FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Scheduling Resource-Bounded Monitoring Devices

Aron Laszka
University of California, Berkeley

joint work with

Waseem Abbas, Yevgeniy Vorobeychik, and Xenofon Koutsoukos
Vanderbilt University

Berkeley UNIVERSITY OF CALIFORNIA

Massachusetts Institute of Technology

NSF

MICHIGAN

VANDERBILT UNIVERSITY

Monitoring Cyber-Physical Systems

* Being able to detect **faults** and **failures** before suffering substantial or irreversible physical damage is fundamental to the resilient operation of cyber-physical systems
* In many spatially-distributed cyber-physical systems, faults and failures may only be detected by **monitoring devices** deployed over the system



FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Resource-Bounded Monitoring Devices

* Using **battery-powered** devices can reduce deployment costs
  * in some cases, battery power is the only feasible option
* Battery-powered devices have limited lifetime
  ⟷ cyber-physical systems may require extended lifetime
* **Sleep scheduling**
  * only a subset of monitoring devices are active at any given time
  * "sleeping" devices conserve battery power
  * however, some events may not be detected
* Finding an **optimal schedule** is challenging:
  tradeoff between <u>system lifetime</u> and <u>detection performance</u>

FORCES
FOUNDATIONS OF RESILIENT
CYBER-PHYSICAL SYSTEMS

Contributions

1. Monitoring networks
2. Simultaneous placement and scheduling
3. Minimizing detection delay

1. Monitoring Networks

* Many spatially-distributed cyber-physical systems can be naturally modeled as **networks**
  * water, wastewater, gas, and oil pipelines
  * electric networks
* Physical topology and device capabilities determine the **set of failures** a monitoring device can detect

→ **Optimal sleep schedule** must take the physical topology of the system into account

Wastewater pipeline network of Norfolk, VA

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Example Application: Water-Distributions Networks

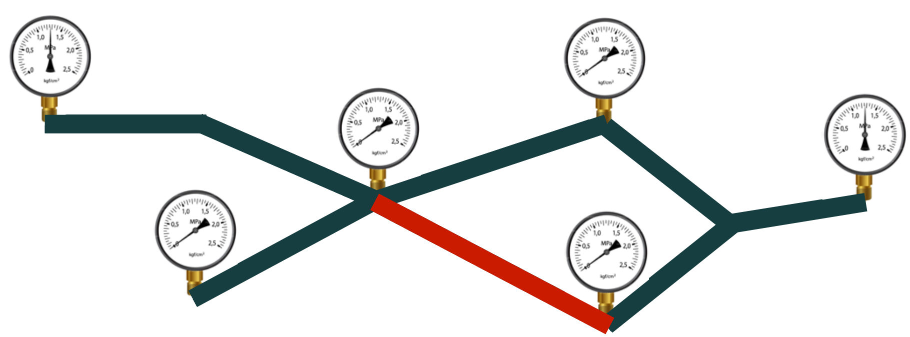
* Leakages in water-distribution networks can cause
  * significant economic losses
  * extra costs for final consumers
  * third-party damage and health risks
  * …

"*6 billion gallons of water per day may be wasted in the U.S.*" (Center for Neighborhood Technology, 2013)

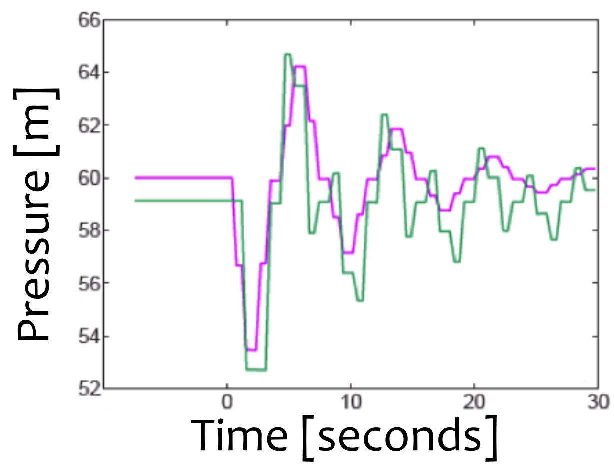
"*worldwide cost of physical losses is over $8 billion*" (World Bank, 2006)

