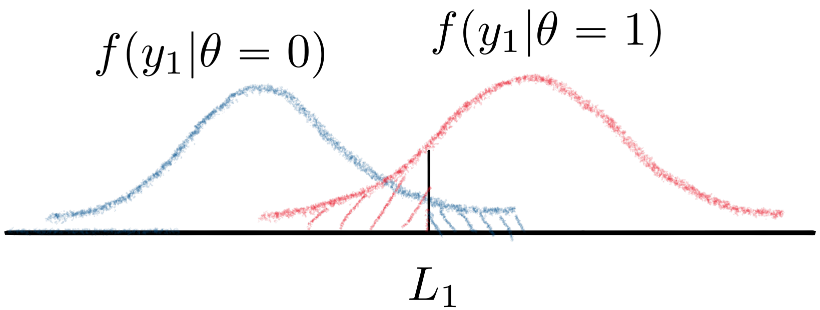
Detection with data manipulation

Problem formulation:

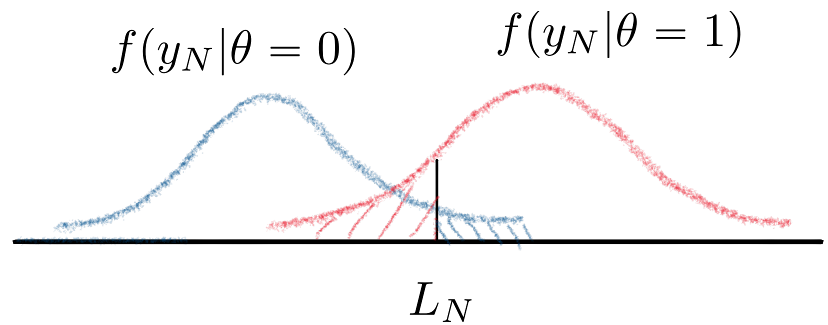


$$\theta = \begin{cases} 1 & \text{attack active} \\ 0 & \text{no attack} \end{cases}$$





…



attacker selects values for

$\theta \in \{0,1\}$ = whether or not to attack (typically mixed policy)

$\sigma \in \{1,2,\dots,N\}$ = which sensor to reveal

detector picks set $Y_1, Y_2, \dots, Y_N$ for each sensor and sets

$$\hat{\theta} = \begin{cases} 1 & y_\sigma \in Y_\sigma \\ 0 & y_\sigma \notin Y_\sigma \end{cases}$$

in 1D example,
$Y_i = [L_i, \infty)$

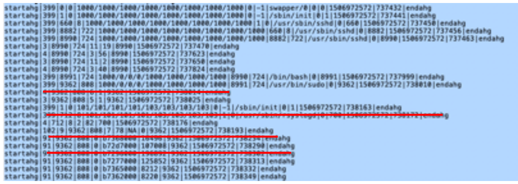
Detection with data manipulation

UC SANTA BARBARA engineering

Problem formulation:

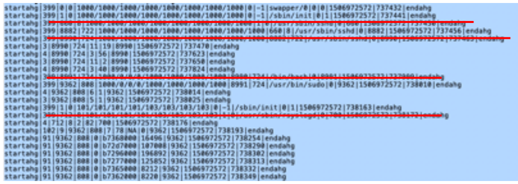
$$\theta = \begin{cases} 1 & \text{attack active} \\ 0 & \text{no attack} \end{cases}$$

sensor selection allows attacker to influence



can be increased by disabling logs

$$p_{\text{fn}} = \mathrm{P}(\hat{\theta}(Y) = 0 | \theta = 1)$$



may be increased by creating “diversions”

$$p_{\text{fp}} = \mathrm{P}(\hat{\theta}(Y) = 1 | \theta = 0)$$

attacker wants to maximize

data manipulation cost (e.g., added risk of being caught by disabling sensors)

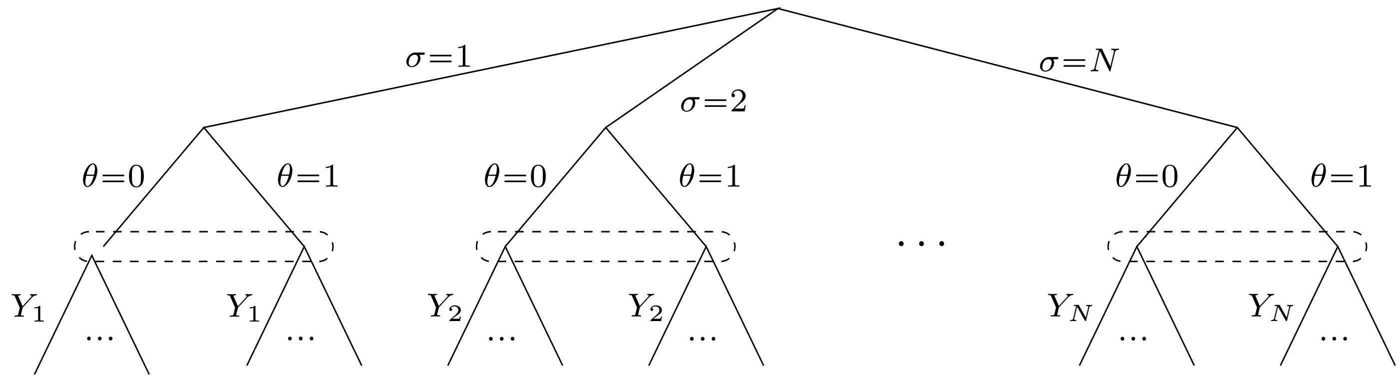
$$J_{\text{attack}} \coloneqq R\,\theta - C\,\mathrm{P}(\hat{\theta} = 1, \theta = 1) + F\,\mathrm{P}(\hat{\theta} = 1, \theta = 0) - S(p_{\text{fn}}, p_{\text{fp}})$$

attacker maximizes reward by “tuning” data manipulation to maximize reward

Extensive Form Representation

attacker selects sensor σ

attacker selects θ



defender selects detection sets Y_i (knowing σ but not θ)

$$\hat{\theta} = \begin{cases} 1 & y_\sigma \in Y_\sigma \\ 0 & y_\sigma \notin Y_\sigma \end{cases}$$

