An Efficient Approach to Fault Identification in Urban Water Networks Using Multi-Level Sensing

Waseem Abbas[1], Polina Sela[2], Saurabh Amin[2], Xenofon Koutsoukos[1]

[1] Vanderbilt University

[2] Massachusetts Institute of Technology

(ACM BuildSys 2015)

VANDERBILT UNIVERSITY

Leakages in Water-Distribution Networks

Leakages in urban water networks can cause

- significant economic losses
- extra costs for final consumers
- third-party damage and health risks
- ...

"Worldwide cost of physical losses is ***over $8 billion/year****."*

(World Bank, 2006)

"Every single day in US, nearly ***six billion gallons*** *of treated water is simply lost due to leaky, aging pipes and outdated systems."* *(Center for Neighborhood Technology, 2013)*

No. of main breaks/yr: 237,600
Revenue loss/yr: $2.8 billion
Small leaks: 500,000 – 1,500,000

(Distribution System Inventory, Integrity and Water Quality, AWWA 2004)

Outline

Objective:

Water loss reduction caused by leaks and bursts by improved **localization** of pipe failures in urban water distribution networks.

Approach:

Design a **sensor placement** that maximizes the detection and identification of link failures through the minimum number of sensors of various types.

Methods:

- Formulation of localization of link failures as a combinatorial **coverage problem** (such as minimum test cover).
- **Efficient algorithm** to solve the (localization) coverage problem.
- **Multi-level sensors' placement**, in which the information collected by sensors is analyzed in more detail.
- **Heterogeneous sensors' placement,** in which different classes of multi-level sensors are placed for a trade-off between the localization performance and the cost entailed.

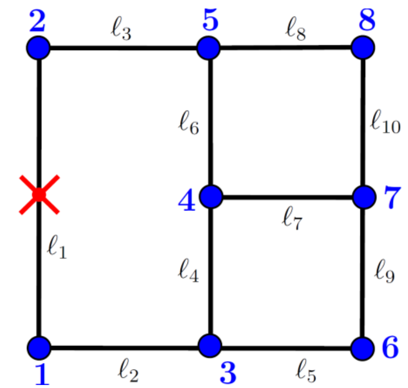
Evaluation:

Simulations of real/benchmark water distribution networks.

System Model

Water distribution network: **Graph (nodes, edges)**

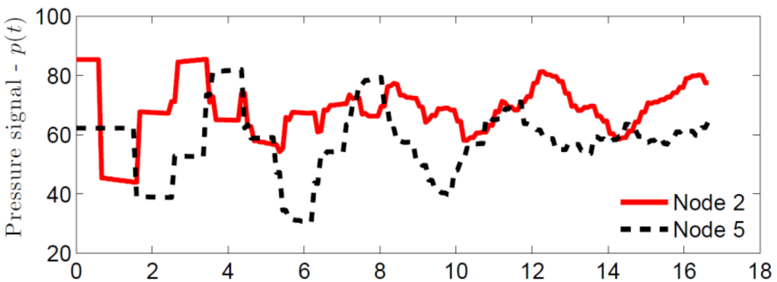
- Nodes: connections and consumers
- Edges: pipes



Event set over links: $\mathcal{L} = \{\ell_1, \ell_2, \cdots, \ell_n\}$

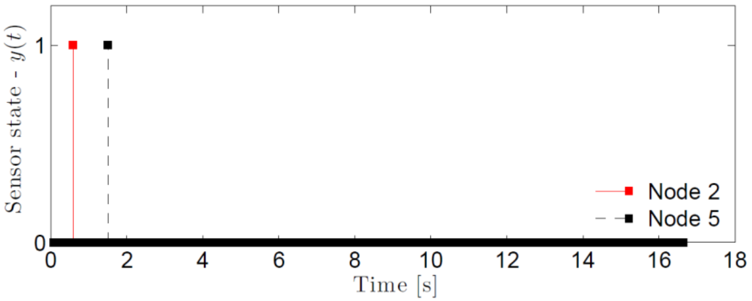
Sensor set over nodes: $\mathcal{S} = \{S_1, \cdots, S_m\}$

Sensed pressure by S_i at time *t* : $p_i(t)$



Single-level sensing model:

A sensor S_i detects a failure (at some link) whenever sensed pressure (or some function of it) is greater than a certain threshold ε.



Sensor output: $\mathbf{y}_{S_i}(\ell)$

$$\mathbf{y}_{S_i}(\ell) = \begin{cases} 1 & \text{if failure at } \ell \text{ is detected by } S_i \\ 0 & \text{otherwise.} \end{cases}$$