Physical Leakage Detection Model

* Pressure sensors can detect nearby events, such as leakages and pipe bursts



Measurements of Nearest Sensors

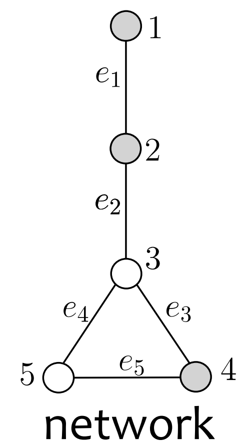


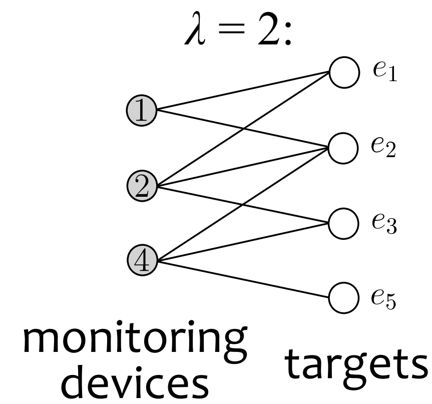
* Continuous monitoring through sensors can significantly reduce physical damage and financial losses
* However, battery-powered sensors have limited lifetime

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Network and Monitoring Model

- Network: $G(V, E)$
  - set of monitoring devices: $X \subseteq V$
  - set of targets: $Y \subseteq (V \cup E)$
- Distance-based monitoring model
  - distance $d(x, y) =$
    - between nodes x and y: number of hops between x and y
    - between node x and edge $y = (u, v)$: $\max\{d(x, u), d(x, v)\}$
  - range of monitoring devices = λ
  - device u can monitor all nodes and edges within λ distance:





$$\{v \subseteq V : \ d(u, v) \leq \lambda\} \cup \{e \subseteq E : \ d(u, e) \leq \lambda\}$$

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Schedules and Detection Performance

- Limited battery power
  - network lifetime = k time intervals
  - active monitoring time = $\sigma < k$
- Schedule: $(S_1, \ldots, S_k)$
  - every $S_i \subseteq X$
  - for every monitoring device s:

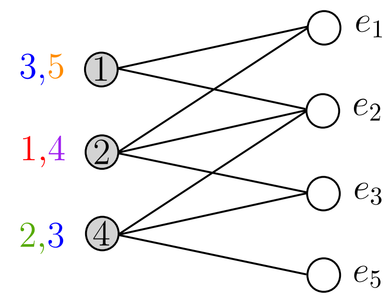
$$\left|\{ S_i \mid s \in S_i \}\right| \leq \sigma$$

- Average detection performance:

$$\mathcal{D} = \frac{1}{k} \sum_{i=1}^{k} \frac{m_i}{|Y|}$$

  - where m_i is the number of monitored targets in time interval i

Optimal schedule for $k = 5$ and $\sigma = 2$:



$S_1 = \{2\}$
$S_2 = \{4\}$
$S_3 = \{1,4\}$
$S_4 = \{2\}$
$S_5 = \{1\}$

Detection performance:

$m_1 = 3$
$m_2 = 3$
$m_3 = 4 \quad \rightarrow \quad \mathcal{D} = 0.75$
$m_4 = 3$
$m_5 = 2$

Computational Complexity

* Number of feasible schedules
  * example: 10 monitoring devices, 30 time intervals, each device may be active in 10 intervals $\rightarrow 847{,}660{,}528^{10} \approx 10^{89}$ > number of atoms in the observable universe

Theorem: Given an instance of the scheduling problem, finding a feasible schedule that maximizes the average detection performance is an **APX-hard** problem.

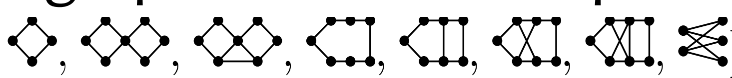
$\rightarrow$ no polynomial-time approximation scheme

* Proof: reduction from the Maximum Cut Problem

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Special Case: Continuous Complete Monitoring

* Continuous complete monitoring: detection performance $\mathcal{D} = 1$ $\longrightarrow$ objective: maximizing lifetime k
* Dominating-set based solution:
  * every set of active monitoring devices S_i is a dominating set

Theorem: Let G be a graph such that
- G has a minimum degree of at least two,
- no subgraph of G is isomorphic to $K_{1,6}$, and
- $G \neq$ {  } .