On each branch $\sigma = i$, we have a 2-player nonzero sum game

attacker plays rows:

$$A^i_{\text{att}} := \begin{bmatrix} -S_i & F-S_i & Fp^i_{\text{fp}}(\mathcal{Y}^1_i)-S_i & \cdots & Fp^i_{\text{fp}}(\mathcal{Y}^M_i)-S_i \\ R-S_i & R-C & S_i-R+C\left(1-p^i_{\text{fn}}(\mathcal{Y}^1_i)\right)-S_i & \cdots & R-C\left(1-p^i_{\text{fn}}(\mathcal{Y}^M_i)\right)-S_i \end{bmatrix} \begin{array}{l} \leftarrow \text{selects } \theta = 0 \\ \leftarrow \text{selects } \theta = 1 \end{array}$$

$$B^i_{\text{def}} := \begin{bmatrix} 0 & A & Ap^i_{\text{fp}}(\mathcal{Y}^1_i) & \cdots & Ap^i_{\text{fp}}(\mathcal{Y}^M_i) \\ B & 0 & Bp^i_{\text{fn}}(\mathcal{Y}^1_i) & \cdots & Np^i_{\text{fp}}(\mathcal{Y}^M_i) \end{bmatrix} \begin{array}{l} \leftarrow \text{selects } \theta = 0 \\ \leftarrow \text{selects } \theta = 1 \end{array}$$

defender plays columns:

$Y_i = \emptyset$ always pick $\hat{\theta} = 0$

$Y_i =$ "*all*" always pick $\hat{\theta} = 1$

intermediate options for Y_i:

$\mathcal{Y}^1_i, \mathcal{Y}^2_i, \ldots, \mathcal{Y}^M_i$

(finite enumeration for simplicity)

Main Results

Theorem: Consider only 3 detection sets (∅, "all", Y_i)

For each sensor selection $\sigma = i$, the game has mixed Nash eq. with detection policy of the form

$$z_{\text{def}}^{*} = \begin{cases} \begin{bmatrix} \frac{\bar{C}^i - R}{\bar{C}^i} \\ 0 \\ \frac{R}{\bar{C}^i} \end{bmatrix} & \bar{C}^i \geqslant R \\ \begin{bmatrix} 0 \\ \frac{R - \bar{C}^i}{C+F-\bar{C}^i} \\ \frac{C+F-R}{C+F-\bar{C}^i} \end{bmatrix} & \bar{C}^i < R \end{cases}$$

ignore Y and declare no attack

ignore Y and declare attack

Y_i-dependent decision

detection cost:

$$J_{\text{def}}^{i\,*} = \begin{cases} \dfrac{AB}{A + B\frac{1-p_{\text{fn}}^i}{p_{\text{fp}}^i}}, & \bar{C}^i \geqslant R \\ \dfrac{AB}{B + A\frac{1-p_{\text{fp}}^i}{p_{\text{fn}}^i}}, & \bar{C}^i < R \end{cases}$$

$$\bar{C}^i := C(1 - p_{\text{fn}}^i) + F p_{\text{fp}}^i$$

detection-set Y_i selected to maximize

$$\begin{cases} \frac{1-p_{\text{fn}}^i}{p_{\text{fp}}^i}, & \bar{C}^i \geqslant R \\ \frac{p_{\text{fn}}^i}{1-p_{\text{fp}}^i}, & \bar{C}^i < R, \end{cases}$$

very different from Bayesian case, where detection-set Y_i is selected to maximize

$$A\,\mathrm{P}(\theta = 0) p_{\text{fp}} + B\,\mathrm{P}(\theta = 1) p_{\text{fn}}$$

but not surprising because now attack probabilities adjust to p_{fp}, p_{fn}

Main Results

UC SANTA BARBARA engineering

Theorem: Consider only 3 detection sets ($\emptyset$, "all", Y_i)

For each sensor selection $\sigma = i$, the game has mixed Nash eq. with detection policy of the form

$$z^*_{\text{def}} = \begin{cases} \begin{bmatrix} \frac{\bar{C}^i - R}{\bar{C}^i} \\ 0 \\ \frac{R}{\bar{C}^i} \end{bmatrix} & \bar{C}^i \geqslant R \\ \begin{bmatrix} 0 \\ \frac{R-\bar{C}^i}{C+F-\bar{C}^i} \\ \frac{C+F-R}{C+F-\bar{C}^i} \end{bmatrix} & \bar{C}^i < R \end{cases}$$

ignore Y and declare no attack

ignore Y and declare attack

Y_i-dependent decision

detection cost:

$$J^i_{\text{def}}{}^* = \begin{cases} \dfrac{AB}{A + B\frac{1-p^i_{\text{fn}}}{p^i_{\text{fp}}}}, & \bar{C}^i \geqslant R \\ \dfrac{AB}{A\frac{1-p^i_{\text{fp}}}{p^i_{\text{fn}}} + B}, & \bar{C}^i < R \end{cases}$$

attacker's reward: $\bar{C}^i \coloneqq C(1 - p^i_{\text{fn}}) + F p^i_{\text{fp}}$

$$J^i_{\text{att}}{}^* = F \begin{cases} \dfrac{R p^i_{\text{fp}}}{\bar{C}^i} - S_i, & \bar{C}^i \geqslant R \\ \dfrac{(R-C)(1-p^i_{\text{fp}}) + C p^i_{\text{fn}}}{C+F-\bar{C}^i} - S_i, & \bar{C}^i < R \end{cases}$$