Influence Matrix

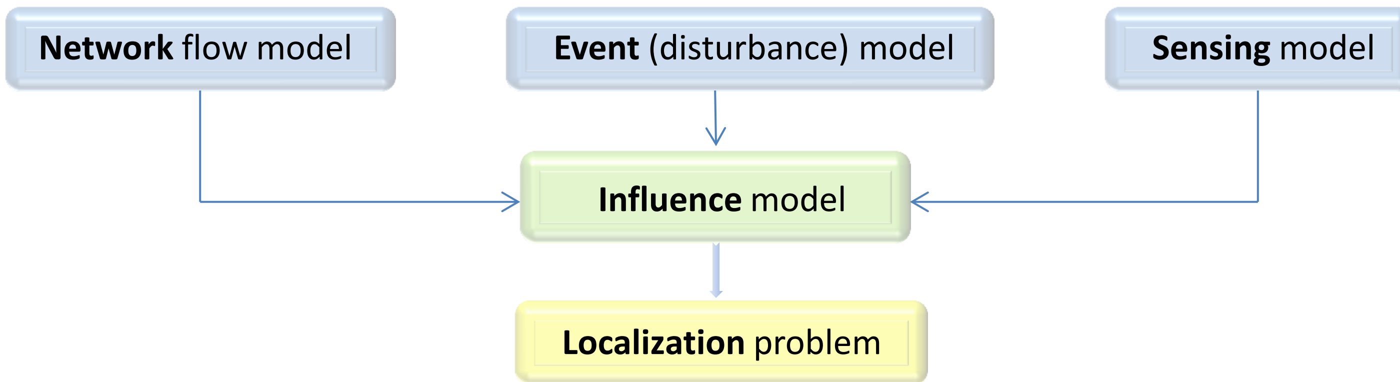
For a network with sensor set $\mathcal{S}$ and event set $\mathcal{L}$, we can write a boolean **influence matrix**, $\mathcal{M}$, of dimensions $|\mathcal{L}| \times |\mathcal{S}|$.

- ℓ_i : i^{th} **row** corresponds to the **event** at the i^{th} link.
- S_j : j^{th} **column** corresponds to the j^{th} **sensor**.
- $\mathcal{M}_{ij}$: j^{th} sensor output in response to the event i.

$$\mathcal{M}(\mathcal{L},\mathcal{S}) = \begin{array}{c} \\ \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \\ \ell_5 \\ \ell_6 \\ \ell_7 \\ \ell_8 \\ \ell_9 \\ \ell_{10} \end{array} \begin{array}{c} \begin{array}{cccccccc} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 & S_7 & S_8 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

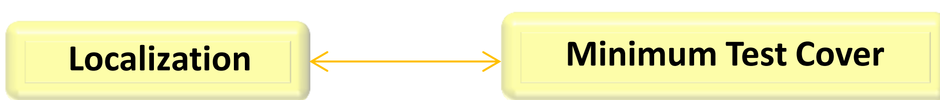
The output of sensor S_8 is 1 when failure at link ℓ_3 occurs

Localization and Min. Test Cover Problem



Localization Problem:

Find the minimum number of sensors and their locations, i.e., $S \subseteq \mathcal{S}$, so that the maximum number of link failures can be uniquely identified and can be distinguished from one another.



Minimum Test Cover:

Given,

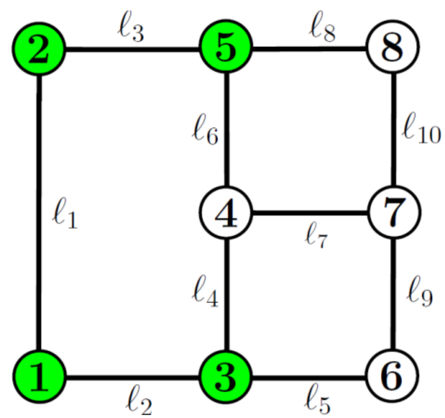
$\mathcal{L}$ = Finite set of elements

$\mathcal{S}$ = Collection of subsets of $\mathcal{L}$

$= \{S_1, S_2, \cdots, S_m\}$

Find a minimum sub collection $S \subseteq \mathcal{S}$ such that if for a pair $x, y \in \mathcal{L}$, there exists some $S_i \in \mathcal{S}$ containing exactly one of x and y, then there exists some $S_j \in S$ also containing exactly one of x and y.

Localization (Example)



$\mathcal{M} =$

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
ℓ_1	1	1	1	0	1	0	0	0
ℓ_2	1	1	1	1	0	1	0	0
ℓ_3	1	1	0	1	1	0	0	1
ℓ_4	1	0	1	1	1	1	1	0
ℓ_5	1	0	1	1	0	1	1	0
ℓ_6	0	1	1	1	1	0	1	1
ℓ_7	0	0	1	1	1	1	1	1
ℓ_8	0	1	0	1	1	0	1	1
ℓ_9	0	0	1	1	0	1	1	1
ℓ_{10}	0	0	0	1	1	1	1	1

- Each link failure can be **uniquely** identified by the output of sensors **1,2,3,** and **5**.
- Thus, sensors in the set **{1,2,3,5}** are sufficient for the localization of failures.

Minimum Test Cover

How can we solve minimum test cover (MTC) for the localization problem?

Minimum test cover (MTC) → Minimum set cover (MSC)

$$\mathcal{L}^T = \{\ell_1, \ell_2, \cdots, \ell_n\}$$

$$\mathcal{S}^T = \{S_1^T, S_2^T, \cdots, S_m^T\}$$

→

For each pair (ℓ_i, ℓ_j), define e_{ij}.

$\mathcal{L}^C$ = Set of all e_{ij}'s

$$\mathcal{S}^C = \{S_1^C, S_2^C, \cdots, S_m^C\}$$

where

$e_{ij} \in S_k^C$ whenever exactly one of ℓ_i and ℓ_j is in S_k^T.

