Resilient Sensor Placement for Fault Localization in Water Distribution Networks

W. Abbas[1], L. Perelman[2], S. Amin[3], and X. Koutsoukos[1]

[1]Vanderbilt University
[2]University of Texas Austin
[3]Massachusetts Institute of Technology

April 19, 2017

VANDERBILT UNIVERSITY

TEXAS
The University of Texas at Austin

Massachusetts Institute of Technology

Motivation

- **Leakages in Water Distribution Networks**:
  - significant economic losses
  - extra costs for final consumers
  - third-party damage and health risks

- **Sensors' failures and attacks:** Sensors for network monitoring can give errors due to
  - faults and failures
  - cyber attacks

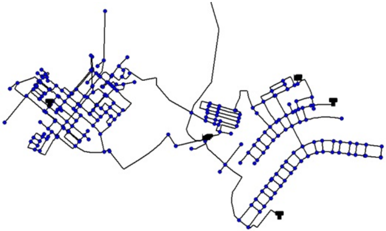
Main Objective

Sensor placement scheme that is efficient in terms of localizing pipe failures, and is also resilient to sensor errors.

Contributions

Challenges

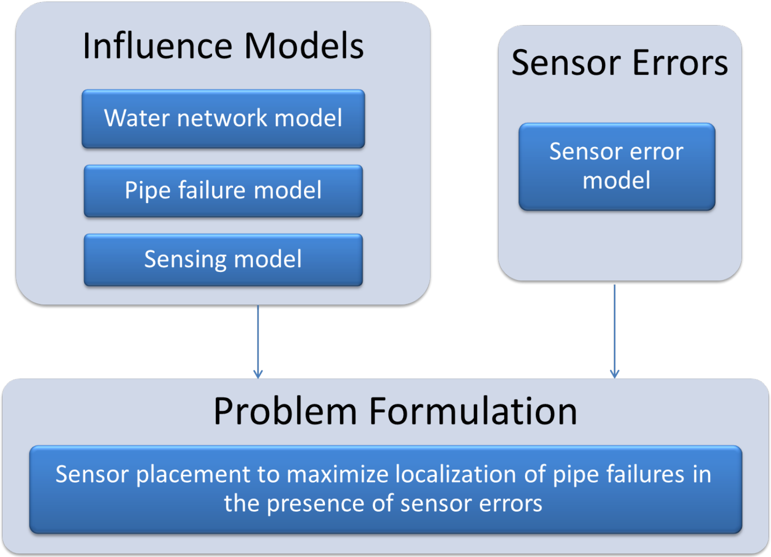
- Pipe failure uncertainty, budget constraints, uncertainty in sensing quality, event detection and localization.



Contributions

- Influence model to capture relationships between failure events and sensors.
- Optimal sensor placement to maximize localization of pipe failures with sensor errors as a combinatorial optimization problem.
- Exploring trade-offs between various system parameters, such as number of sensors, features extracted from failure signals, number of sensors with errors.

Overview



Influence Models
Water network model
Pipe failure model
Sensing model
Sensor Errors
Sensor error model
Problem Formulation
Sensor placement to maximize localization of pipe failures in the presence of sensor errors

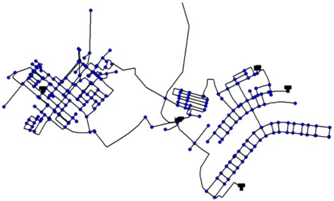
Influence Model

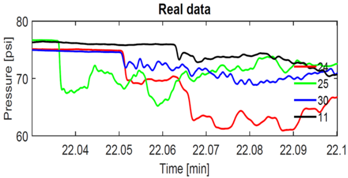
Water network $G(V, E)$

- links E model pipes
- nodes V model junctions of pipes, reservoirs, sensors

Transient model for pipe failures

- Pipe bursts propagates as a pressure wave.
- High velocity $(500 - 1400[\frac{m}{s}])$.
- Wave signal dissipates depending on traveled distance, network topology and characteristics
- At different locations, pressure signals with different magnitudes, arrival times, and shapes are observed.