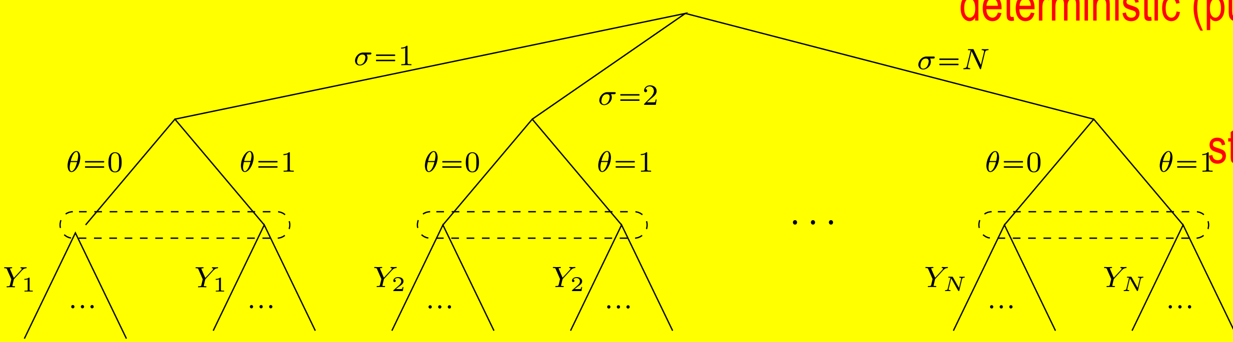
Attacker's reward determines (deterministic) optimal sensor selection form attacker

$$\sigma = \arg\max_i F \begin{cases} \dfrac{R p^i_{\text{fp}}}{\bar{C}^i} - S_i & \bar{C}^i \geqslant R \\ \dfrac{(R-C)(1-p^i_{\text{fp}}) + C p^i_{\text{fn}}}{C+F-\bar{C}^i} - S_i & \bar{C}^i < R, \end{cases}$$

Main Results

attacker selects sensor σ

deterministic (pure policy)



attacker selects θ

stochastic (mixed policy)

defender selects detection sets Y_i

stochastic (mixed policy)

$$\hat{\theta} = \begin{cases} 1 & y_\sigma \in Y_\sigma \\ 0 & y_\sigma \notin Y_\sigma \end{cases}$$

$$J_{\text{def}}^{i\;*} = \begin{cases} A+B\,\frac{\cdots}{p_{\text{fp}}^i} & \\ \frac{AB}{A\,\frac{1-p_{\text{fp}}^i}{p_{\text{fn}}^i}+B}, & \bar{C}^i < R \end{cases} \qquad J_{\text{att}}^{i\;*} = F \begin{cases} \frac{\cdots p_{\text{fp}}}{\bar{C}^i} - S_i, & \bar{C}^i \geqslant R \\ \frac{(R-C)(1-p_{\text{fp}}^i)+Cp_{\text{fn}}^i}{C+F-\bar{C}^i} - S_i, & \bar{C}^i < R \end{cases}$$

Attacker's reward determines (deterministic) optimal sensor selection form attacker

$$\sigma = \arg\max_i F \begin{cases} \frac{Rp_{\text{fp}}^i}{\bar{C}^i} - S_i & \bar{C}^i \geqslant R \\ \frac{(R-C)(1-p_{\text{fp}}^i)+Cp_{\text{fn}}^i}{C+F-\bar{C}^i} - S_i & \bar{C}^i < R, \end{cases}$$

Main Results

UC SANTA BARBARA engineering

Theorem: Consider only 3 detection sets ($\emptyset$, "all", Y_i)

For each sensor selection $\sigma = i$, the game has mixed Nash eq. with detection policy of the form

$$z^*_{\text{def}} = \begin{cases} \begin{bmatrix} \frac{\bar{C}^i - R}{\bar{C}^i} \\ 0 \\ \frac{R}{\bar{C}^i} \end{bmatrix} & \bar{C}^i \geqslant R \\ \begin{bmatrix} 0 \\ \frac{R - \bar{C}^i}{C + F - \bar{C}^i} \\ \frac{C + F - R}{C + F - \bar{C}^i} \end{bmatrix} & \bar{C}^i < R \end{cases}$$

ignore Y and declare no attack

ignore Y and declare attack

Y_i-dependent decision

detection cost:

But this policy depends crucially on opponent's goal…

$$J_{\text{attack}} \coloneqq R\theta - C\,\mathrm{P}(\hat{\theta} = 1, \theta = 1) + F\,\mathrm{P}(\hat{\theta} = 1, \theta = 0)$$

$$\bar{C}^i \coloneqq C(1 - p^i_{\text{fn}}) + F p^i_{\text{fp}}$$

Learning Through Fictious Play

Over repeated instances of the game, defender keeps track of

$$\bar{y}^i(t) = \begin{bmatrix} \text{fraction of times } \theta = 0 \text{ in } [0,t] \text{ when } \sigma = i \\ \text{fraction of times } \theta = 1 \text{ in } [0,t] \text{ when } \sigma = i \end{bmatrix}$$

at time $t+1$ defender makes optimal Bayesian decision assuming

$$\begin{bmatrix} \mathrm{P}(\theta = 0) \\ \mathrm{P}(\theta = 1) \end{bmatrix} = \bar{y}^{\sigma}(t)$$

empirical distribution observed so far for sensor $\sigma(t+1)$ is the correct prior

***Theorem**:*

1. If attacker is using a fixed policy (Nash or not), then $\bar{y}^{\sigma}(t)$ converges to optimal best response

trivial

2. If attacker is also using fictious play based on the observation of detectors policy, then $\bar{y}^{\sigma}(t)$ converges to optimal best response

non-trivial, based on results of convergence of 2x*N* player games

Learning Through Fictious Play

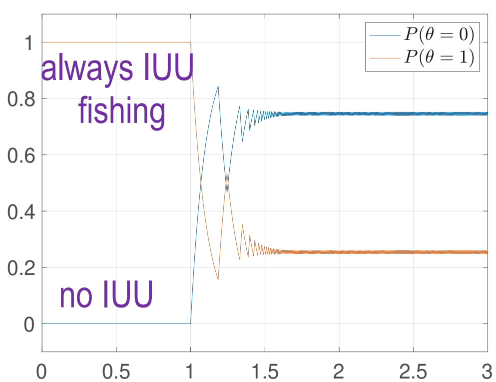
UC SANTA BARBARA engineering

Theorem:

1. If attacker is using a fixed policy (Nash or not), then $\bar{y}^{\sigma}(t)$ converges to optimal best response

2. If attacker is also using fictious play based on the observation of detectors policy, then $\bar{y}^{\sigma}(t)$ converges to optimal best response