Then, there exists a schedule for any $k < \frac{5}{2}\sigma$ such that $\mathcal{D} = 1$.

* Proximity graphs are always $K_{1,6}$-free

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Potential Game Formulation

* Game $\Gamma(\mathcal{P}, \mathcal{A}, \mathcal{U})$:
  * $\mathcal{P} = \{1, 2, ..., n\}$: set of players
  * $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_n$: actions spaces
  * $\mathcal{U} = \{\mathcal{U}_1, \mathcal{U}_2, ..., \mathcal{U}_n\}$: utility functions
* Potential game: $\Gamma(\mathcal{P}, \mathcal{A}, \mathcal{U})$ is a potential game if there exists a potential function φ: $\mathcal{A} \rightarrow \mathrm{R}$ such that

$$\mathcal{U}_x(a_x, a_{-x}) - \mathcal{U}_x(a'_x, a_{-x}) = \phi(a_x, a_{-x}) - \phi(a'_x, a_{-x})$$

* Potential games are extensively used for distributed control optimization problems

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Scheduling Problem as a Potential Game

- **Scheduling game** $\Gamma(\mathcal{P}, \mathcal{A}, \mathcal{U})$:
  - players $\mathcal{P} = X$ monitoring devices
  - actions space $\mathcal{A}_i = \sigma$-subsets of $\{1, 2, \ldots, k\}$
  - utility functions:

$$U_x(a_x, a_{-x}) \triangleq \sum_{j=1}^{k} a_{xj} \left| N(x) \setminus \bigcup_{z \in S_j \setminus \{x\}} N(z) \right|$$

  where $N(x)$ is the set of targets in range of device x, and a_{xj} is 1 if device x is active in time interval j and 0 otherwise
- Potential function:

$$\phi(a) \triangleq \sum_{j=1}^{k} \left| \bigcup_{x \in S_j} N(x) \right|$$

Theorem: The scheduling game is a potential game.

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Binary Log-Linear Learning

- Algorithm
  - start with random actions
  - in each iteration, a player and an action is chosen at random
  - action is updated with some probability, which depends on the resulting utilities

Algorithm 2 Binary Log-Linear Learning

1: **Initialization:** Pick a small $\epsilon \in \mathbb{R}_+$, an $a \in \mathcal{A}$, and total iterations.
2: **While** $i \leq$ iterations **do**
3: Pick a random node $x \in \mathcal{X}$, and a random $a'_x \in \mathcal{A}_x$.
4: Compute $P_\epsilon = \frac{\epsilon^{U_x(a'_x, a_{-x}(t))}}{\epsilon^{U_x(a'_x, a_{-x}(t))} + \epsilon^{U_x(a_x, a_{-x}(t))}}$.
5: Set $a_x \leftarrow a'_x$ with probability P_ϵ.
6: $i \leftarrow i + 1$
7: **End While**

- Only the joint action profiles that **maximize the potential function** form stochastically stable equilibria
  $\longrightarrow$ algorithm will converge to a global optimum
- Polynomial running time

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Simple Heuristic: Greedy Algorithm

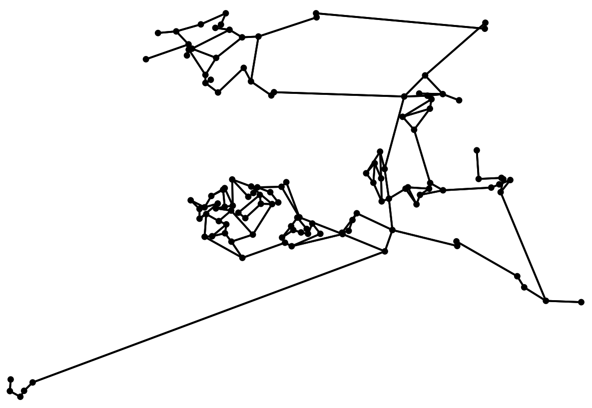
- Scheduling problem resembles set covering since we have to "cover" targets with monitoring devices in every time interval
- Algorithm
  - start with an empty schedule ($S_1 = S_2 = \ldots = S_k = \emptyset$)
  - in each iteration, activate an additional monitoring device x in interval l (i.e., add x to S_l)
  - choose (x, l) such that the increase in detection performance is maximum
- Polynomial running time