Sensing model

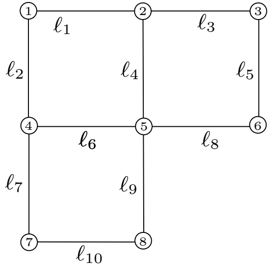
- Set of sensors: $S = \{S_1, \cdots, S_m\}$
- Set of events (pipe failures): $\mathcal{L} = \{\ell_1, \cdots, \ell_n\}$
- Set of features sensed in a transient signal: $\mathcal{Y} = \{1, \cdots, \eta\}$
- Each feature is represented by a boolean string.
- The output of a sensor i as a result of event ℓ_j :

$$S_i(\ell_j) = \left[\begin{array}{cccc} s_1(\ell_j) & s_2(\ell_j) & \cdots & s_\eta(\ell_j) \end{array} \right]$$

- The array consisting of m sensor outputs for event ℓ_j is the **signature of event ℓ_j** :

$$S(\ell_j) = \left[\begin{array}{cccc} S_1(\ell_j) & S_2(\ell_j) & \cdots & S_m(\ell_j) \end{array} \right]$$

Example:



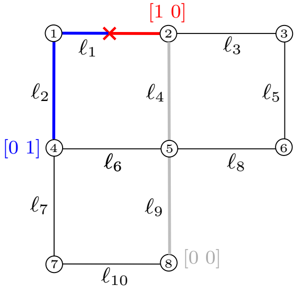
- Pipe length $= 900[m]$
- Wave propagation velocity $= 1000[\frac{m}{s}]$
- Wave dissipates after $1500[m]$.
- A sensor output consists of two bits, and has three possible outputs – [0 0], [0 1], and [1 0].
- Influence matrix =

	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	1	0	1	0	0	0	0	0	0
ℓ_2	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0
ℓ_3	0	1	1	0	1	0	0	0	0	1	0	1	0	0	0	0
ℓ_4	0	1	1	0	0	1	0	1	1	0	0	1	0	0	0	1

Influence Model

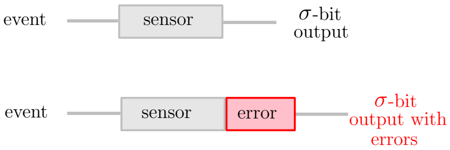
Example:

- Pipe length = 900[m]
- Wave propagation velocity = 1000[$\frac{m}{s}$]
- Wave dissipates after 1500[m].
- A sensor output consists of two bits, and has three possible outputs – [0 0], [0 1], and [1 0].
- Influence matrix =



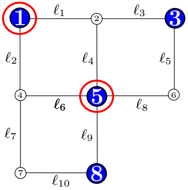
	S_1		S_2		S_3		S_4		S_5		S_6		S_7		S_8	
ℓ_1	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0
ℓ_2	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0
ℓ_3	0	1	1	0	1	0	0	0	0	1	0	1	0	0	0	0
ℓ_4	0	1	1	0	0	1	0	1	1	0	0	1	0	0	0	1

Sensor Errors



- **Error:** One or more of the output bits are flipped.
- **Error sources:** Sensor degradation, Cyber attacks
- At most e sensors can give incorrect outputs.

- **Multiple sensor errors:** Given a set of m sensors, at most e of them can give incorrect outputs for an event.
- **Example:** $e = 2$



Sensors	S_1	S_3	S_5	S_8
correct op of l_1	1 0	0 1	0 1	0 0
Possible o/p with 2 errors	0 1	1 0	1 0	0 0
	...	...	...	...

Problems

Resilient sensor placement

- How to place **m** sensors, each with a $\boldsymbol{\sigma}$**-bit** output, to maximize the number of events that can be localized accurately, even if **e** of the deployed sensors give errors?
- At the same time, how can we evaluate such a sensor placement in water distribution networks?

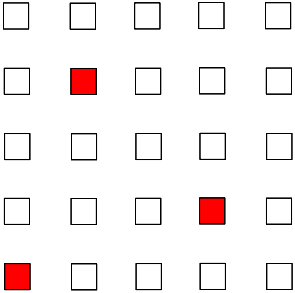
Tradeoffs

What is the trade-off between **m**, **e**, $\boldsymbol{\sigma}$, and the localization performance in the context of sensor placement for fault localization. In particular, fixing any two variables, what is the relationship between the remaining two?