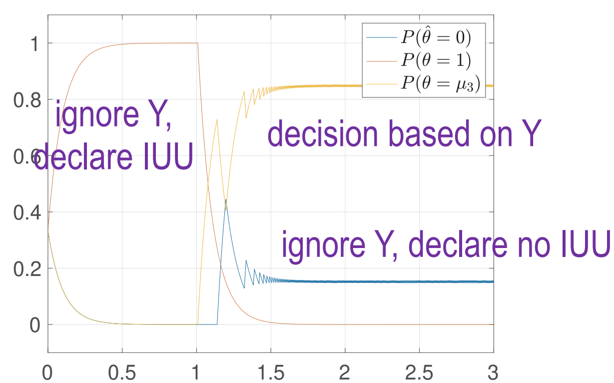
attacker's mixed policy



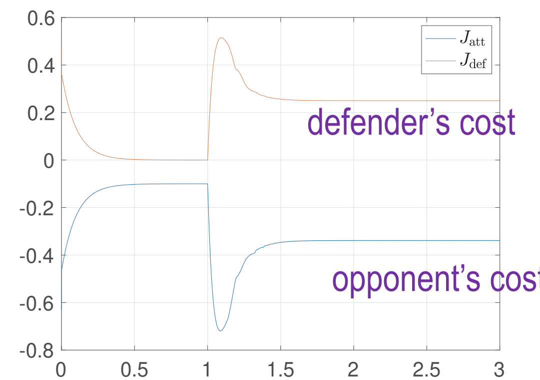
opponent uses fixed (non optimal) policy

opponent uses optimal policy

defender's mixed policy



player's costs



defender adjusts to opponent's policy, regardless of optimal or not

Outline

UC SANTA BARBARA engineering

Partial information games

- Sensor manipulation
  - Sensor-reveal game
  - Existence/computation of Nash equilibrium
  - Data-driven approach to detection

with D. Garajic (BAE)

[based on CDC'20, WeC09.4]

- Asymmetric information
  - online learning of the attacker's best response

with G. Yang (UCSB), R. Poovendran (UW)

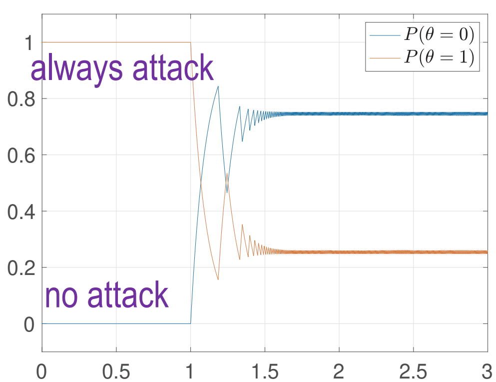
[based on CDC'20, FrB09.3]

Learning Through Fictious Play

Impact:
- Robustify attack detection with respect to potential sensor manipulation
- Algorithm adapts to changes in opponent's intent

2. If attacker is also using fictious play based on the observation of detectors policy, then $\bar{y}^{\sigma}(t)$ converges to optimal best response

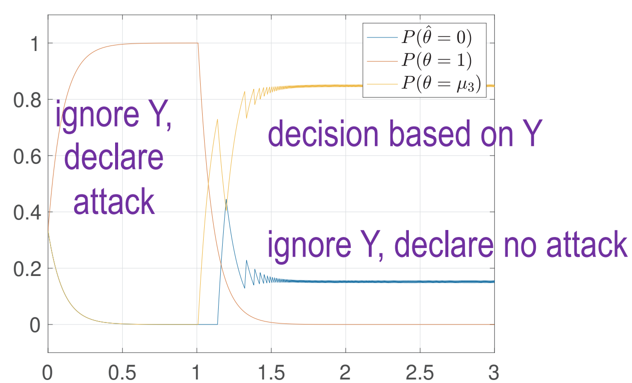
attacker's mixed policy



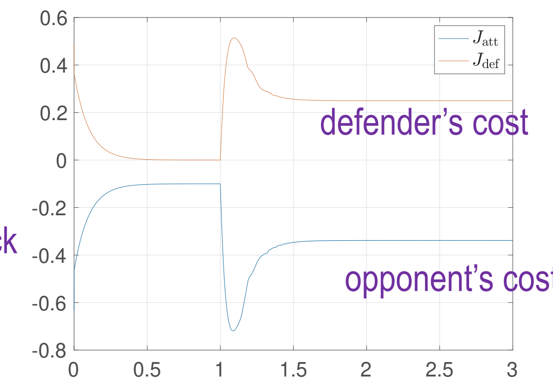
opponent uses fixed (non optimal) policy

opponent uses optimal policy

defender's mixed policy



player's costs



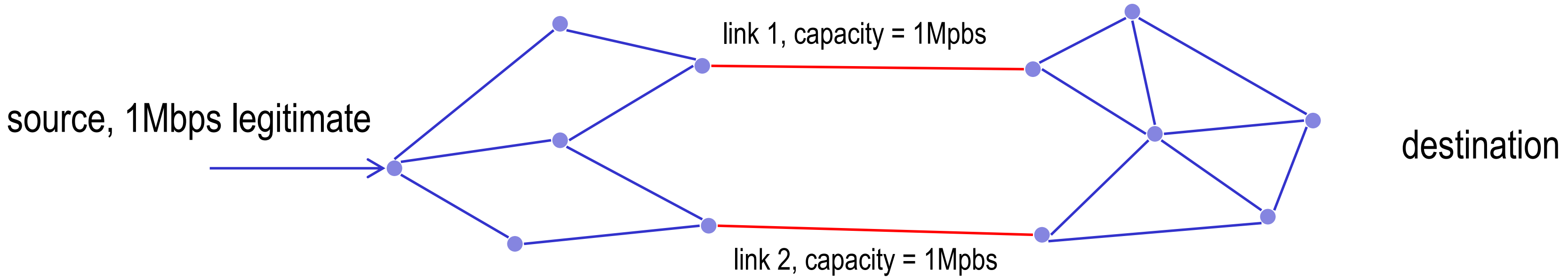
defender adjusts to opponent's policy, regardless of optimal or not

Asymmetric Information

- Defender often unaware of the attack objective ahead of time
- Attacker can observe/probe the defense strategy prior to attack

↳ Lack of no-regret policies for defender

routing game example *(based on cross fire attack [Kang-Lee-Gligor-13])*



- *attacker has resources to create 1Mpbs traffic through links 1, 2, or a combination*
- *router must decide how to route 1Mpbs of traffic through links 1, 2, or a combination*