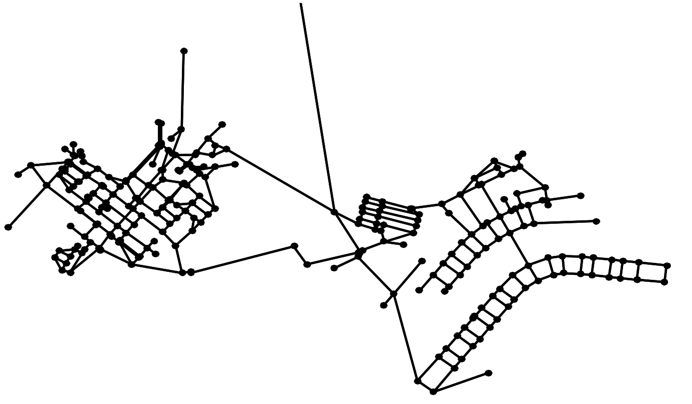
Algorithm 1 Greedy Heuristic

Given: σ, $\mathcal{K} = \{1, 2, \cdots, k\}$
Initialization: $\mathcal{X}' \leftarrow \mathcal{X}$, $f(x) \leftarrow \emptyset$, $\forall x \in \mathcal{X}$
While $|\mathcal{X}'| \neq \emptyset$ **do**
    $(x, \ell) \leftarrow \underset{x \in \mathcal{X}', \ell \in \mathcal{K}}{\operatorname{argmax}} \sum_{y \in \mathcal{Y}} |f(y)|$
    $f(x) \leftarrow f(x) \cup \{\ell\}$
    If $|f(x)| = \sigma$ **do**
        $\mathcal{X}' \leftarrow \mathcal{X}' \setminus \{x\}$
    End If
End While

Numerical Evaluation – Water Networks

* Real-world water-distribution networks





Water network 1:

- 126 nodes, 168 links, one reservoir, one pump, and two storage tanks
- extensively studied in the sensor placement literature

Water network 2:

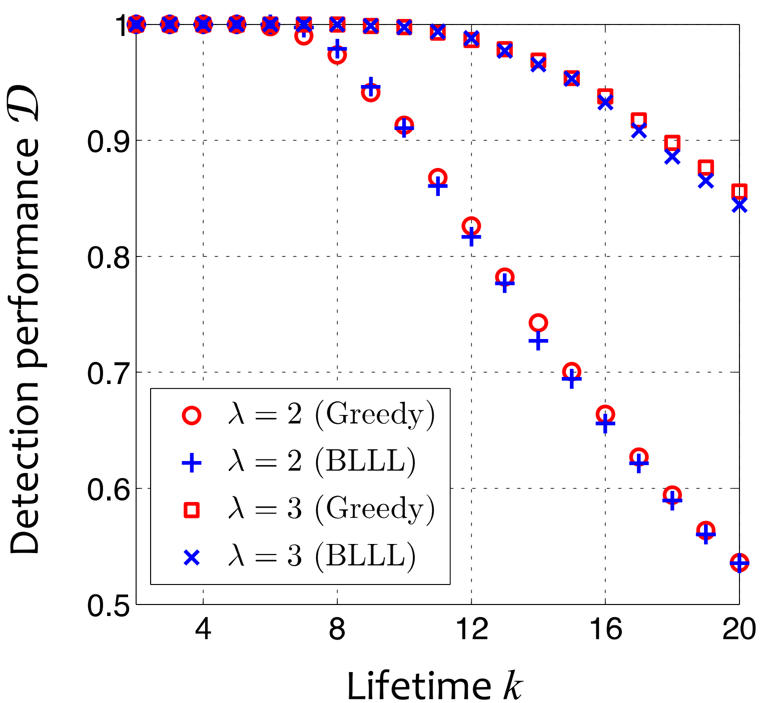
- 270 nodes, 366 links, three tanks, and five pumps
- grid system in Kentucky