Group Testing and Sensor Placement

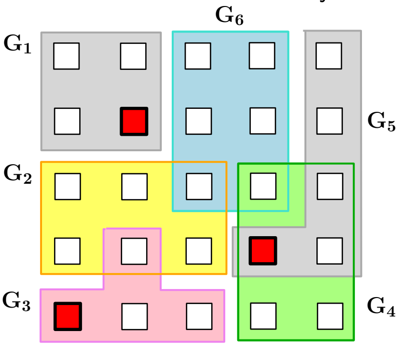
Group testing

- Set of elements out of which few are 'defective'.
- Determine the defective elements efficiently.



Group testing

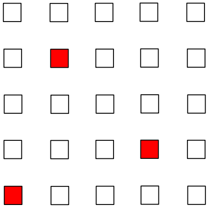
- Set of elements out of which few are 'defective'.
- Determine the defective elements efficiently.



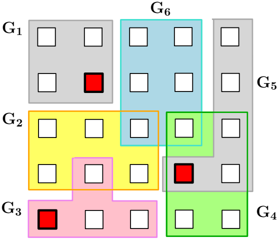
Query: Does G_i contain a defective element?
Answer: 'Yes' or 'No'.

Group Testing and Sensor Placement

- **Group testing (GT)**
  - Set of elements with some **defective** ones.
  - Elements are divided into **groups**.
  - Questions are asked, *"if G_i contains a defective element?"*
  - Answers are either *"yes"* or *"no"*.



- **Non-adaptive group testing (NAGT)**
  - All groups are made a priori.
- **NAGT with Unreliable Tests**
  - Some questions answered incorrectly.



- **Results**
  - Necessary queries: $O\left((d^2/\log d)\log n\right)$
  - Sufficient queries: $O(d^2 \log n)$

(e.g., Macula 1997, Porat and Rothschild 2011, Mazumdar and Mohajer 2014)

Group Testing and Sensor Placement

NAGT

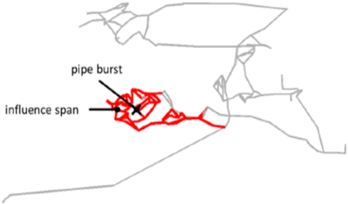
- elements
- defective elements
- groups
- tests (queries)
- unreliable tests

Resilient sensor placement

- pipes
- pipes with failures
- sensors
- sensors outputs
- sensors outputs with errors

However, there is a major difference.

- Typically in NAGT, any set of elements can be grouped together to make a test.
- In sensor placement, groups (tests) are coming from the physical system, i.e., any set of pipes cannot be grouped together.

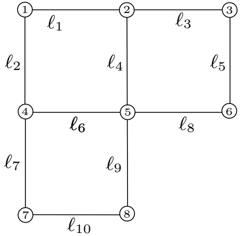


Localization of Events

- Each sensor has a σ-bit output, $S_i(\ell_x)$
- Array of sensors' outputs is $S(\ell_x) = [\ S_1(\ell_x) \ \cdots \ S_{tot}(\ell_x)\ ]$.
- **Hamming distance:** $H(S(\ell_x), S(\ell_y))$
- **Example:**
  - Assume $\sigma = 2$, then four possible outputs: $a = [0\ 0]$, $b = [0\ 1]$, $c = [1\ 0]$, and $d = [1\ 1]$.

$$S(\ell_1) = [\ c \quad c \quad b \quad b \quad b \quad a \quad a \quad a\ ]$$
$$S(\ell_2) = [\ c \quad b \quad a \quad c \quad b \quad a \quad b \quad a\ ]$$
$$H(S(\ell_1), S(\ell_2)) = 4$$

- In the case of **no errors**, sensors' output S is always a signature of some event.



ℓ_x can be distinguished from ℓ_y as long as the Hamming distance between $S(\ell_x)$ and $S(\ell_y)$ is at least one.

Localization of Events – No Sensor Errors

Select m sensors (budget) such that the number of event pairs ℓ_x, ℓ_y whose signatures have a Hamming distance of at least one is maximized.