Asymmetric Information

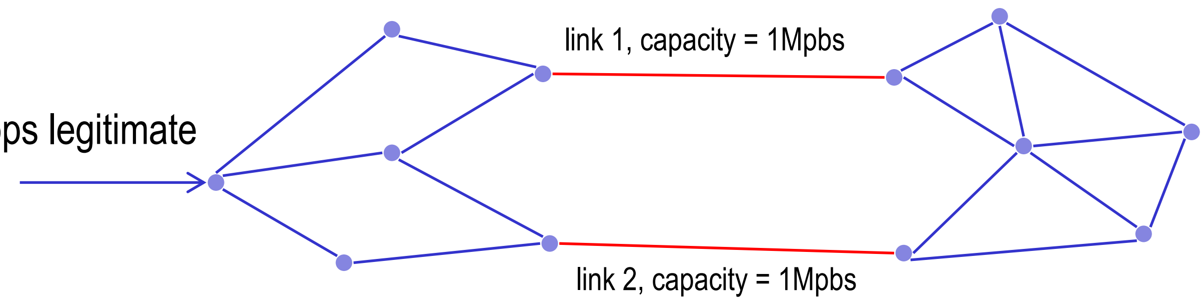
UC SANTA BARBARA engineering

- Defender often unaware of the attack objective ahead of time
- Attacker can observe/probe the defense strategy prior to attack

↳ Lack of no-regret policies for defender

routing game example *(based on cross fire attack)*

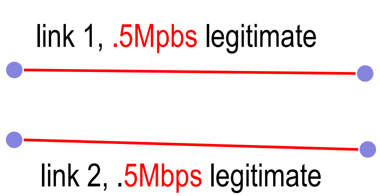
Source, 1Mbps legitimate



destination

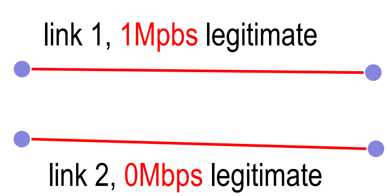
- *attacker has resources to create 1Mpbs traffic through links 1, 2, or a combination*
- *router must decide how to route 1Mpbs of traffic through links 1, 2, or a combination*

option I



Optimal attack floods link 2 and compromises .5Mpbs of legitimate traffic

option II



Optimal attack floods link 1 and compromises 1Mpbs of legitimate traffic

Asymmetric Information

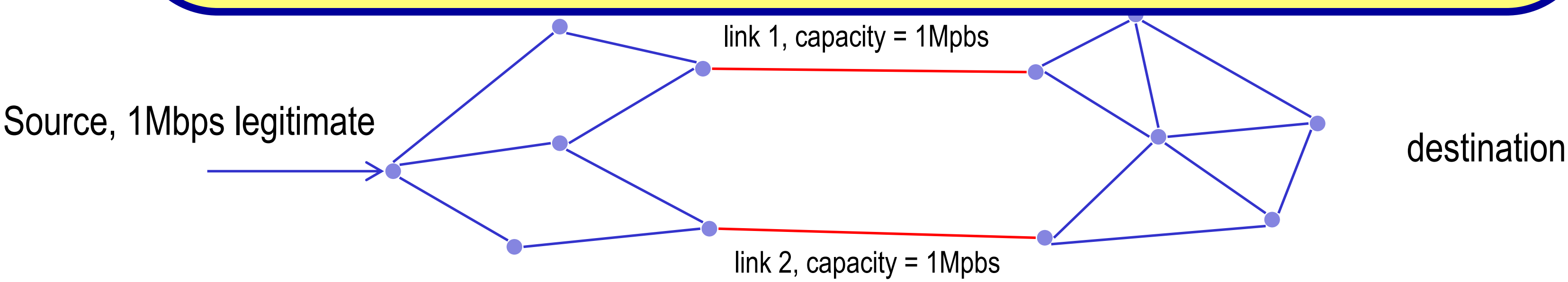
UC SANTA BARBARA
Engineering

1. Option I is optimal for router but leads to regret:

   "if router knew link 2 was going to be flooded, it would have routed all legitimate traffic through link 1"

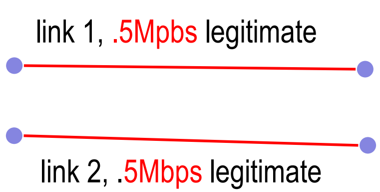
2. But option II also leads to regret…

Asymmetric information ⇒ regret



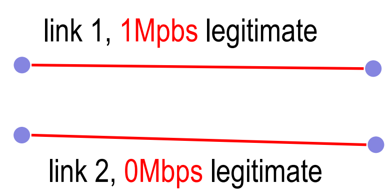
- *attacker has resources to create 1Mpbs traffic through links 1, 2, or a combination*
- *router must decide how to route 1Mpbs of traffic through links 1, 2, or a combination*

option I



Optimal attack floods link 2 and compromises .5Mpbs of legitimate traffic

option II



Optimal attack floods link 1 and compromises 1Mpbs of legitimate traffic

Asymmetric Information

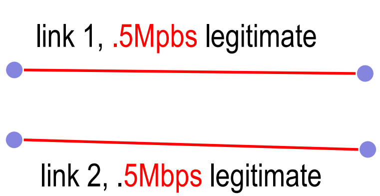
- Defender often unaware of the attack objective ahead of time
- Attacker can observe/probe the defense strategy prior to attack

🙁 *Naif defender that reacts to "current" environment regrets choice (no Nash equilibrium, fictitious play will not converge)*

🙂 *Defender must learn attacker's "response" and plan accordingly (need to consider Stackelberg equilibrium)*

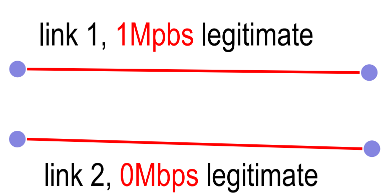
- *attacker has resources to create 1Mpbs traffic through links 1, 2, or a combination*
- *router must decide how to route 1Mpbs of traffic through links 1, 2, or a combination*

option I



Optimal attack floods link 2 and compromises .5Mpbs of legitimate traffic

option II



Optimal attack floods link 1 and compromises 1Mpbs of legitimate traffic

Stackelberg equilibrium

UC SANTA BARBARA engineering

defender's (leader) cost

$$u^* = \arg\min_{u \in \mathcal{U}} J_u\big(u, \beta(u)\big)$$

defender's action space

attacker (follower) cost

$$\beta(u) := \arg\min_{a \in \mathcal{A}} J_a(u, a)$$

attacker's best response

attacker's action space

Challenge in CPS sec.: attacker's best response $\beta(\cdot)$ is unknown to defender