* For both networks, we let $X = V$, $Y = E$, and $\sigma = 2$

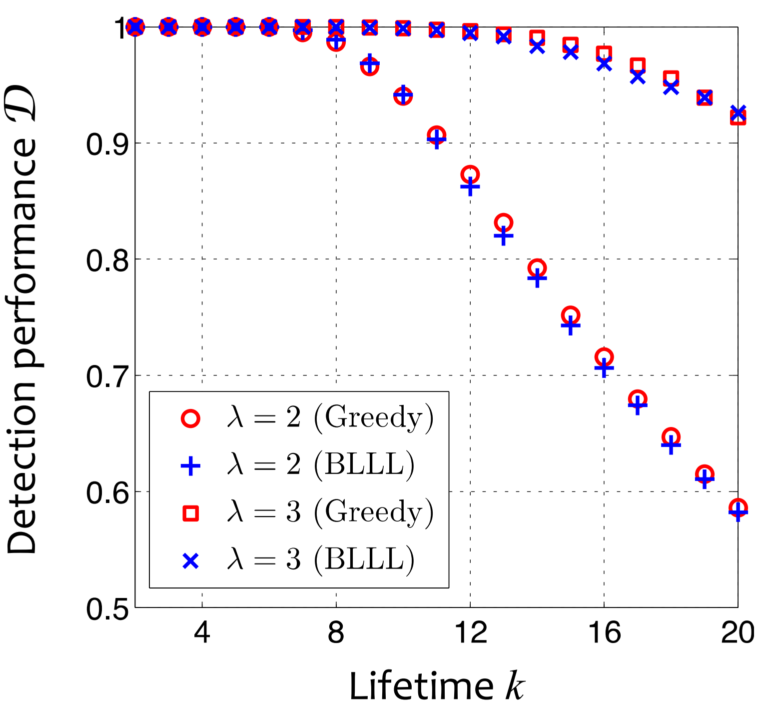
FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Numerical Results – Water Networks

Water network 1

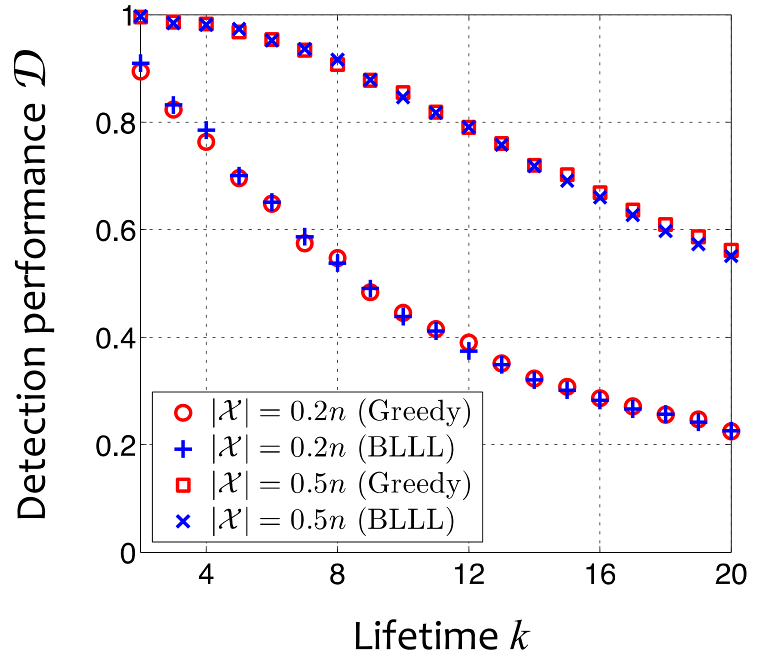


Water network 2



Numerical Results – Geometric Graphs

- Random geometric graphs
  - 100 nodes are distributed uniformly at random in a unit square
  - nodes are connected if their Euclidean distance is at most 0.2
  - certain fraction (either 20% or 50%) of the nodes are selected to be monitoring devices
- We let $X = Y = V$ and $\sigma = 2$



FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

2. Simultaneous Placement and Scheduling

* So far, we assumed that the set of monitoring devices X is given
* Placement problem:

  where to place monitoring devices in a network?
* Simple "solution":
  1. find a placement maximizing some "target coverage" metric
  2. find a schedule (e.g., greedy algorithm or BLLL)
* Simultaneous placement and scheduling
  * find a placement and schedule **simultaneously**
  * advantage: placement can take the feasible schedules into account $\rightarrow$ higher detection performance or longer lifetime