Note that $|\mathcal{L}^S| = \binom{n}{2}$

	S_1^T	S_2^T	S_3^T	S_4^T	S_5^T	S_6^T	S_7^T	S_8^T
ℓ_1	1	1	1	0	1	0	0	0
ℓ_2	1	1	1	1	0	1	0	0
ℓ_3	1	1	0	1	1	0	0	1
ℓ_4	1	0	1	1	1	1	1	0

...

Influence matrix

	S_1^C	S_2^C	S_3^C	S_4^C	S_5^C	S_6^C	S_7^C	S_8^C
ℓ_1, ℓ_2	0	0	0	1	1	1	0	0
ℓ_1, ℓ_3	0	0	1	1	0	0	0	1
ℓ_1, ℓ_4	0	1	0	1	0	1	1	0

...

Transformed matrix

Set Cover: Minimum number of columns that cover the maximum number of rows.

Solving Minimum Test Cover

Greedy approach:

- Obtain a **transformed matrix** (containing pair-wise link failures)
- In each iteration, pick a column (sensor) that **covers the maximum** number of uncovered pair-wise link failures.

The greedy approach gives the **best approximation ratio**, which is **(1 + 2 ln n).**

However, the approach **computationally expensive**, and not suitable for large scale networks.

Link failures = n $\longrightarrow$ Pairwise link failures = $\binom{n}{2}$

No. of comparisons in an iteration $\longrightarrow$ $\mathcal{O}\left(\binom{n}{2}\right)$

Solving Minimum Test Cover

Greedy approach:

- Obtain a **transformed matrix** (containing pair-wise link failures)
- In each iteration, pick a column (sensor) that **covers the maximum** number of uncovered pair-wise link failures.

The greedy approach gives the **best approximation ratio**, which is **(1 + 2 ln n).**

However, the approach **computationally expensive**, and not suitable for large scale networks.

Link failures = n $\longrightarrow$ Pairwise link failures = $\binom{n}{2}$

No. of comparisons in an iteration $\longrightarrow$ $\mathcal{O}\left(\binom{n}{2}\right)$

Proposed approach (Main result):

We propose a (greedy) solution that **does not** require a transformation of **all links to pair-wise link failures.**

The proposed approach gives the **same solution as the greedy** approach, thus the same best approximation ratio.

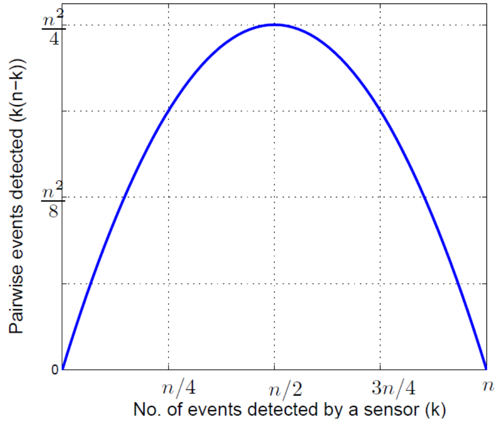
If *k* is the maximum no. of link failures detected by any sensor, then

No. of comparisons in an iteration $\longrightarrow$ $\mathcal{O}\left(\frac{k}{n}\binom{n}{2}\right); \quad k << n$

Solving Minimum Test Cover

Two basic observations used in the solution are:

- A sensor that detects k link failures, detects k (n-k) pairwise link failures.



- If a sensor detects link failures ℓ_i and ℓ_j, then it can not detect the pair-wise link failure $\ell_i \ell_j$.

  e.g., if $S_1 = \{\ell_2, \ell_3, \ell_5\}$, then S_1 cannot detect pair-wise failures $\{\ell_2\ell_3,\ \ell_2\ell_5,\ \ell_3\ell_5\}$.

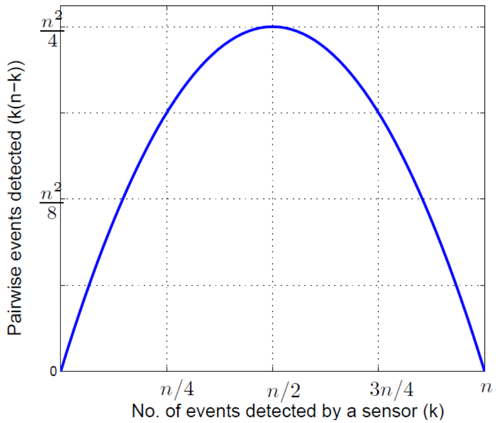
  Thus, if S_1 is selected in the test cover, we need to select sensor(s) in the next iteration(s) that also detect pairwise link failures corresponding to the links in S_1 (covered links of S_1) .

Both of these factors contribute to the selection of a sensor in each iteration of the solution.

Solving Minimum Test Cover

Two basic observations used in the solution are:

- A sensor that detects k link failures, detects k (n-k) pairwise link failures.



- If a sensor detects link failures ℓ_i and ℓ_j, then it can not detect the pair-wise link failure $\ell_i\ell_j$.

  e.g., if $S_1 = \{\ell_2, \ell_3, \ell_5\}$, then S_1 cannot detect pair-wise failures $\{\ell_2\ell_3,\ \ell_2\ell_5,\ \ell_3\ell_5\}$.

  Thus, if S_1 is selected in the test cover, we need to select sensor(s) in the next iteration(s) that also detect pairwise link failures corresponding to the links in S_1 (covered links of S_1) .

Both of these factors contribute to the selection of a sensor in each iteration of the solution.