- attacker's intent $J_a(\cdot)$ not known a priori
- attacker's capabilities $\mathcal{A}$ not known a priori

Stackelberg equilibrium

UC SANTA BARBARA engineering

defender's (leader) cost

$$u^* = \arg\min_{u \in \mathcal{U}} J_u\big(u, \beta(u)\big)$$

defender's action space

attacker (follower) cost

$$\beta(u) := \arg\min_{a \in \mathcal{A}} J_a(u, a)$$

attacker's best response

attacker's action space

Challenge in CPS sec.: attacker's best response $\beta(\cdot)$ is unknown to defender

- attacker's intent $J_a(\cdot)$ not known a priori
- attacker's capabilities $\mathcal{A}$ not known a priori

Online learning approach:

1. estimate best response function $\hat{\beta}(\cdot)$ based on observation of attacker's actions a
2. select defender action

$$\hat{u}^* = \arg\min_{u \in \mathcal{U}} J_u\big(u, \hat{\beta}(u)\big)$$

Stackelberg learning

defender's (leader) cost

$$u^* = \arg\min_{u \in \mathcal{U}} J_u\big(u, \beta(u)\big)$$

attacker (follower) cost

$$\beta(u) := \arg\min_{a \in \mathcal{A}} J_a(u, a)$$

Assume

$$\beta(\cdot) \in \Big\{ f(\theta, \cdot) : \theta \in \Theta \Big\}$$

linearly parameterized function approximator on compact set $\mathcal{U}$
(results extend to only approx. match)

Stackelberg learning

UC SANTA BARBARA engineering

defender's (leader) cost

$$u^* = \arg\min_{u\in\mathcal{U}} J_u\big(u,\beta(u)\big)$$

attacker (follower) cost

$$\beta(u) := \arg\min_{a\in\mathcal{A}} J_a(u,a)$$

Assume

$$\beta(\cdot) \in \Big\{ f(\theta,\cdot) : \theta \in \Theta \Big\}$$

linearly parameterized function approximator on compact set $\mathcal{U}$
(results extend to only approx. match)

attack best-response learning rule:

$$\dot{\theta} = -\lambda_e(t)\left[\nabla_\theta \Big\| f(\theta,u) - a \Big\|^2\right]_{T_\Theta}$$

hysteresis switching
stops adaptation when error
$\|f(\theta,u)-a\|$ is smaller than ε/2

gradient descent learning with projection

- can be computed without knowing θ
- can be computed without even observing a, just $J_u(u,a)$

$$\lambda_e(t) := \begin{cases} \lambda_\theta & \text{if } \|f(\theta,u)-a\| \geq \varepsilon; \\ \lim_{s\nearrow t}\lambda_e(s) & \text{if } \|f(\theta,u)-a\| \in (\varepsilon/2,\varepsilon); \\ 0 & \text{if } \|f(\theta,u)-a\| \leq \varepsilon/2 \end{cases}$$

guarantees:

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\theta-\theta^*\|^2 \leq -2\lambda_e\|f(\theta,u)-a\|^2 \leq 0$$

$\|\theta-\theta^*\|$ is monotonically decreasing;
stops only if $\|f(\theta,u)-a\| < \varepsilon$
↳ will prove: stops in finite time

Stackelberg learning

defender's (leader) cost

attacker (follower) cost

$$u^* = \arg\min_{u \in \mathcal{U}} J_u\big(u, \beta(u)\big) \qquad \beta(u) := \arg\min_{a \in \mathcal{A}} J_a(u, a)$$

Assume

$$\beta(\cdot) \in \Big\{ f(\theta, \cdot) : \theta \in \Theta \Big\}$$

linearly parameterized function approximator on compact set $\mathcal{U}$ (results extend to only approx. match)

attack best-response learning rule: $$\dot{\theta} = -\lambda_e(t) \Big[\nabla_\theta \Big\| f(\theta, u) - a \Big\|^2 \Big]_{T_\Theta}$$

defender's adaptation rule: $$\dot{u} = -\lambda_u \Big[\nabla_u J_u\big(u, f(\theta, u)\big) \Big]_{T_\mathcal{U}}$$

gradient descent with projection (can be generalized by other adaptation mechanisms)

guarantees:

$$\dot{\theta} = 0 \implies \frac{\mathrm{d}}{\mathrm{d}t} J_u(u, f(\theta, u)) \leq -\lambda_u \Big\| \big[-\nabla_u J_u(u, f(\theta, u)) \big]_{T_\mathcal{U}} \Big\|^2 \leq 0$$

leader's cost $J_u(u, f(\theta, u))$ is monotonically decreasing;
stops only if $\big[-\nabla_u J_u(u, f(\theta, u))\big]_{T_\mathcal{U}} = 0$
↳ will prove: convergence

Convergence analysis

UC SANTA BARBARA engineering

- Follower: $\min_{a\in\mathcal{A}} J_a(u,a)$ Best response: $\beta(u) := \arg\min_{a\in\mathcal{A}} J_a(u,a)$
- Leader: $\min_{u\in\mathcal{U}} J_a(u,\beta(u))$ Assume $\beta(\cdot) = f(\theta^*,\cdot) \in \{f(\theta,\cdot) : \theta\in\Theta\}$
- Est.: $\dot\theta = \lambda_e(t)\big[-\nabla_\theta\|f(\theta,u)-a\|^2\big]_{T_\Theta}$ Opt.: $\dot u = \lambda_u\big[-\nabla_u J_u(u,f(\theta,u))\big]_{T_\mathcal{U}}$

hysteresis switching:
stops only if error is smaller than $\varepsilon > 0$

Theorem.