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Problem Formulation

- Input
  - network $G(V, E)$, range λ, lifetime k, power σ
  - number of monitoring devices: n
  - feasible monitoring device **locations:** $S \subseteq V$
- Solution: placement and schedule $(X, S_1, \ldots, S_k)$,
  where $X \subseteq S$, $|X| = \mathrm{n}$, and every $S_i \subseteq X$
- Objective: detection performance $\mathcal{D}$
- Complexity:
  at least as hard as the scheduling problem

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Adapting Binary Log-Linear Learning

- Action space $\mathcal{A}_i$ of scheduling game Γ: σ-subsets of $\{1, 2, ..., k\}$
- **Scheduling and placement game** $\Gamma^*(\mathcal{P}, \mathcal{A}^*, \mathcal{U})$:
  - we extend each action space with a <u>position</u>: $\mathcal{A}_i^* = \mathcal{A}_i \times \mathcal{S}$
  - everything else is the same as in the scheduling game
- Scheduling and placement game is a potential game
  $\longrightarrow$ convergence properties still hold

Algorithm 3 Simultaneous Placement and Scheduling

1: **Initialization:** Pick a small $\epsilon \in \mathbb{R}_+$ and the number of iterations. Select randomly a subset of nodes $\mathcal{X} \subseteq \mathcal{S}$, and assign labels to nodes in $\mathcal{X}$, i.e, select $a \in \mathcal{A}$.
2: **While** $i \leq$ iterations **do**
3: Randomly select a node $x \in \mathcal{X}$.
4: Randomly select a node $s \in (\mathcal{S} \setminus \mathcal{X}) \cup \{x\}$, and $a_s \in \mathcal{A}_s$.
5: Compute $P_\epsilon = \frac{\epsilon^{U_s(a_s, a_{-x})}}{\epsilon^{U_s(a_s, a_{-x})} + \epsilon^{U_x(a_x, a_{-x})}}$.
6: With probability P_ϵ, set $\mathcal{X} \leftarrow (\mathcal{X} \setminus \{x\}) \cup \{s\}$, and select a_s for node s.
7: $i \leftarrow i + 1$
8: **End While**

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

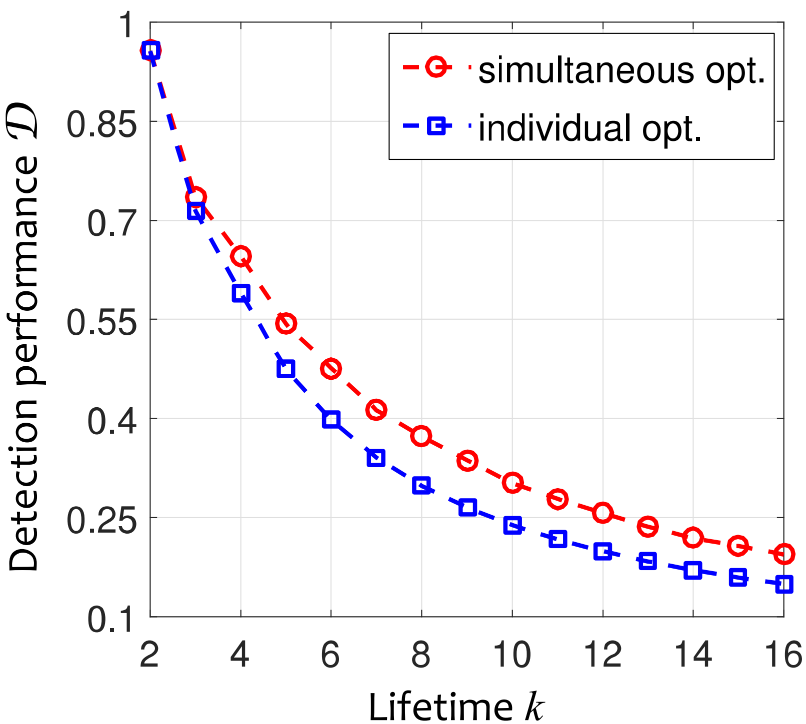
Numerical Evaluation

* Baseline: **individual optimization**
  1. select a set of monitoring devices $X \subseteq S$ maximizing the number of targets that are monitored by at least one device
  2. find a schedule using Algorithm 2 (BLLL)
* Networks
  * Water network 1:
    place monitoring devices at 25 nodes (out of 126 total)
  * Random geometric graph:
    place monitoring devices at 10 nodes (out of 50 total)