Influence matrix =

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
ℓ_1	c	c	b	b	b	a	a	a
ℓ_2	c	b	a	c	b	a	b	a
ℓ_3	b	c	c	a	b	b	a	a
ℓ_4	b	c	b	b	c	b	a	b

Pair-wise Influence matrix =

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
$\ell_{1,2}$	0	1	1	1	0	0	1	0
$\ell_{1,3}$	1	0	1	1	0	1	0	0
$\ell_{1,4}$	1	0	0	0	1	1	0	1
$\ell_{2,3}$	1	1	1	1	0	1	1	0

Maximum coverage problem (MCP):

- Given a set of elements $\mathcal{U}$, a collection $\mathcal{C}$ of subset of $\mathcal{U}$, that is $\mathcal{C} = \{C_1, \cdots, C_t\}$, where $C_i \subset \mathcal{U}$, and a positive integer m; then select a sub-collection $\mathcal{C}_s \subset \mathcal{C}$ containing m subsets (C_i's) such that the union of subsets in $\mathcal{C}_s$ is maximized.

Sensor Placement for Localization

For the sensor placement problem,

- $\mathcal{U}$: set of all pair-wise events.
- C_i: set of pair-wise events 'detected' by the i^{th} sensor,
- $\mathcal{C} = \{C_1, C_2, \cdots, C_{tot}\}$.

Optimal sensor placement

Finding an optimal sensor placement that maximizes the localization of events is equivalent to solving the maximum coverage problem.

- It is well known that MCP is **NP-hard**.
- **Greedy** approach gives the best approximation algorithm.[1]
- For faster implementation, the overall event space (set of all pair-wise events) can be reduced.[2]

1. Vazirani, *Approximation Algorithms*, 2001.
2. Perleman, Abbas, Amin, and Koutsoukos, *Sensor placement for fault location identication in water networks: a minimum test cover approach*, Automatica, 2016.

Localization with Sensor Errors

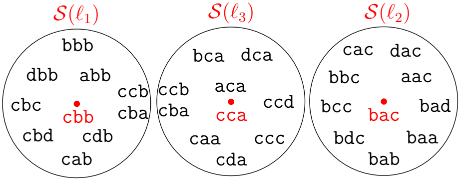
- Some sensors might give incorrect outputs, **errors**.
- We assume that at most **e** sensors can give errors.
- As a result of event ℓ_x, output $\tilde{S}(\ell_x)$ is produced such that

$$H(S(\ell_x),\ \tilde{S}(\ell_x)) \ \leq \ e$$

signature of ℓ_x

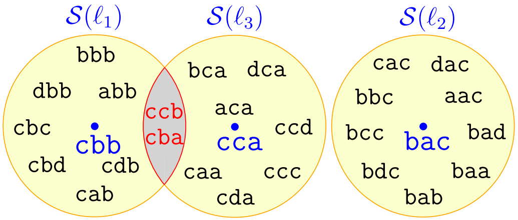
- $\mathcal{S}(\ell_x)$: set of all possible outputs corresponding to ℓ_x.

- **Example:**
  - Consider events $\{\ell_1, \ell_2, \ell_3\}$,
  - sensors $S = [S_2 \ \ S_3 \ \ S_4]$.



Localization with Sensor Errors

- Note that ℓ_1 can always be distinguished from ℓ_2 as $\mathcal{S}(\ell_1) \cap \mathcal{S}(\ell_2) = \emptyset$.
- However, ℓ_1 cannot *always* be distinguished from ℓ_3 as $\mathcal{S}(\ell_1) \cap \mathcal{S}(\ell_3) \neq \emptyset$.



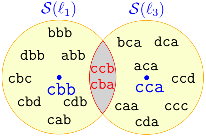
Detection of pair-wise events

In the presence of e sensor errors, ℓ_x can always be distinguished from ℓ_y (the pair-wise event $\ell_{x,y}$ is detectable) if and only if

$$H(S(\ell_x),\ S(\ell_y)) \ \geq\ (2e\ +\ 1).$$

Sensor Placement with Sensor Errors

- If $0 < H(S(\ell_x), (\ell_y)) < (2e+1)$, then still there can be outputs in $\mathcal{S}(\ell_x)$ that accurately distinguish ℓ_x from ℓ_y.