1. After finite time T, estimate will accurately predict $a(t)$
$$\big\|f(\theta(t),u(t)) - a(t)\big\| < \varepsilon \qquad \forall t\ge T$$
2. Leader will converge to 1st-order optimality condition
$$\big[-\nabla_u J_u(u,f(\theta,u))\big]_{T_\mathcal{U}} \to 0$$
3. One can use probing to guarantee correct estimation
$$\|\theta(t)-\theta^*\| < \varepsilon_\theta \qquad \forall t\ge T$$

generalize Barbalat's lemma for hysteresis switching

establish invariance principle for projected gradient descent

persistent excitation (PE)
$$\int_t^{t+\tau_0} \nabla_\theta f(s)^\top \nabla_\theta f(s)\,\mathrm{d}s \ge \alpha_0 I$$
$$\varepsilon_\theta = \varepsilon\sqrt{\tau_0/\alpha_0}$$

Model mismatch

UC SANTA BARBARA engineering

- Follower: $\min_{a \in \mathcal{A}} J_a(u,a)$ Best response: $\beta(u) := \arg\min_{a \in \mathcal{A}} J_a(u,a)$
- Leader: $\min_{u \in \mathcal{U}} J_a(u, \beta(u))$ Assume $\cancel{\beta(\cdot) = f(\theta^*, \cdot) \in \{f(\theta, \cdot) : \theta \in \Theta\}}$
  $\exists\, \theta^* \in \Theta : \max_{u \in \mathcal{U}} \| f(\theta^*, u) - \beta(u) \| \leq \varepsilon_f$
- Est.: $\dot{\theta} = \lambda_e(t) \big[-\nabla_\theta \| f(\theta, u) - a \|^2 \big]_{T_\Theta}$ Opt.: $\dot{u} = \lambda_u \big[-\nabla_u J_u(u, f(\theta, u)) \big]_{T_\mathcal{U}}$

hysteresis switching:
stops only if error is smaller than $\varepsilon > \kappa \varepsilon_f$

Theorem.

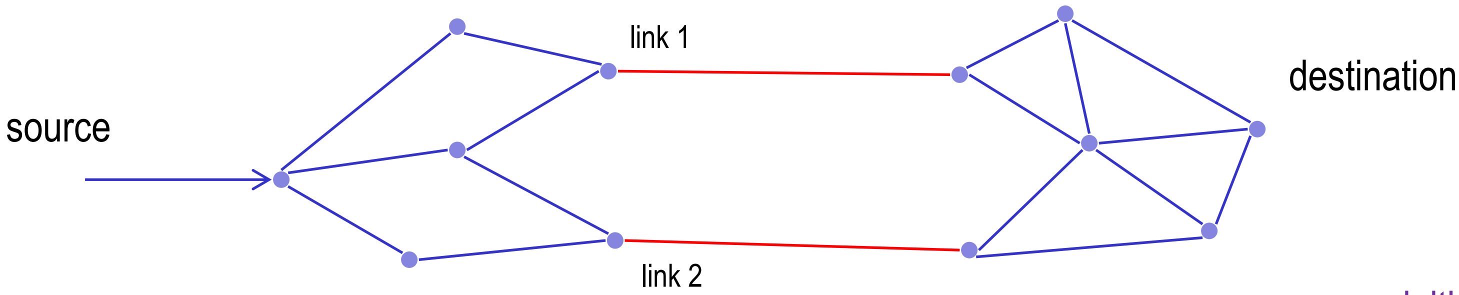
1. After finite time T, estimate will accurately predict $a(t)$
$$\| f(\theta(t), u(t)) - a(t) \| < \varepsilon \qquad \forall\, t \geq T$$
2. Leader will converge to 1st-order optimality condition
$$\big[-\nabla_u J_u(u, f(\theta, u)) \big]_{T_\mathcal{U}} \to 0$$
3. One can use probing to guarantee correct estimation
$$\| \theta(t) - \theta^* \| < \varepsilon_\theta \qquad \forall\, t \geq T$$

generalize Barbalat's lemma for hysteresis switching

establish invariance principle for projected gradient descent

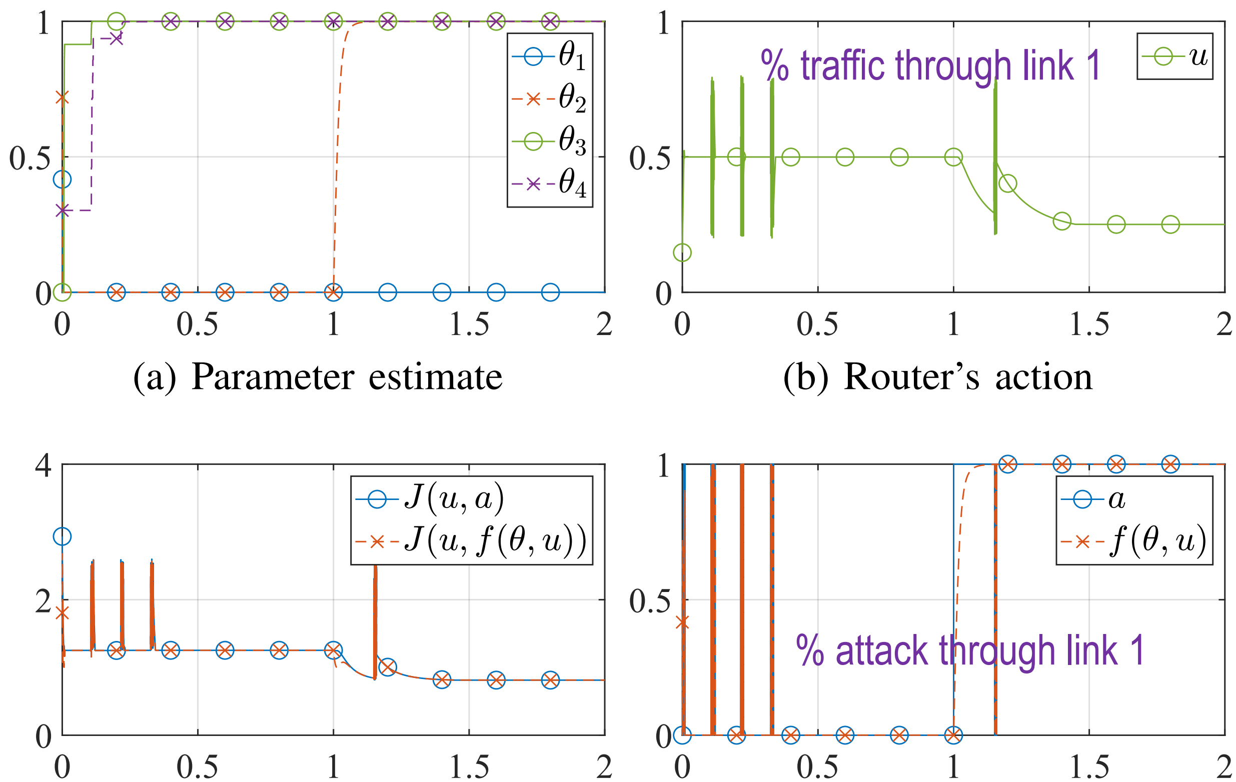
persistent excitation (PE)
$$\int_t^{t+\tau_0} \nabla_\theta f(s)^\top \nabla_\theta f(s)\, \mathrm{d}s \geq \alpha_0 I$$
$$\varepsilon_\theta = 2\kappa \varepsilon_f \sqrt{\tau_0/\alpha_0} > 0$$