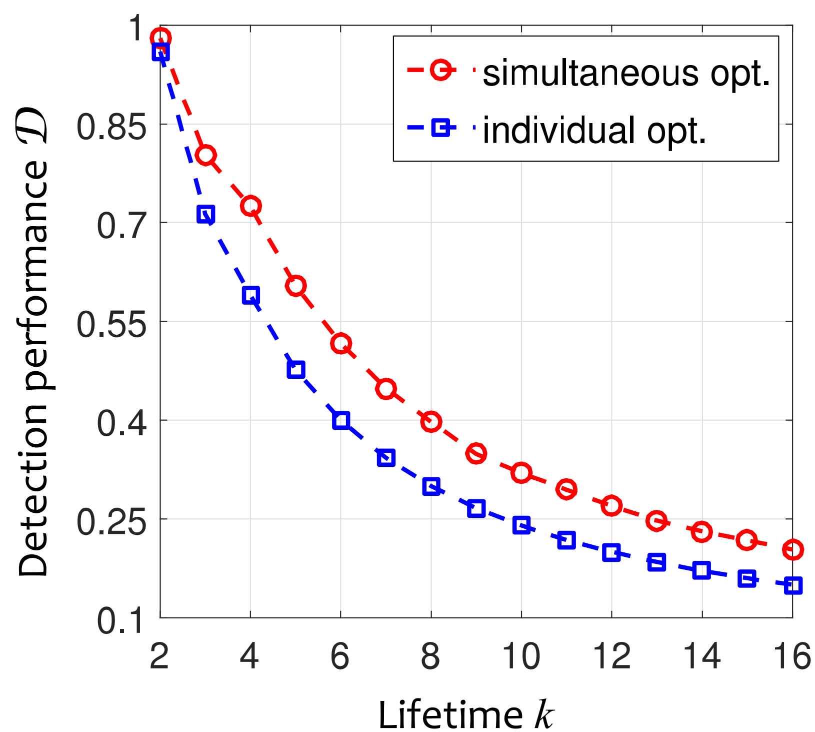
FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Numerical Results

Real-world water network



Random geometric graph



FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

3. Minimizing Detection Delay

- Maximizing detection performance $\mathcal{D}$
  = maximizing the **number of time intervals** in which each target is monitored
  - without considering **which time intervals**
- Detection performance does not guarantee timely detection
  - example: if a target could be monitored in 5 out of 10 intervals,
    $\{1, 2, 3, 4, 5\} \rightarrow \mathcal{D} = 0.5$, and average time until detection is 1.5
    $\{1, 3, 5, 7, 9\} \rightarrow \mathcal{D} = 0.5$, but average time until detection is 0.5
- When losses depend on the time between a failure and its mitigation, we need to minimize the average <u>time until detection</u>

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Problem Formulation

* Detection delay $\mathcal{T}$:
  expected number of time intervals until a failure at a uniformly randomly chosen target in a uniformly randomly chosen time interval is detected

$$\mathcal{T} = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{|Y|} \sum_{y \in Y} \left[\min \left\{ j \;\middle|\; j \geq i \wedge y \in \bigcup_{x \in S_{(j \bmod k)}} N(x) \right\} - i \right]$$