In each iteration, select the sensor that has the *maximum* score,

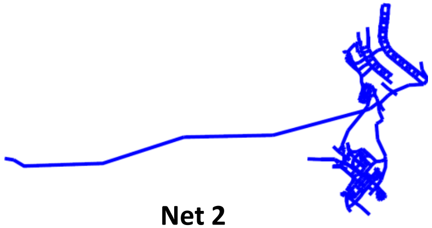
Score = $k_i(n_i - k_i)$ + (no. of undetected pairwise link failures corresponding to the covered links by the sensors selected in the cover)

No. of uncovered links in the i^{th} iteration.

No. of uncovered links that are covered by the sensor in the i^{th} iteration.

Simulations

Net 1

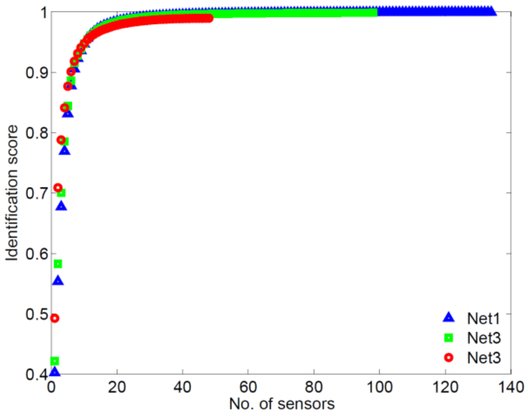


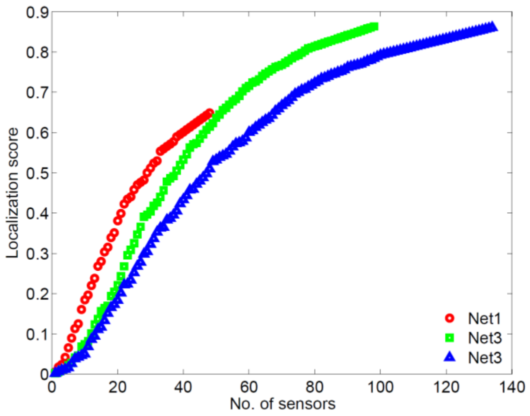
Net 2

Net 3

Network	No. of pipes	No. of nodes	No. of sensors	TG [sec]	FG [sec]
Net1	168	126	48	14.03	4.94
Net2	366	269	98	143.69	34.77
Net3	496	420	134	415.83	98.76

TG - transformed greedy; FG - fast greedy;





Multi-level Sensing

In **single-level sensing (1-bit)**, output of the sensor is

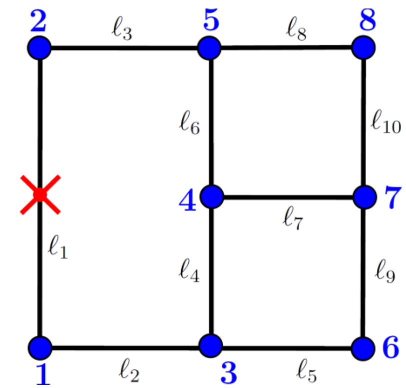
$$\begin{cases} 1 & \text{if failure event is detected, or} \\ 0 & \text{otherwise.} \end{cases}$$

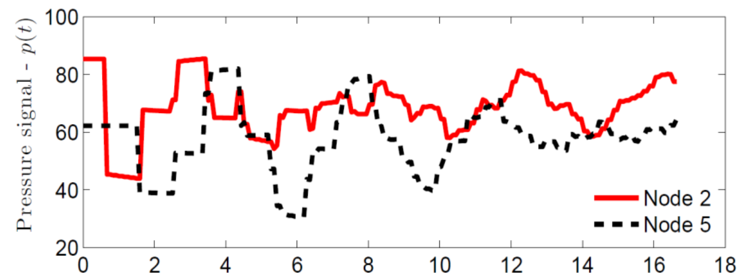
In **multi-level sensing (σ - bit)** , a sensor in case of detection, captures some *extra information* about the failure event, such as time taken to detect the event, intensity of the event, etc.

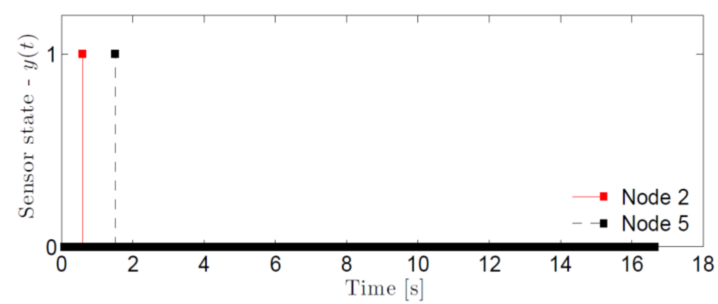
Output of sensor consists of *multiple bits*.