Routing game example



- initially attacker wants to disrupt both links equally

  router's policy adjusts to balance traffic



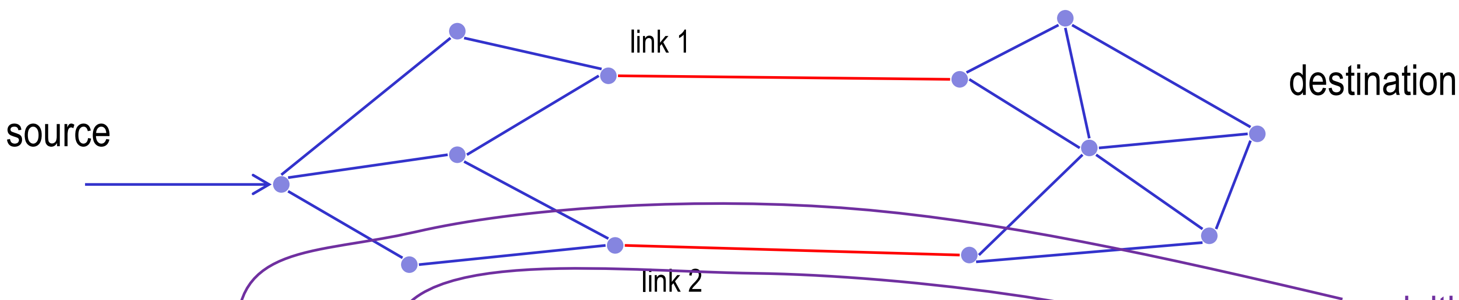
(a) Parameter estimate

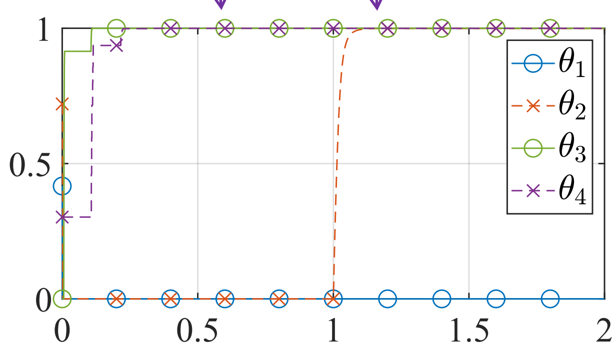
(b) Router's action

(c) Router's cost and estimate

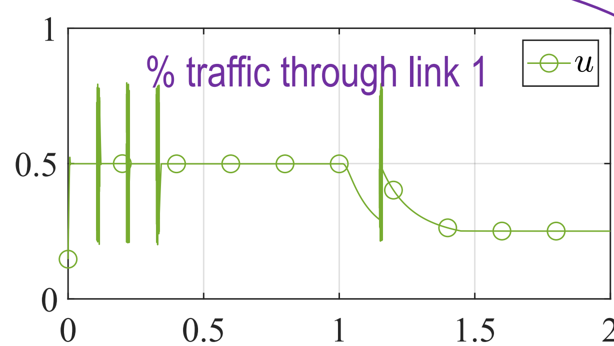
(d) Attacker's action and estimate

Routing game example

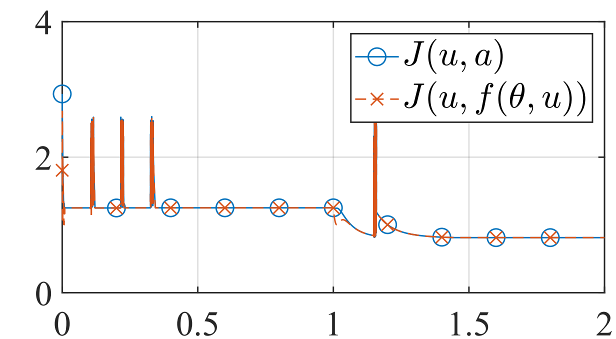




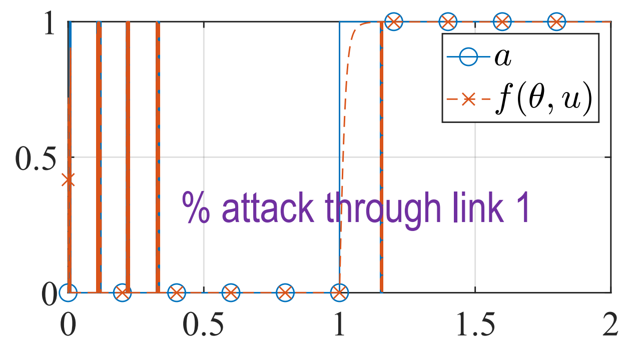
(a) Parameter estimate



(b) Router's action



(c) Router's cost and estimate



(d) Attacker's action and estimate

- initially attacker wants to disrupt both links equally

  router's policy adjusts to balance traffic

- at time 1, attacker's goal changes to disrupt link 1

  router's policy shifts more traffic to link 2

On both cases, attacker's best response is to focus all attack traffic on single link, but router learns intent through probing

Outline

Partial information games

- ➳ Sensor manipulation
  - Sensor-reveal game
  - Existence/computation of Nash equilibrium
  - Data-driven approach to detection

with D. Garajic (BAE)

[based on CDC’20, WeC09.4]

- ➳ Asymmetric information
  - online learning of the attacker’s best response

with G. Yang (UCSB),
R. Poovendran (UW)

[based on CDC’20, FrB09.3]