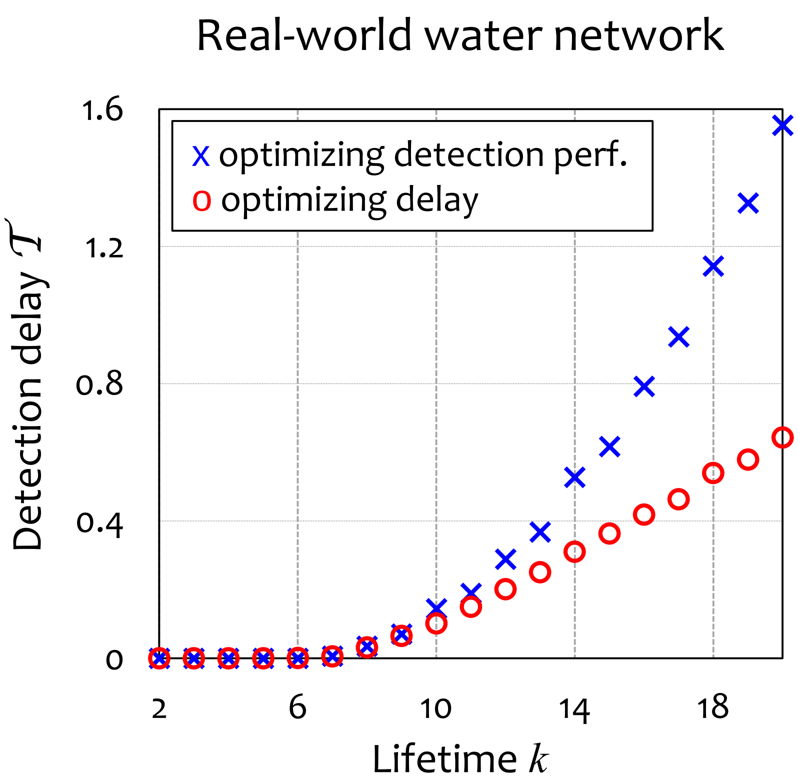
* Computational complexity
  * minimizing detection delay is an **APX-hard** problem
  * using the same argument as for maximizing detection performance

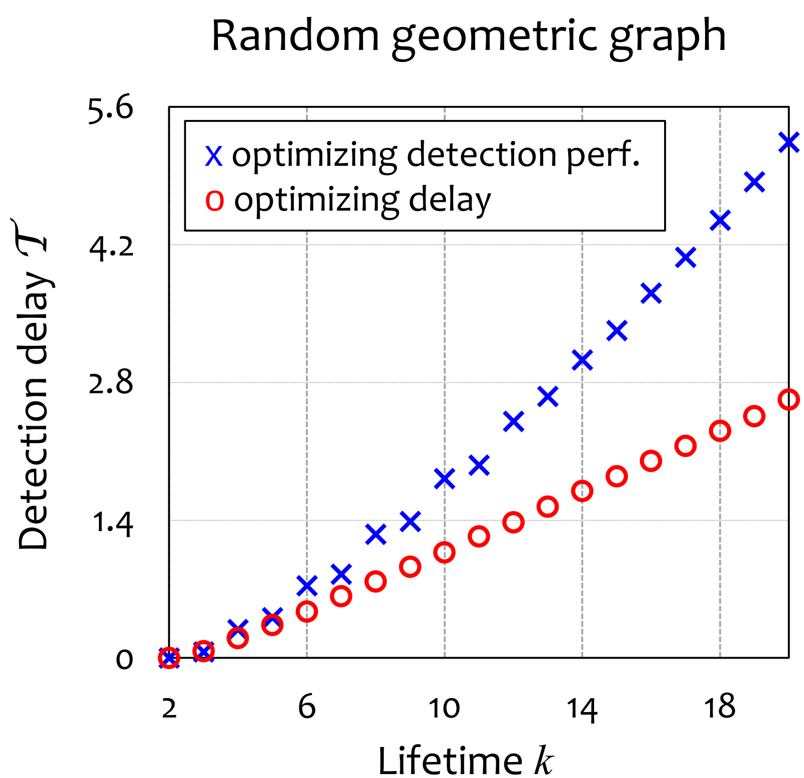
FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Numerical Evaluation

* **Simulated annealing**
  * start with a random schedule
  * in each iteration, choose a random monitoring device and a new σ-set of time intervals, and switch to the new assignment with probability depending on the difference in detection delay $\mathcal{T}$
* Baseline solution:
  maximizing detection performance $\mathcal{D}$ (using BLLL or greedy)
* Networks:
  * Water network 1 with $X = V$, $Y = E$, $\sigma = 2$, and $\lambda = 2$
  * Random geometric graph with $X = Y = V$, $\sigma = 2$, and $\lambda = 1$

Numerical Results





FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Summary

* Sleep scheduling enables prolonging the lifetime of battery-powered monitoring devices
* When detection depends on the physical topology of a cyber-physical system, optimal sleep schedules must take the **physical topology** into account
* **Simultaneous placement and scheduling** of monitoring devices may lead to a substantial increase in detection performance
* For certain applications, **minimizing detection delay** can lead to significantly better schedules

Thank you for your attention!

Questions?

Numerical Results – Water Networks