Case: Bi-level sensing

$$\begin{cases} 0 \quad 0 & \text{failure event is not detected,} \\ 1 \quad 0 & \text{event is detected early, i.e., in } [0 \quad t_1) \\ 0 \quad 1 & \text{event is detected later, i.e., in } [t_1 \quad T] \end{cases}$$

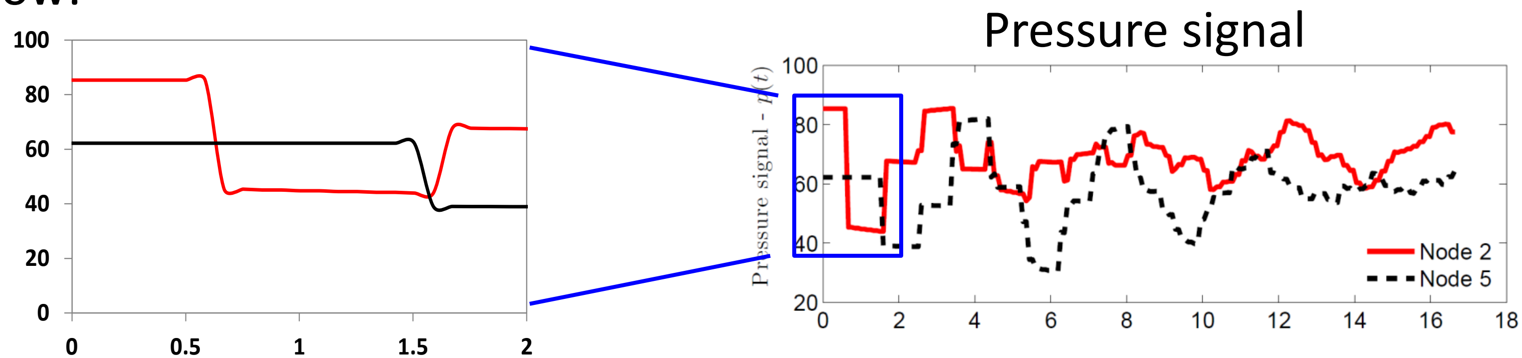






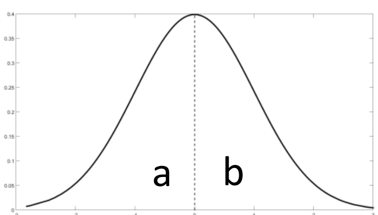
System Model

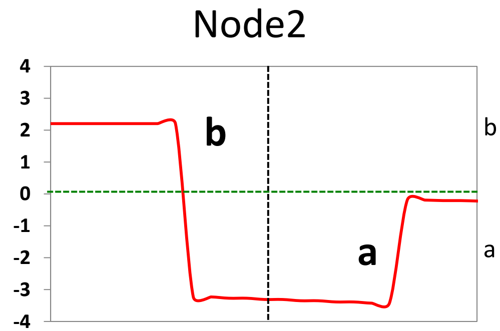
1. Time window:

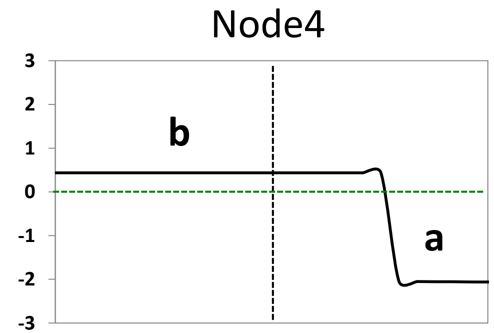


2. Standardized signal:

3. Symbolic representation:







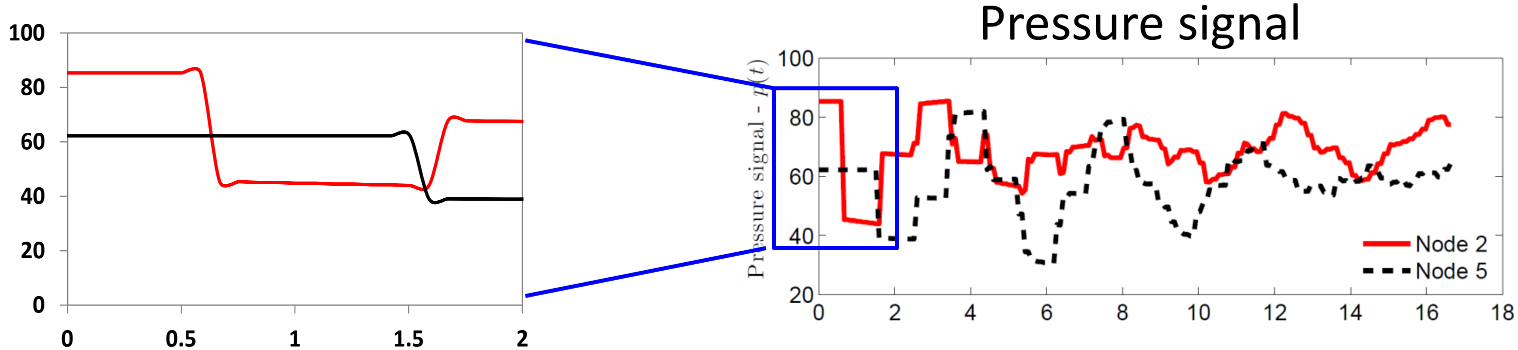
ba | ba

4. Boolean representation:

1-bit 1 | 1

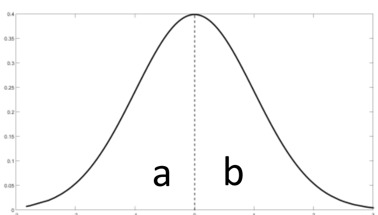
System Model

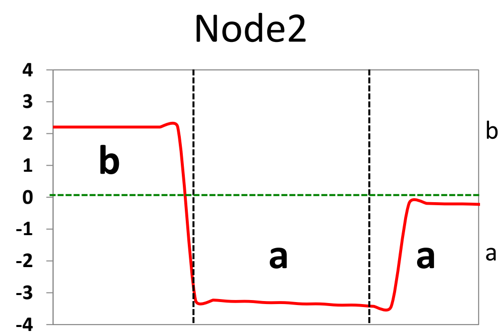
1. Time window:

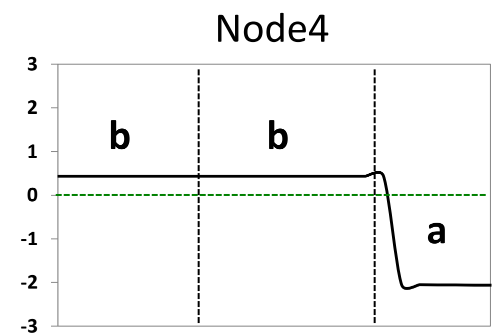


2. Standardized signal:

3. Symbolic representation:







baa

bba

4. Boolean representation:

	Node2	Node4
	baa	bba
1-bit	1	1
2-bit	10	01

Bi-level Sensing - Example

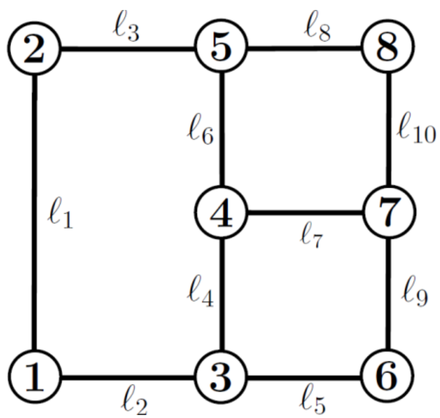
For a single failure event occurring at the center of each pipe, the output of a **2-bit sensor** S_i, denoted by $\mathbf{y}_{S_i}(\ell)$ is

$$\mathbf{y}_{S_i}(\ell) = \begin{cases} (1\ 0) & \text{if } d(S_i,\ell) < d_1 \\ (0\ 1) & \text{if } d_1 \leq d(S_i,\ell) \leq d_2 \\ (0\ 0) & \text{otherwise} \end{cases}$$

$d(S_i,\ell)$ is the length of the shortest path between S_i and ℓ.

$d_1 = 0.5[km], \quad d_2 = 1[km]$

(We note that $d = vt$)



$\mathcal{M} =$

	S_1		S_2		S_3		S_4		S_5		S_6		S_7		S_8	
ℓ_1	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0
ℓ_2	1	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0
ℓ_3	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1
ℓ_4	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0
ℓ_5	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
ℓ_6	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	1
ℓ_7	0	0	0	0	0	1	1	0	0	1	0	1	1	0	0	1
ℓ_8	0	0	0	1	0	0	0	1	1	0	0	0	0	1	1	0
ℓ_9	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	1
ℓ_{10}	0	0	0	0	0	0	0	1	0	1	0	1	1	0	1	0

Influence matrix

1-bit vs. σ-bit Sensors

- The **maximum number of pair-wise link failures** that can be detected by σ-bit sensors is greater than in the case of 1-bit sensors.
- For a **given number of sensors**, more pair-wise link failures can be detected by σ-bit sensors as compared to 1-bit case.

1-bit

k : No. of ***link failures*** detected

$\mathcal{P}_1$: No. of ***pair-wise link failures*** detected

$$\mathcal{P}_1 = k(n-k)$$

σ-bit ($\sigma > 1$)

k_i : No. of link ***failures detected*** by the i^{th} bit such that $\sum_{i=1}^{\sigma} k_i = k$.

$\mathcal{P}_\sigma$: No. of ***pair-wise link failures*** detected

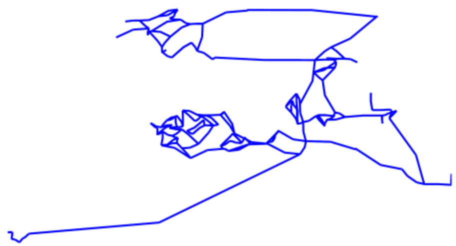
$$\mathcal{P}_\sigma = \mathcal{P}_1 + \left(\sum_{x=1}^{\sigma-1} \sum_{y>x}^{\sigma} k_x k_y \right)$$

e.g, for 2-bit: $\mathcal{P}_2 = \mathcal{P}_1 + k_1 k_2$

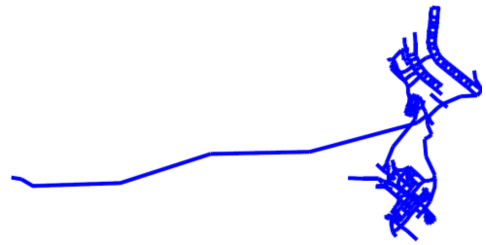
Moreover, we also have a following bound

$$\mathcal{P}_1 + (\sigma - 1)\left(k - \frac{\sigma}{2}\right) \leq \mathcal{P}_\sigma \leq \mathcal{P}_1 + \left(\frac{k^2(\sigma-1)}{2\sigma}\right)$$