- We define

$$f(\ell_{i,j}) = \begin{cases} 1 & \text{if } H(S(\ell_i), S(\ell_j)) \geq 2e+1 \\ \frac{H(S(\ell_i), S(\ell_j))}{2e+1} & \text{otherwise.} \end{cases}.$$

Sensor placement problem

- S_{tot}: set of all sensors,
- The sensor placement problem is,

$$\underset{\mathcal{A} \subset S_{\text{tot}}}{\text{argmax}} \left(\frac{\sum_{\ell_{i,j}} f(\ell_{i,j})}{\text{Total number of pair-wise links}} \right)$$

subject to $|\mathcal{A}| \leq m$.

Sensor Placement with Sensor Errors

In terms of the pair-wise influence matrix, we need to select sensors such that each pair-wise event $\ell_{x,y}$ is covered $k = 2e + 1$ times.

- Influence matrix =

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
ℓ_1	c	c	b	b	b	a	a	a
ℓ_2	c	b	a	c	b	a	b	a
ℓ_3	b	c	c	a	b	b	a	a
ℓ_4	b	c	b	b	c	b	a	b

- Pair-wise influence matrix =

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
$\ell_{1,2}$	0	1	1	1	0	0	1	0
$\ell_{1,3}$	1	0	1	1	0	1	0	0
$\ell_{1,4}$	1	0	0	0	1	1	0	1
$\ell_{2,3}$	1	1	1	1	0	1	1	0

Set multicover problem (SMP):

Given a set of elements $\mathcal{U}$, a collection $\mathcal{C}$ of subsets of $\mathcal{U}$, that is $\mathcal{C} = \{C_1, \cdots, C_t\}$, where $C_i \subset \mathcal{U}$, and a positive integer k; then select a sub-collection $\mathcal{C}_s \subset \mathcal{C}$ such that for every $in \in \mathcal{U}$, we get $|C_j \in \mathcal{C}_s : i \in C_j| \geq k$.

For $k = 1$, the problem is a a well known set cover problem.

Sensor Placement with Sensor Errors

For set cover, and set multicover problem, greedy heuristics gives the best approximation ratios.[1,2]

Greedy heuristics to place sensors

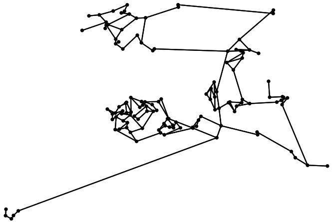
1. For each sensor S_i, compute the set of pair-wise link failures covered by the sensor.
2. In each iteration, select a sensor covering the maximum number of pair-wise link failures that are not yet covered for at least $k = 2e + 1$ times in previous iterations.
3. Perform m such iterations.

1. Feige, *A threshold of ln n for approximating set cover*, J. ACM, 1998.
2. Berman et al., *Randomized approximation algorithms for set multicover problems with applications to reverse engineering of protein and gene networks*, Discrete Applied Mathematics, 2007.

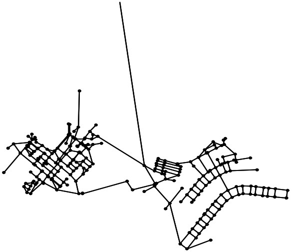
Performance Evaluation

Performance metrics

- Number of detectable pair-wise events
- Localization sets
- Identification score



- Water network 1
  (126 nodes, 168 pipes)



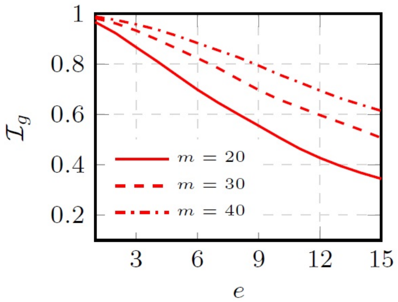
- Water network 2
  (270 nodes, 366 pipes)

Performance – Identification Score

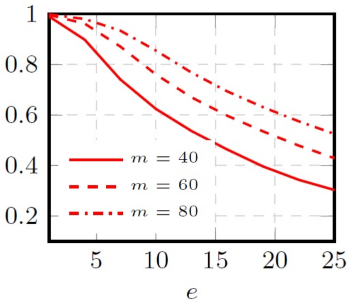
For a pair-wise event $\ell_{i,j}$, we have defines,

$$f(\ell_{i,j}) = \begin{cases} 1 & \text{if } H(S(\ell_i), S(\ell_j)) \geq 2e+1 \\ \frac{H(S(\ell_i),S(\ell_j))}{2e+1} & \text{otherwise.} \end{cases}.$$