Bi-level Sensing - Simulations

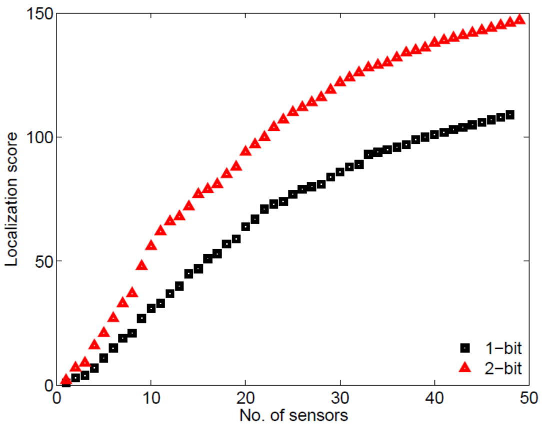


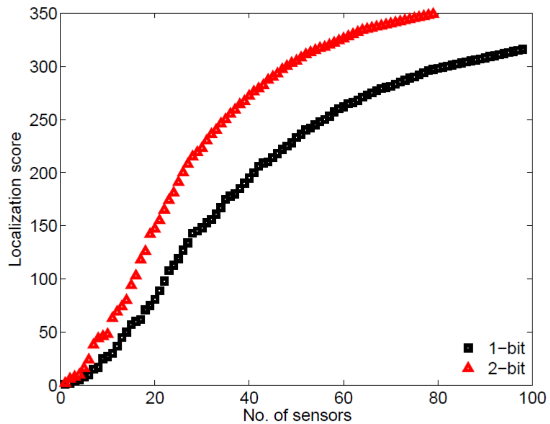
Net 1

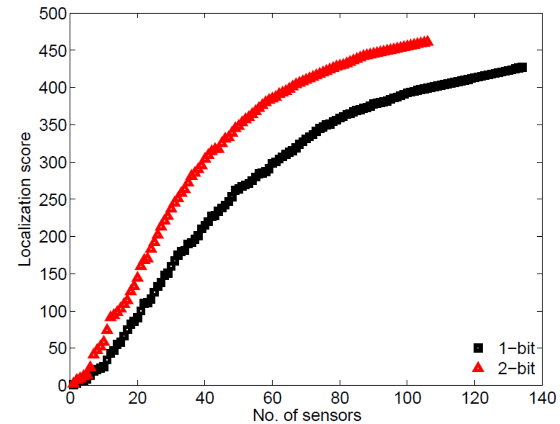


Net 2

Net 3



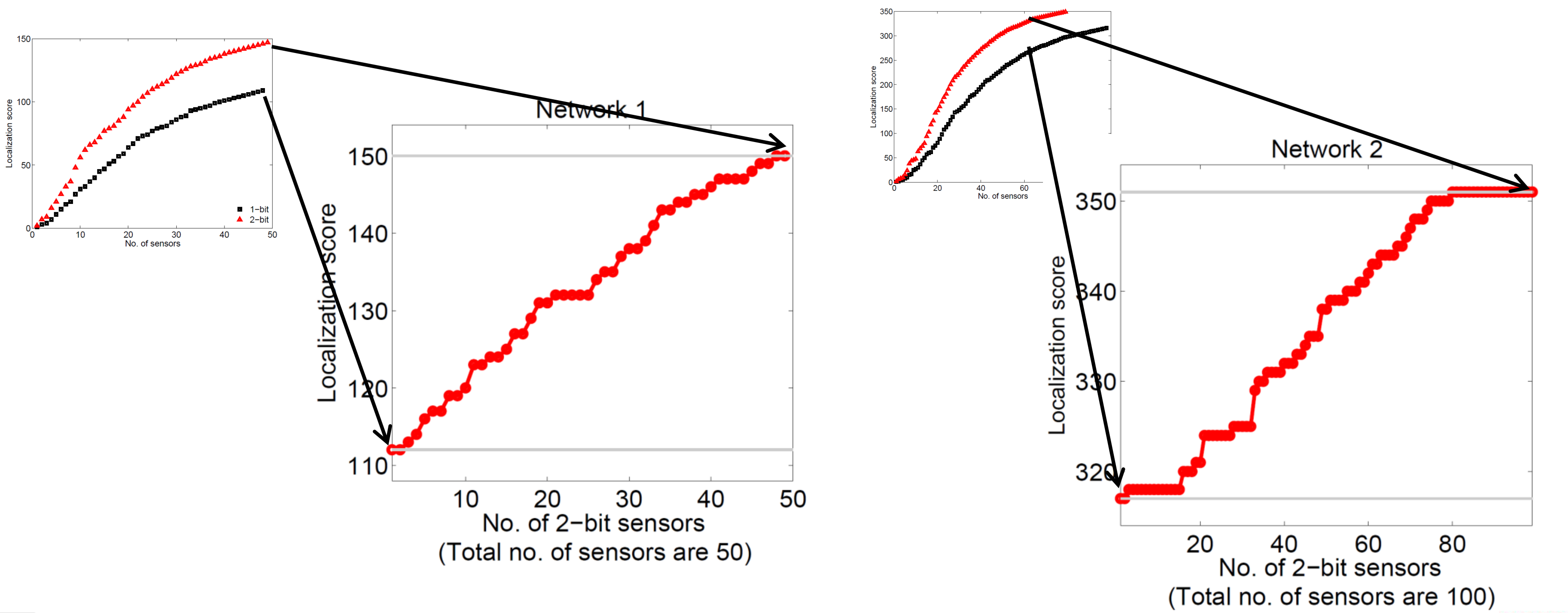




Heterogeneous Sensing

- **Homogeneous** sensors – we consider all sensors to have the same information structure, i.e. 1-bit or k-bit
- **Heterogeneous** sensors – we consider mixed information structures and explore the trend (trade-off) of the localization performance as a function of the sensing models.

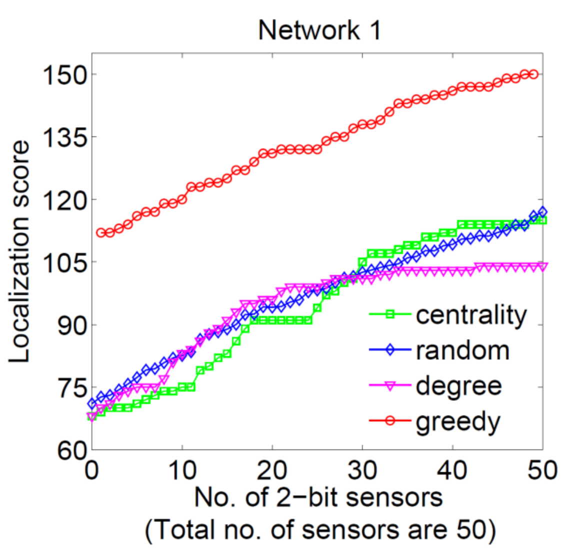
Simulations:
We use a mix of 1-bit and 2-bit sensors for the localization of link failures in Networks 1 and 2.

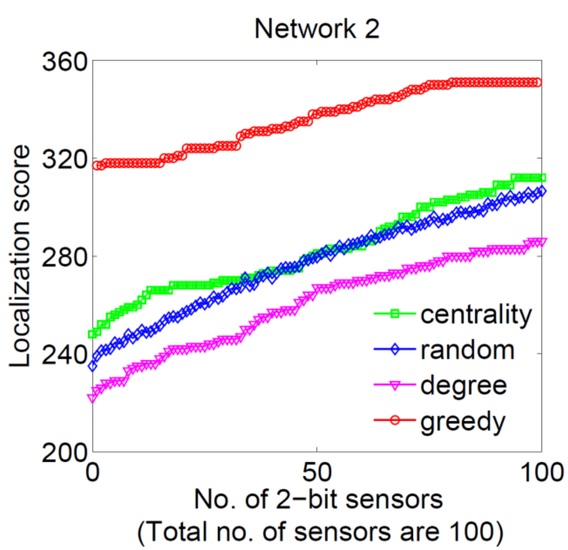


Heterogeneous Sensing

- Where should these heterogeneous sensors be deployed within the network? How can we use the **underlying network structure** to determine potential locations for heterogeneous sensors?

Simulations:
Our simulations illustrate that **purely network based metrics serve as a bad choice** for the sensor placement, in which higher level sensors are placed on the central locations in the underlying network graph.



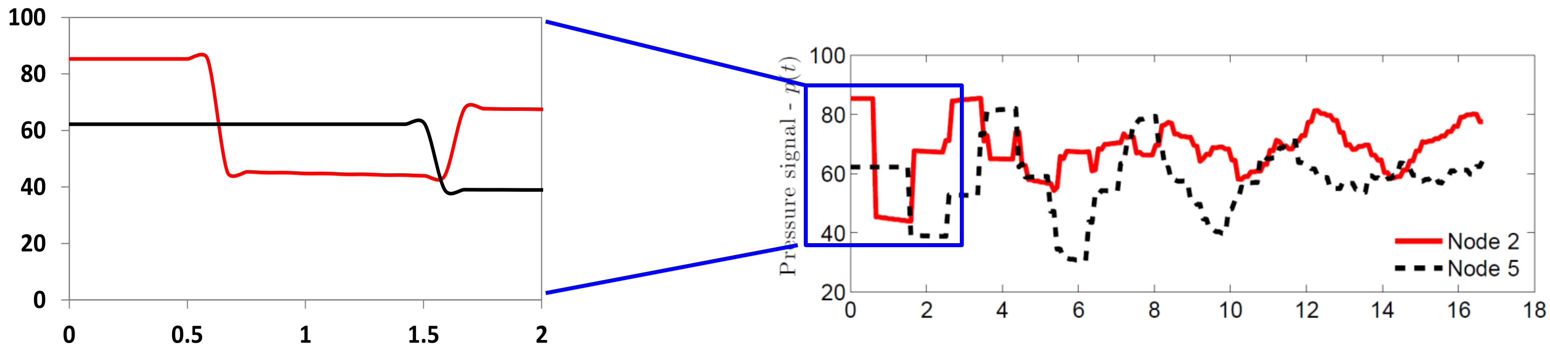


Summary and Conclusions

- The problem of identification of link failures can be posed as the **minimum test cover problem**.
- Minimum test cover for the identification of link failures in water networks can be solved using an **efficient fast greedy** algorithm.
- **Multi-level sensors** capture some extra information about the failure events, and are better for the identification of link failures as compared to the single-level sensors.
- Deploying a **combination of various types of sensors** (e.g., 1-bit and 2-bit) allow a trade-off between the localization performance and the cost entailed.

Thank You

System Model



We can introduce more symbols to represent the pressure signal. But this requires better calibrated model and more complex representation.