Identification score: $\mathcal{I}_g = \frac{\sum_{\ell_{i,j}} f(\ell_{i,j})}{\text{total no. of pair-wise events}}$



Water network 1

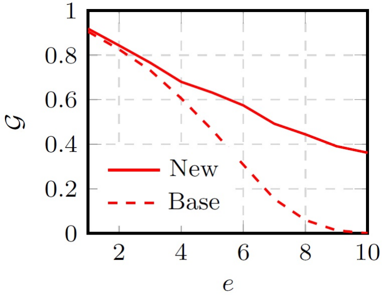


Water network 2

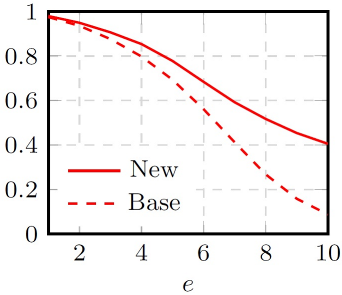
Performance – Pair-wise Events

For the pair-wise event $\ell_{i,j}$,

- **Good ($\mathcal{G}$):** $H(\tilde{S}(\ell_i),\ S(\ell_i)) > H(\tilde{S}(\ell_i),\ S(\ell_j));$ $\forall\ \tilde{S}(\ell_i)$

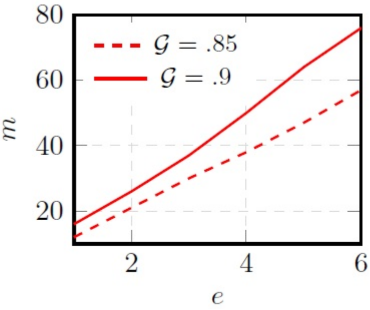


Water network 1

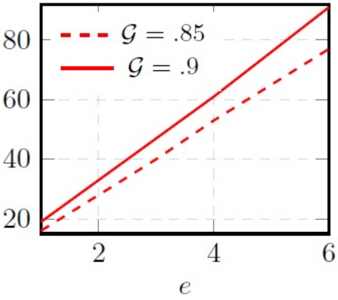


Water network 2

$\mathcal{G}$ as a function of sensor errors (e) for a fixed number of sensors (m).



Water network 1



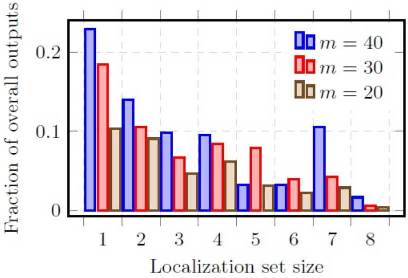
Water network 2

Number of sensors (m) as a function of sensor errors (e) for fixed $\mathcal{G}$.

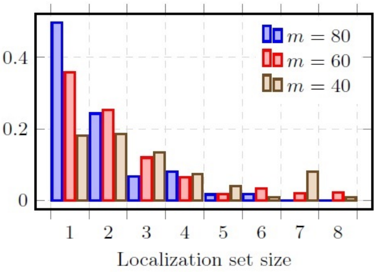
Performance – Localization Sets

In the case of event ℓ_i, $\tilde{S}(\ell_i)$ is generated.

- **Localization set:** set of signatures that are at the same Hamming distance from $\tilde{S}(\ell_i)$.



Water network 1



Water network 2

- In WN-2 the percentage of outputs with localization sets of sizes at most 5 is about 90% and 80% for $m = 80$ and 60 respectively.

Conclusions

- Optimal sensor placement to maximize localization of pipe failures with sensor errors can be formulated as a maximum k-cover problem.
- We can efficiently compute approximate solutions.
- We can further improve localization by exploiting trade-offs between the number of sensors (m), features extracted from the failure signal (σ), possible number of sensors with errors (e).

Future Work

- An integrated approach to resilient localization (redundancy + diversity + hardening).

Acknowledgments

- National Science Foundation (CNS-1238959)
- Air Force Research Laboratory (FA 8750-14-2-0180)

Thank You